Memory & Cognition
2007, 35(1), 166-175

The spread of the phonological neighborhood influences spoken word recognition

Michael S. Vitevitch
University of Kansas, Lawrence, Kansas

In three experiments, the processing of words that had the same overall number of neighbors but varied in the spread of the neighborhood (i.e., the number of individual phonemes that could be changed to form real words) was examined. In an auditory lexical decision task, a naming task, and a same–different task, words in which changes at only two phoneme positions formed neighbors were responded to more quickly than words in which changes at all three phoneme positions formed neighbors. Additional analyses ruled out an account based on the computationally derived uniqueness points of the words. Although previous studies (e.g., Luce & Pisoni, 1998) have shown that the *number* of phonological neighbors influences spoken word recognition, the present results show that the *nature* of the relationship of the neighbors to the target word—as measured by the spread of the neighborhood—also influences spoken word recognition. The implications of this result for models of spoken word recognition are discussed.

In research on spoken language processing, *neighborhood density* refers to the number of words that sound similar to a given word: Words with many neighbors, or similar words, are said to have dense neighborhoods, whereas words with few neighbors are said to have sparse neighborhoods. Several studies in English have demonstrated that neighborhood density influences various aspects of spoken language processing, including lexical acquisition (e.g., Storkel, 2002, 2004), speech production (e.g., Vitevitch, 1997, 2002b; Vitevitch & Sommers, 2003), and spoken word recognition (Luce & Pisoni, 1998; see also Vitevitch & Rodríguez, 2005, for a discussion of the influence of neighborhood density on spoken word recognition in Spanish).

In several laboratory-based spoken word recognition tasks, Luce and Pisoni (1998) demonstrated that English words with sparse neighborhoods are responded to more quickly and accurately than those with dense neighborhoods, suggesting that multiple word forms are activated and compete with each other during spoken word recognition. Words with large numbers of phonological neighbors (i.e., dense neighborhoods) are subject to greater competition and therefore recognized more slowly and less accurately than words with few phonological neighbors (i.e., sparse neighborhoods).

Vitevitch (2002c) observed a similar processing disadvantage for words with dense neighborhoods in an analysis of a corpus containing speech perception errors, known as "slips of the ear," that were collected via naturalistic observation. An example of a slip of the ear is erroneously hearing the correctly produced question "What's wrong with her *bike*?" as "What's wrong with her *back*?" (Bond, 1999). In analyzing the misperceived words in Bond's corpus, Vitevitch (2002c) found that slips of the ear tended to occur in words with dense phonological neighborhoods, further suggesting that multiple word forms are activated and compete during spoken word recognition.

The previously discussed studies clearly demonstrate that the *number* of phonologically related word forms that are activated influences spoken word recognition: Words with few neighbors are recognized more quickly and more accurately than words with many neighbors in English. Now, consider two words with the same number of phonological neighbors. Does some other factor, such as the distribution of the neighbors in the lexical neighborhood, influence the speed and accuracy of spoken word recognition? By way of illustration, consider the words *mop* (/mɑp/) and *mob* (/mɑb/). When a single phoneme substitutes any of the phonemes in the word *mop*, phonological neighbors are formed (e.g., *h̲op*, *ma̲p*, *moc̲k̲*). However, similar substitutions in the word *mob* produce phonological neighbors at only two of the three phoneme positions (e.g., *r̲ob*, *m*b*, *moc̲k̲*); no real word in English is formed when the phoneme in the medial position of the word *mob* is substituted. Note that each word has the same total number of phonological neighbors, but that the number of phoneme positions in the word that produce a neighbor differs between the two words.

To investigate the possible influence of the distribution of similar-sounding neighbors in the phonological neighborhood on spoken word recognition, a phonological analogue of a metric used in studies of visual word recognition—the *spread* of a neighborhood, or the *P-metric* (Andrews, 1997; Johnson & Pugh, 1994; Pugh,

M. S. Vitevitch, mvitevit@ku.edu

Rexer, Peter, & Katz, 1994)—was manipulated in several behavioral tasks. *Spread* refers to the number of phoneme positions (or letter positions, as in Johnson & Pugh, 1994) in a word that can be changed to form a neighbor. In the examples above, the word *mop* has a *P*-metric value of 3 ($P = 3$) because changes at three phoneme positions produce phonological neighbors, whereas the word *mob* has a *P*-metric value of 2 ($P = 2$) because changes at two phoneme positions produce phonological neighbors. If the distribution of phonological neighbors in the lexical neighborhood influences spoken word recognition, then differences should be observed in terms of the speed and accuracy with which these two types of words ($P = 2$ vs. $P = 3$) are responded to. The same stimuli were presented in three different laboratory-based tasks to evaluate the influence of neighborhood spread on the speed and accuracy of spoken word recognition.

EXPERIMENT 1

To examine how the spread of phonological neighbors in the similarity neighborhood might affect spoken word recognition, a lexical decision task was used. In the lexical decision task, participants are presented with a stimulus item and must decide as quickly and accurately as possible if that item is a real word in English or a nonsense word. In the present experiment, the stimuli were presented auditorily rather than visually as in Johnson and Pugh (1994) and varied in phonological *P* rather than orthographic *P*. The stimuli that the participants heard consisted of three-phoneme words, with a consonant–vowel–consonant (CVC) syllable structure, that had the same number of phonological neighbors but differed in how those neighbors were spread about the neighborhood. For half of the words, $P = 2$, meaning that a change in either of two phoneme positions produced a neighbor; for the remaining words, $P = 3$, meaning that a change in any of all three phonemes in the word produced a neighbor.

Method

Participants. Forty right-handed native English speakers from the pool of introductory psychology students at the University of Kansas participated in partial fulfillment of a course requirement. None of the participants reported a history of speech or hearing problems, and none of them participated in either of the other experiments reported in the present study.

Stimuli. Ninety-two CVC words were used as stimuli in the experiment (see Appendix A). The stimuli were divided into two sets of 46 words each. One set contained words that formed a neighbor when a single phoneme was substituted (e.g., Landauer & Streeter, 1973; Luce & Pisoni, 1998) at any of the three phoneme positions of the word ($P = 3$). The other set contained words that formed a neighbor when a single phoneme substitution could be made at one of only two phoneme positions of the word to form a neighbor ($P = 2$). Words for which $P = 1$ were not examined because of the paucity of words in this category.

Although the two sets of words differed in the number of phonemes that could be changed to form a neighbor, they did not differ [all Fs$(1,90) < 1$] in the overall number of neighbors (i.e., neighborhood density), word familiarity (Nusbaum, Pisoni, & Davis, 1984), word frequency (Kučera & Francis, 1967), the frequency with which the neighbors occurred (i.e., neighborhood frequency; Kučera & Francis, 1967), or phonotactic probability (Vitevitch & Luce, 1998, 1999, 2005). Note that information related to the familiarity, frequency, neighborhood density, and neighborhood frequency for each word can be obtained from a Web-based interface maintained by Mitchell Sommers at Washington University (128.252.27.56/neighborhood/Home.asp). Information related to the phonotactic probability of each word can be obtained from a Web-based interface (www.people.ku.edu/~mvitevit/PhonoProbHome.html) described in Vitevitch and Luce (2004). The mean values for these characteristics for each set of words are presented in Table 1. The same number of initial segments appeared in each condition.

In addition, onset density did not differ between the two conditions of words [$F(1,90) < 1$]. *Onset density* refers to the proportion of neighbors that share the same initial phoneme as the target word (Vitevitch, 2002a). For words for which $P = 2$, the mean proportion of neighbors that shared the same initial phoneme as the target word was .60, whereas for words for which $P = 3$ the mean proportion was .59.

Although the stimuli were presented auditorily rather than visually (cf. Johnson & Pugh, 1994), the two conditions of words did not differ in the number of letters comprising the words [$F(1,90) < 1$]. Words for which $P = 2$ had a mean of 4.5 letters per word ($SD = 0.81$), and words for which $P = 3$ had a mean of 4.4 letters per word ($SD = 0.77$). The two conditions of words also did not differ in the number of orthographic neighbors [$F(1,90) < 1$]. Words for which $P = 2$ had a mean of 4.9 orthographic neighbors ($SD = 4.2$), and words for which $P = 3$ had a mean of 5.5 orthographic neighbors ($SD = 4.3$), on the basis of calculations from the N-Watch program described by Davis (2005).

The stimuli were spoken in isolation and recorded by the author in an IAC sound-attenuated booth on high-quality audio-recording equipment. The stimuli were digitized at a sampling rate of 20 kHz using a 16-bit analog-to-digital converter. All words were edited into individual digital files and stored on a computer disk for later presentation. Stimuli in the $P = 2$ condition had a mean file duration of 863 msec ($SD = 105$), and stimuli in the $P = 3$ condition had a mean file duration of 851 msec ($SD = 100$); this difference was not statistically significant [$F(1,90) = 0.30$, $p > .5$]. Ninety-two nonsense words were also used (see Appendix B). The method used to create nonwords in previous studies (e.g., Vitevitch, 2002a) was used in the present experiment: The last phonemes of words not found in the stimulus set were changed to create the nonwords that were used. Only the last phoneme was changed to increase the likelihood that the participants would listen to the entire stimulus item

Table 1
Mean Values (and Standard Deviations) for the Lexical Characteristics of the Stimuli

Characteristic	$P = 2$ *M*	$P = 2$ *SD*	$P = 3$ *M*	$P = 3$ *SD*
Frequency of occurrence (log)	1.00	0.76	1.10	0.62
Familiarity*	6.86	0.28	6.88	0.20
Neighborhood density†	8.7	3.50	9.20	1.90
Neighborhood frequency	1.23	0.37	1.24	0.31
Phonotactic probability				
Sum of phones	.116	.05	.113	.04
Sum of biphones	.004	.004	.004	.003

Note—No differences were statistically significant [all Fs$(1,90) < 1$]. *Based on a 7-point scale. †A word was considered a neighbor if a substitution of a phoneme in the target word formed that word and it appeared in the computer-readable phonemically transcribed *Webster's Pocket Dictionary* (Nusbaum et al., 1984). This method of determining neighborhood size was consistent with the method employed by Johnson and Pugh (1994), with the exception that phonemes rather than letters were substituted (i.e., the *N*-metric commonly attributed to Coltheart, Davelaar, Jonasson, & Besner, 1977).

before making a response. The nonwords were recorded and treated in the same manner as the real word stimuli.

Procedure. The participants were tested in groups of 4 or fewer. Each participant was seated in a booth equipped with an iMac running PsyScope 1.2.2 (Cohen, MacWhinney, Flatt, & Provost, 1993) that controlled stimulus randomization and presentation, a set of Beyerdynamic DT-100 headphones, and a PsyScope buttonbox with a dedicated timing board. Each trial proceeded as follows: The word READY appeared in the center of the computer screen for 500 msec to indicate the beginning of the trial. The participants were then presented with one of the randomly selected stimuli at a comfortable listening level over the headphones. The left button on the response box was labeled NONWORD, and the right button (i.e., that for the dominant hands of the participants) was labeled WORD. The participants responded as quickly and accurately as possible by pushing the appropriately labeled button. Reaction time was measured from the onset of the stimulus file to the onset of the response. Prior to the experimental trials, each participant received 10 practice trials. These trials were used to familiarize the participants with the task and were not included in the final data analysis.

Results and Discussion

Separate ANOVAs were performed on response latency and accuracy rates with participants and items treated as random factors. Although there is some debate about whether or not to treat stimulus items as a random factor in statistical analyses (Cohen, 1976; Hino & Lupker, 2000; Keppel, 1976; Raaijmakers, 2003; Raaijmakers, Schrijnemakers, & Gremmen, 1999; Smith, 1976; Wike & Church, 1976), it is the current practice in psycholinguistic research to conduct both types of analyses. For consistency with this convention, both types of analyses will be reported; however, the discussion and interpretation of the results will be based only on the analyses in which participants were treated as a random factor. Also, estimates of effect size will be conducted only on the analyses in which participants were treated as a random factor.

Only correct responses to the stimulus items within 2 *SD*s of the mean response time were used in the analyses of response latency. A significant difference in response latencies was found in the lexical decision task [$F(1,39) = 39.76$, $p < .001$] given that the participants responded more quickly to words for which $P = 2$ ($M =$ 1,080 msec, $SD = 99$) than to words for which $P = 3$ ($M =$ 1,115 msec, $SD = 95$). The same pattern of results was obtained when stimulus items were treated as a random factor [$F(1,90) = 4.57$, $p < .05$]. An estimate of effect size using Cohen's *d* shows that this can be considered a medium-sized effect ($d = 0.36$).

No significant difference was found for the accuracy rate in the lexical decision task (both *F*s < 1), suggesting that the participants did not sacrifice speed for accuracy in making their responses. The participants responded to words for which $P = 2$ with 90% accuracy ($SD = 4.8$) and to words for which $P = 3$ with 91% accuracy ($SD =$ 5.0).

The results of the auditory lexical decision task showed that words for which $P = 2$ were responded to more quickly than words for which $P = 3$, even though these two sets of words had comparable numbers of neighbors overall. These results extend the work of Johnson and Pugh (1994), who examined neighborhood spread in visual word recognition, to the auditory domain. Recall that in the present experiment the number of phoneme positions, rather than the number of letter positions, was manipulated, and an auditory lexical decision task was employed rather than a visual lexical decision task. To further examine the influence of neighborhood spread on spoken word recognition, an auditory naming task was performed in Experiment 2, and an auditory same–different task was performed in Experiment 3.

EXPERIMENT 2

In the present experiment, an auditory naming task was used to further examine how the spread of phonological neighbors in the similarity neighborhood might affect spoken word recognition. In the auditory naming task, a word is presented to participants over a set of headphones, and they must simply repeat the word as quickly and accurately as possible. This task, as well as the same–different task in Experiment 3, was used to better generalize the results observed in Experiment 1 (and those of Johnson & Pugh, 1994), in which the lexical decision task was employed. Because every task used in laboratory settings has advantages and disadvantages, replication across a variety of tasks increases our confidence that the observed effect was not due to the assumptions of a particular task employed in a particular experiment. Furthermore, Wike and Church (1976) recommended replication as a means of generalizing results without resorting to statistical techniques that might be inappropriate, such as analyses that treat stimulus items as a random factor.

Method

Participants. Thirty native English speakers from the pool of introductory psychology students at the University of Kansas participated in partial fulfillment of a course requirement. None of the participants reported a history of speech or hearing problems, and none of them had participated in either of the other experiments reported in the present study.

Stimuli. The stimuli consisted of the same 92 words manipulated for neighborhood spread that were used as stimuli in Experiment 1.

Procedure. The participants were tested 1 at a time. Each participant was seated in a booth equipped with an iMac running PsyScope 1.2.2 (Cohen et al., 1993), which controlled stimulus randomization and presentation; a set of Beyerdynamic DT-109 headphones; and a PsyScope buttonbox with a dedicated timing board. Each trial proceeded as follows: The word READY appeared in the center of the computer screen for 500 msec to indicate the beginning of the trial. The participant was then presented with one of the randomly selected stimuli at a comfortable listening level over the headphones. Response latency was measured from the onset of the stimulus file to the onset of the participant's response. When a response was made, the word READY appeared on the screen and the next trial began. Responses were also recorded on digital audio tape for later accuracy analyses. Prior to the experimental trials, each participant received 10 practice trials. None of the items used in the practice session was used in the experiment. The practice trials were used to familiarize the participants with the task, and the data collected from them were not included in the final analysis. The participants were instructed to respond as quickly and accurately as possible.

Results and Discussion

As in Experiment 1, separate ANOVAs were performed on response latency and accuracy rates with participants and items treated as random factors. Only correct responses within 2 *SD*s of the mean response time were used in the analyses of response latency. An accurate response was one in which each phonological segment in the verbal response made by a participant matched the segments in a phonological transcription of the stimulus word as judged by a trained speech scientist (see Vitevitch & Luce, 2005).

A significant difference in response latencies was found [$F(1,29) = 126.04$, $p < .001$] given that the participants responded more quickly to words for which $P = 2$ (M = 1,018 msec, $SD = 144$) than to words for which $P = 3$ (M = 1,056 msec, $SD = 140$).[1] The same pattern of results was observed when stimulus items were treated as a random factor [$F(1,90) = 6.78$, $p < .01$]. An estimate of effect size using Cohen's *d* shows that this can be considered an effect of small to medium size ($d = 0.26$).

No significant differences were found for the accuracy rates in the naming task (both *F*s < 1), suggesting that the participants did not sacrifice speed for accuracy in making their responses. The participants responded to words for which $P = 2$ with 94% accuracy ($SD = 4.2$) and to words for which $P = 3$ with 95% accuracy ($SD = 5.1$).

The results of the auditory naming task are consistent with the results obtained in Experiment 1 using the auditory lexical decision task: Words for which $P = 2$ were responded to more quickly than words for which $P = 3$, even though the two sets of words had comparable numbers of neighbors overall. These results further extend the work of Johnson and Pugh (1994), who examined only neighborhood spread with the lexical decision task (and only in the visual modality). An auditory same–different task was performed in Experiment 3 to further generalize the results observed in Experiments 1 and 2.

EXPERIMENT 3

In the present experiment, an auditory same–different task was used to further examine how the spread of phonological neighbors in the similarity neighborhood might affect spoken word recognition. In the auditory same–different task, participants hear two words presented close together in time and must decide as quickly and accurately as possible whether the two words were the same (e.g., *dog–dog*) or different (e.g., *dog–doll*).

Method

Participants. Thirty-eight right-handed native English speakers from the pool of introductory psychology students at the University of Kansas participated in partial fulfillment of a course requirement. None of the participants reported a history of speech or hearing problems.

Stimuli. The stimuli consisted of the same 92 words manipulated for neighborhood spread that were used as stimuli in Experiments 1 and 2, and 184 additional English words that were recorded and edited in the same fashion as the other stimuli.

Procedure. The equipment used in Experiment 1 was also used in the present experiment. Each experimental trial proceeded as follows: The word READY appeared in the center of the computer screen for 500 msec to indicate the beginning of the trial. The participants were then presented with two of the spoken stimuli at a comfortable listening level. The interstimulus interval was 50 msec. Reaction times were measured from the onset of the second sound file in the pair to the buttonpress response. The participants were instructed to respond as quickly and accurately as possible on each trial. The buttonbox had the label DIFFERENT on the left button and the label SAME on the right button (the middle response button was deactivated). Half of the trials consisted of two presentations of the stimulus items (constituting "same" trials), and half consisted of nonmatching stimuli (constituting "different" trials). For the "different" stimulus pairs (listed in Appendix C), items with the same initial phoneme and (when possible) the same vowel were paired to increase the likelihood that the participants would listen to both words in the stimulus pair and base their decisions on both words rather than adopt a strategy of simply listening for the match (or mismatch) of the initial phonemes of each word in the pair. Each participant was allowed 10 practice trials prior to the experimental trials. These trials were used to familiarize the participants with the task and were not included in the final analysis.

Results and Discussion

As in Experiments 1 and 2, separate ANOVAs were performed on response latency and accuracy rates with participants and stimulus items treated as random factors. Only correct responses within 2 *SD*s of the mean response time were used in the analyses of response latency. A significant difference in response latencies was found [$F(1,37) = 35.288$, $p < .001$] given that the participants responded "same" more quickly to words for which $P = 2$ ($M = 819$ msec, $SD = 87$) than to words for which $P = 3$ ($M = 859$ msec, $SD = 92$). The same pattern of results was obtained when stimulus items were treated as a random factor [$F(1,90) = 7.43$, $p < .01$]. An estimate of effect size using Cohen's *d* shows that this can be considered a medium-sized effect ($d = 0.44$).

No significant differences were found for the accuracy rates in the auditory same–different matching task (both *F*s < 1), suggesting that the participants did not sacrifice speed for accuracy in making their responses. The participants responded to both types of words with 96% accuracy ($SD = 4$ in both cases).

The results of the auditory same–different task in the present experiment are consistent with the results of Experiments 1 and 2: Words for which $P = 2$ were responded to more quickly than words for which $P = 3$, even though the two sets of words had comparable numbers of neighbors overall. The results of the present set of experiments further generalize the work of Johnson and Pugh (1994), who examined only neighborhood spread with the lexical decision task, and only in the visual modality.

GENERAL DISCUSSION

Previous studies demonstrated that the *number* of words in the phonological neighborhood influences the speed and accuracy of spoken word recognition. In English, words with few neighbors (i.e., those with sparse phonological neighborhoods) are recognized more quickly and accurately than words with many neighbors (i.e., those with dense phonological neighborhoods) (Luce & Pisoni, 1998; Vitevitch, 2002a; cf. Vitevitch & Rodríguez, 2005). The results of our Experiments 1–3 clearly demonstrate that the *spread* of the neighborhood also influences spoken word recognition. Specifically, words with two phoneme positions that can be changed to form a neighbor ($P = 2$) were responded to more quickly than words with three phoneme positions that can be changed to form a neighbor ($P = 3$), despite their having comparable numbers of neighbors overall. Although current models of spoken word recognition can account for processing differences that result from different numbers of competitors (see, e.g., Auer & Luce, 2005; Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris, 1994), it is not clear whether or not each of these models can account for the results of the present set of experiments, in which words with equal numbers of neighbors were differentially responded to as a function of the spread of the neighborhood.

We shall first consider cohort theory because Johnson and Pugh accounted for their findings with a model of visual word recognition based on the assumptions of the cohort theory of spoken word recognition proposed by Marslen-Wilson and Welsh (1978). Recall that Marslen-Wilson and Welsh suggested that acoustic–phonetic information activates a set of lexical candidates (i.e., the *cohort*) that is consistent with the input. As more of the word is heard, additional acoustic–phonetic information accumulates. Candidates that are no longer consistent with the additional input drop out of the cohort. Once sufficient information has accrued to distinguish the input from all other words in the cohort of partially activated candidates, word recognition is said to occur. Using a gating task, in which listeners attempt to identify the stimulus word as increasingly larger portions of the word are presented auditorily, Grosjean (1980) found that words in which this *recognition point* occurred early were correctly identified sooner (i.e., with fewer "gates") than words in which the recognition point occurred later. Thus, one might hypothesize that, in the present set of experiments, words for which $P = 2$ had earlier recognition points than words for which $P = 3$, thereby accounting for the difference in response times in all three experiments.

To examine the possibility that in the present set of experiments words for which $P = 2$ had earlier recognition points than words for which $P = 3$, the recognition points, or the computationally derived *uniqueness points*, of the stimulus items were examined. Note that use of the terms *isolation point*, *uniqueness point*, and *recognition point* is not consistent in the literature (cf. Bölte & Uhe, 2004; Grosjean, 1996; Radeau & Morais, 1990). In the present context, the term *uniqueness point* will be used to refer to the point in a word at which it becomes unique from all other words in the lexicon, as assessed via computational search through a corpus of English words (i.e., the same corpus used to estimate phonological neighborhood density in the present study). Uniqueness points differ from recognition or isolation points, which are empirically derived via the gating task (see, e.g., Grosjean, 1980, 1996). Note furthermore that there is some debate about the psychological validity of such constructs in the processing of fluent speech (cf. Bölte & Uhe, 2004, and Radeau, Morais, Mousty, & Bertelson, 2000).

In an analysis of computationally derived uniqueness points, Luce (1986) found that the uniqueness point for monosyllabic words in English—such as those used in the present set of experiments—typically occurred after the end of the word. That is, the sound sequences that comprise many monosyllabic words are also part of longer words (e.g., *car–card*, *cat–cattle–catalog*), which means that listeners need to hear the beginning of the next word before they can be sure they have reached the end of the present monosyllabic word and correctly recognize it.

In the stimuli used in the present experiments, an analysis of the uniqueness points of the stimulus items showed that words for which $P = 2$ had a mean uniqueness point at 3.6 phonemes ($SD = 0.6$) and words for which $P = 3$ had a mean uniqueness point at 3.7 phonemes ($SD = 0.5$); this difference was not statistically significant [$F(1,90) = 1.95$, $p = .17$]. Recall that the stimuli used in the present experiments consisted of monosyllabic words that were three phonemes long. Uniqueness points greater than three indicate that the three-phoneme-long monosyllabic stimulus items did not diverge from other words in the lexicon until after the offset of the word, which is consistent with the results reported by Luce (1986) for monosyllabic words. Furthermore, words for which $P = 2$ did not diverge from other words in the lexicon sooner than did words for which $P = 3$, suggesting that differences in the uniqueness points of the stimulus words cannot account for the results of the present set of experiments. Although Johnson and Pugh (1994) interpreted their results in terms of a cohort-based model, it is unlikely that such an account can explain the results of the present set of experiments.

Rather than being a proxy measure for the uniqueness point,[2] the spread of the neighborhood, or P-metric, seems to measure some other lexical construct. That is, P measures the distribution of phonological neighbors in the similarity neighborhood. As was demonstrated in Experiments 1–3, spoken word recognition is significantly affected by the distribution of phonological neighbors in the similarity neighborhood. Specifically, words with neighbors that are "packed" into fewer regions of the neighborhood are responded to more quickly than words with neighbors spread throughout the neighborhood. Given the emphasis that cohort theory places on the initial portion of word forms and the clear evidence (provided in the present set of experiments) that neighbors located in other parts of the word influence processing, it is unlikely that cohort theory can account for the present results.

Given that TRACE (McClelland & Elman, 1986)—a computational model that accounts for numerous effects observed in studies of spoken word recognition—incorporates several assumptions of cohort theory into its design, it is logical to consider this model next. As McClelland and Elman discussed, potential lexical candidates in TRACE, as in cohort theory, are activated as the acoustic–phonetic input accrues over time. Thus, as in cohort theory, the initial portion of the word is important for activating potential lexical candidates in TRACE. As described above, relying on the initial portion of the word proved problematic for cohort theory in accounting for the present results, and, thus, one might expect that TRACE would also fail to account for these results.

In contrast to the earlier cohort theory, however, other parts of the word can also partially activate lexical candidates, enabling TRACE to correctly retrieve a lexical item despite a distortion in the beginning of the word (e.g., recognizing *dwibble* as the word *dribble*). Indeed, Allopenna, Magnuson, and Tanenhaus (1998) used an eyetracking task to provide evidence that the initial parts of a word (i.e., the cohort) and the rhyme portion of a word activated lexical competitors. Furthermore, the time course and probabilities of eye movements obtained by Allopenna et al. closely corresponded to the response probabilities derived from simulations of TRACE. Given the fact that other portions of the acoustic–phonetic input can continuously activate lexical candidates in TRACE, it is possible that this model might be able to account for the effects observed in the present set of experiments—that is, TRACE might be able to account not only for the influence of the number of lexical competitors on processing, but also for the influence that the location of those neighbors in the neighborhood has on processing that has been demonstrated in the present study. The previous statement should be interpreted cautiously, however, given the inherent difficulty of predicting exactly how complex computational models might perform without examining an actual simulation (Lewandowsky, 1993).

In response to the interactive nature of TRACE, Norris (1994) developed *Shortlist*, a feedforward model of spoken word recognition (see also MERGE, the feedforward model of speech recognition; Norris, McQueen, & Cutler, 2000). Although Shortlist differs from TRACE with regard to the existence of feedback between levels, the models are similar in that initial and subsequent input influence lexical retrieval in both models. Indeed, Norris demonstrated that Shortlist correctly activates the word *cigarette* (/sɪgərɛt/) even when it is presented with input that contains a mispronunciation in the initial portion of the word (e.g., /ʃɪgərɛt/). Thus, despite the noninteractive architecture of Shortlist, it too might be able to account for the present set of results. Again, however, caution should be exercised when the computational simulation is not actually performed.

Luce and Pisoni (1998) described another model of spoken word recognition—the *neighborhood activation model* (NAM)—which accounts for the influence of the intelligibility of the stimulus words, the frequency of occurrence of the stimulus words, and the number of lexical competitors (as well as the frequency of occurrence of the competitors) on processing. In assessing the confusability of the stimulus word and its competitors, NAM places equal weight on each phoneme (regardless of whether it is a consonant or a vowel) and on the position of each phoneme (regardless of whether the phoneme occurs in the onset, the nucleus, or the coda position) in a word. Given that all phoneme positions are treated equally in NAM, it is unclear whether NAM would be able to account for the results of the present experiment (or for those of Vitevitch, 2002a), which demonstrate that some phoneme positions do influence spoken word recognition more than others. As the present experiments demonstrate, phoneme positions that form a neighbor influence spoken word recognition differently than do those that do not form a neighbor.

Although the original NAM might have problems accounting for the results of the present experiments, a more recent connectionist instantiation of NAM, dubbed *PARSYN* (Auer & Luce, 2005), might be able to account for the present findings (as well as for those of Vitevitch, 2002a). In PARSYN, *par*adigmatic and *syn*tagmatic representations are activated (hence the name) as a spoken word is presented. Paradigmatic states refer to the number of alternatives active at a given point in time, whereas syntagmatic states refer to patterns that occur over time. In the case of the word *cat*, the paradigmatic representations activated would include the initial phoneme /k/ as well as other related phonemes, such as /b/ (a stop that differs from /k/ in place of articulation and voicing) and /g/ (a stop that differs from /k/ in voicing). Syntagmatic states that would be highly activated in the case of /kæt/ would include representations of the pattern of sounds /kæ/ and /æt/, whereas related but less common sequences of segments (such as /ki/ or /æv/) would be less active. By considering the dynamic interaction of paradigmatic and syntagmatic states, PARSYN can account for many aspects of spoken word recognition (Auer & Luce, 2005; see also Luce, Goldinger, Auer, & Vitevitch, 2000). Given that PARSYN takes the number of competitors (i.e., paradigmatic information) as well as the distribution of those representations over time (i.e., syntagmatic information, which would convey some information about phoneme position), it is possible that PARSYN could account for the results observed in the present set of experiments. However, as was stated in the discussions of TRACE and Shortlist, we must be cautious in predicting exactly how a complex computational model might perform without examining an actual simulation (Lewandowsky, 1993).

Previous research on spoken word recognition (as well as speech production and word learning) has focused much attention on the influence that the number of phonological neighbors has on processing. The present set of studies (see also Vitevitch, 2002a) demonstrates that the *distribution* of neighbors in the neighborhood also influences processing. Models of spoken word recognition must account not only for the influence of the *number* of competitors on processing, but—in the absence of a difference in the number of competitors—also for the influence of the *location* of competitors on processing. Thus, the number of neighbors, as well as the relationship among the neighbors, appears to provide an important, but different, kind of constraint on spoken word recognition.

AUTHOR NOTE

This research was supported in part by Grants NIDCD R03 DC 04259 and NIDCD R01 DC 006472 from the National Institutes of Health to the University of Kansas through the Schiefelbusch Institute for Life Span Studies; National Institute of Child Health and Human Development Grant P30 HD002528 from the Mental Retardation and Developmental Disabilities Research Center; and Grant NIDCD P30 DC 005803 from the Center for Biobehavioral Neurosciences in Communication Disorders. I thank Shinying Chu, David Levine, and Thu Vo for their assistance in data collection, and Steven B. Chin, Luis Hernandez, Lorin Lachs, David Pisoni, and Holly Storkel for helpful discussions. Correspondence regarding this article should be addressed to M. S. Vitevitch, Spoken Language Laboratory, Department of Psychology, 1415 Jayhawk Blvd., University of Kansas, Lawrence, KS 66045 (e-mail: mvitevit@ku.edu).

Note—This article was accepted by the previous editorial team, when Colin M. MacLeod was Editor.

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 & Language*, **38**, 419-439.

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

Auer, E. T., Jr., & Luce, P. A. (2005). Probabilistic phonotactics in spoken word recognition. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 610-630). Oxford: Blackwell.

Bölte, J., & Uhe, M. (2004). When is all understood and done? The psychological reality of the recognition point. *Brain & Language*, **88**, 133-147.

Bond, Z. S. (1999). *Slips of the ear: Errors in the perception of casual conversation*. San Diego: Academic Press.

Cohen, J. (1976). Random means random. *Journal of Verbal Learning & Verbal Behavior*, **15**, 261-262.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, **25**, 257-271.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-556). New York: Academic Press.

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, **37**, 65-70.

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

Grosjean, F. (1996). Gating. *Language & Cognitive Processes*, **11**, 597-604.

Hino, Y., & Lupker, S. J. (2000). Effects of word frequency and spelling-to-sound regularity in naming with and without preceding lexical decision. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 166-183.

Johnson, N. F., & Pugh, K. R. (1994). A cohort model of visual word recognition. *Cognitive Psychology*, **26**, 240-346.

Keppel, G. (1976). Words as random variables. *Journal of Verbal Learning & Verbal Behavior*, **15**, 263-265.

Kučera, 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 & Verbal Behavior*, **12**, 119-131.

Lewandowsky, S. (1993). The rewards and hazards of computer simulations. *Psychological Science*, **4**, 236-243.

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

Luce, P. A., Goldinger, S. D., Auer, E. T., Jr., & 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 & Hearing*, **19**, 1-36.

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

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

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 & Brain Sciences*, **23**, 299-370.

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: Indiana University, Psychology Department, Speech Research Laboratory.

Pugh, K. R., Rexer, K., Peter, M., & Katz, L. (1994). Neighborhood effects in visual word recognition: Effects of letter delay and nonword context difficulty. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 639-648.

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 & Language*, **41**, 416-426.

Radeau, M., & Morais, J. (1990). The uniqueness point effect in the shadowing of spoken words. *Speech Communication*, **9**, 155-164.

Radeau, M., Morais, J., Mousty, P., & Bertelson, P. (2000). The effect of speaking rate on the role of the uniqueness point in spoken word recognition. *Journal of Memory & Language*, **42**, 406-422.

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

Storkel, H. L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language*, **29**, 251-274.

Storkel, H. L. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, **25**, 201-221.

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

Vitevitch, M. S. (1998). *All neighborhoods are not created equal: The phonological P-metric and spoken word recognition* (Research on Speech Perception, Progress Report No. 22). Bloomington: Indiana University, Psychology Department, Speech Research Laboratory.

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

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

Vitevitch, M. S. (2002c). 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 perception of spoken words. *Psychological Science*, **9**, 325-329.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory & 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, & Computers*, **36**, 481-487.

Vitevitch, M. S., & Luce, P. A. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory & Language*, **52**, 193-204.

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.
Wike, E. L., & Church, J. D. (1976). Comments on Clark's "The language-as-fixed-effect fallacy." *Journal of Verbal Learning & Verbal Behavior*, **15**, 249-255.

NOTES

1. These results replicate the findings of an auditory naming task described in Vitevitch (1998) with a set of stimuli that were also manipulated in terms of neighborhood spread, but which were not as well controlled as the present stimuli.

2. For the stimulus words in the present set of experiments, the correlation between *P* and uniqueness point was not significant [$r = .15$, $Z(92) = 1.4$, $p = .17$]. Furthermore, $r^2 = .02$, meaning that 2% of the variability in *P* was accounted for by the uniqueness point.

(*Continued on next page*)

APPENDIX A
Stimulus Items Used in Experiments 1–3, and Examples of Their Neighbors

Phonological $P = 2$				Phonological $P = 3$			
	Neighbors				Neighbors		
Stimulus	P1	P2	P3	Stimulus	P1	P2	P3
chalk	hawk	check	*	cheese	tease	choose	cheap
chill	fill	*	chip	chess	guess	chase	check
church	search	*	chirp	chose	rose	cheese	chore
deaf	chef	*	dead	curb	verb	cub	curl
dodge	lodge	*	dot	dish	wish	dash	dip
doll	*	dull	dock	dog	log	dug	dawn
dose	*	dice	dove	doubt	shout	dirt	down
fetch	retch	*	fed	dove	cove	dive	dome
fish	wish	*	fib	firm	term	fame	fern
five	dive	*	fine	foam	home	firm	phone
foul	howl	feel	*	fog	hog	fig	fall
gab	cab	*	gag	foot	soot	fight	full
geese	peace	gas	*	gauze	pause	gaze	gone
good	wood	guide	*	germ	term	gem	jerk
gouge	gauge	*	gown	gown	down	gun	gouge
hedge	wedge	*	head	guide	wide	god	guise
hen	den	*	hem	guise	size	gauze	guide
jade	wade	*	jail	hive	dive	heave	hike
joke	poke	jerk	*	hog	dog	hug	haul
judge	fudge	*	jug	jab	cab	job	jack
king	ring	*	kick	jerk	work	joke	germ
league	*	log	lease	ledge	hedge	lodge	leg
leash	*	lash	leap	lobe	robe	lob	load
loaf	*	life	lobe	lodge	dodge	ledge	lock
lull	hull	*	lush	lurch	church	leach	learn
mesh	*	mash	met	mop	hop	map	mock
mob	lob	*	mock	mouse	house	mace	mouth
moth	*	mouth	moss	mouth	south	moth	mouse
noise	poise	nose	*	neck	wreck	knock	net
noun	down	nun	*	niece	piece	nurse	need
nudge	fudge	*	nut	nurse	purse	noose	nerve
palm	psalm	*	pop	pause	cause	poise	pawn
path	math	*	pad	peg	beg	pig	pen
poise	noise	pause	*	pouch	vouch	pitch	pout
sash	rash	*	sack	sauce	toss	cease	sought
shawl	wall	shell	*	shave	wave	shove	shape
sheath	wreath	*	sheep	shop	top	ship	shot
shine	dine	shun	*	shove	love	shave	shun
thought	fought	*	thong	theme	beam	thumb	thief
tube	*	tub	tune	toad	road	tide	tote
vague	*	vogue	vase	van	man	vein	vat
verb	curb	*	verse	verse	terse	voice	verb
verge	surge	*	verb	vote	boat	vet	vogue
wing	sing	*	wish	weave	leave	wove	weep
womb	tomb	worm	*	wedge	hedge	wage	well
worse	curse	*	worth	worth	birth	with	word

Note—P1, change at first phoneme position; P2, change at second phoneme position; P3, change at third phoneme position. *No English word in the corpus (see Nusbaum, Pisoni, & Davis, 1984) is formed by changing the stimulus phoneme at this position.

APPENDIX B
Nonwords (Transcribed in IPA) Used in Experiment 1

bæf	dɛʒ	læɵ	pin
bæv	daɪt	lɑd	pɪp
bæb	dʌp	liɵ	pɪʃ
bæz	fod	luɵ	pɪv
bæp	faɪm	laɪɵ	pob
bɔn	hæb	meɵ	pod
bɔp	hæð	meg	pot
bef	hɛb	mep	pʌv
bɛf	hɛk	mig	ræb
bɛʤ	hɪf	mʌp	rɑɵ
bɛp	hɪb	nɑp	rʌp
bɛv	hɪʒ	nɪs	rʌz
bib	hɪʤ	naɪp	sæz
biɵ	hɪʃ	pæb	sɛk
big	kæk	pæg	sib
bɪk	kɑʤ	pæv	ʃɪd
bɪɵ	ked	peg	siv
bog	keb	pɛp	sʌt
bʌp	kɪf	peɵ	tæt
bʌɵ	kɪʤ	pid	teʤ
dɑz	kɪɵ	piʤ	tev
deb	kɪz	pɪf	tɪʃ
dɛʤ	kof	pig	taɪv

APPENDIX C
"Different" Stimulus Pairs Used in Experiment 3

bad/badge	fool/food	match/mass	safe/save
bake/base	fuzz/fun	maze/main	sane/same
batch/bat	game/gaze	met/mess	sang/sake
beam/beach	gate/gain	moan/mole	sat/sag
beige/bait	gum/gun	mood/moon	scene/seal
bell/bed	hack/hash	mug/mud	shape/share
birch/bird	head/hem	net/nerve	shock/shot
bowl/boil	heard/heap	note/nose	soup/suit
cat/can	hole/hope	patch/pack	tag/tack
chip/chin	hot/hop	peach/peel	talk/taught
code/comb	hum/hut	perch/perk	tall/toss
coil/coin	hype/height	pipe/pike	term/terse
core/cone	kick/kin	pub/puff	tide/tight
cove/coat	knife/nice	pun/puck	toll/tone
curve/curl	knit/nick	rage/race	ton/tough
date/dame	leaf/leak	rash/rat	tour/took
dial/dire	lean/leap	reef/reek	tub/tuck
dill/dim	lease/leave	ride/ripe	weak/weep
duck/dug	less/leg	rip/riff	well/web
dull/done	life/light	roach/road	wine/wipe
fame/fake	load/loan	roam/rope	wing/whip
feet/feel	make/mate	run/rug	wise/wife
fig/fin	man/map	sack/sad	yell/yawn

(Manuscript received September 6, 2004;
revision accepted for publication November 7, 2005.)