Influence of Onset Density on Spoken-Word Recognition

Michael S. Vitevitch
Indiana University Bloomington

Previous research has suggested that the initial portion of a word activates similar sounding words that compete for recognition. Other research has shown that the number of similar sounding words that are activated influences the speed and accuracy of recognition. Words with few neighbors are processed more quickly and accurately than words with many neighbors. The influences of the number of lexical competitors in the initial part of the word were examined in a shadowing and a lexical-decision task. Target words with few neighbors that share the initial phoneme were responded to more quickly than target words with many neighbors that share the initial phoneme. The implications of onset-density effects for models of spoken-word recognition are discussed.

The process of spoken-word recognition involves the discrimination of a single candidate from the many possible lexical alternatives activated in memory by the acoustic–phonetic input (e.g., Forster, 1979; Luce, Pisoni, & Goldinger, 1990; Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris, 1994). In an early and influential theory of spoken-word recognition, Marslen-Wilson and Welsh (1978) proposed that the similar competitors activated in memory—a group of words often called the *cohort*—consist of words that share the initial portion of the incoming acoustic–phonetic input. For example, on receiving the input /hʌ/ the words *hull, Hun, hundred, hump, hunt, hustle, huff,* and *hut,* among others, compete with each other for recognition. As additional acoustic–phonetic input is received, such as /hʌn/, competitors that are not consistent with the input—in this case *hull, hump, hustle, huff,* and *hut*—drop out of the cohort and no longer compete with the words that are consistent with the acoustic-phonetic input. Competitors that mismatch the incoming acoustic-phonetic input continue to drop out of the cohort until a single item—the word to be recognized—remains.

Evidence from Marslen-Wilson and Zwitserlood (1989) supported the special status of the initial portion of a word in lexical access (cf. Connine, Blasko, & Titone, 1993). Marslen-Wilson and Zwitserlood used a cross-modal priming task in which participants heard a word over a set of headphones and made a lexical decision on a visual target presented at the offset of the spoken stimulus.

Reaction times to the visually presented word were compared across several different prime conditions varying in lexicality and the amount of phonological overlap. For example, the Dutch word *honing* (honey) is semantically related to the visually presented word *bij* (bee). In this condition, the *original-word condition,* the semantic relationship between the auditorily and visually presented words facilitated the lexical-decision response. The degree of priming when presented with the word *honing* (original-word condition) was compared with the degree of priming when presented with the real-word rhyme prime *woning* (dwelling), the real-word control prime *pakket,* the nonword rhyme prime *foning,* and the nonword control prime *dakket.*

Marslen-Wilson and Zwitserlood (1989) reasoned that if the overall goodness of phonological match rather than the initial portion of the word determined which words entered into the cohort, then the rhyme primes (both word and nonword) should facilitate the response to the visual target as much as the original word. That is, *woning* (real-word rhyme prime) and *foning* (nonword rhyme prime) share all but the initial phoneme with the word *honing* and should facilitate the response to *bij* just as presentation of the original word *honing* should facilitate the response to *bij.* However, they found that both word and nonword rhyme primes did not facilitate lexical decision as much as the original word primes did, suggesting that the initial portion of a word determines entry into the cohort and may, therefore, have special status in spoken-word recognition.

Cole and Jakimik (1980; see also Cole, 1973) also demonstrated that the initial portion of words may have special status in spoken-word recognition by using a speeded mispronunciation detection task. Participants detected mispronunciations in the second syllable of a word more quickly than they detected mispronunciations in the first syllable of the word. Cole and Jakimik (1980) argued that when the mispronunciation occurred in the second syllable of the word, the listener used the correct information in the first syllable to recognize the intended word and detect the mispronunciation in the second syllable. In contrast, when the mispronunciation was in the initial portion of the word, the listener activated a set of competitors that was consistent with the mispronounced syllable and inconsistent with the intended pronunciation, resulting in the listener requiring additional input to recognize the intended input and detect the word-initial mispronunciation.

Linguistic evidence also suggests that the initial portion of a word may have a special status in the processing of spoken words. Treiman and colleagues have found that errors in short-term memory for spoken syllables (e.g., Treiman & Danis, 1988) and phoneme sequences formed in word games (e.g., Treiman, 1983, 1986; see also MacKay, 1972) are affected by the linguistic structure of the syllable. The results from these tasks are consistent with a linguistic perspective that syllables are coded in terms of a linguistic onset and a linguistic rime, a term often used to refer to the final part of a word (Fudge, 1969). In short, the initial portion of a word or syllable is "psychologically important" (Treiman & Danis, 1988, p. 147) to various language-related processes.

The psychological importance of the initial portion of a word has been observed in children as well as adults. Walley (1987) demonstrated that the language processing abilities of young children are influenced by the information in the initial portion of a word. More specifically, Walley found that 4- and 5-year-old children were more accurate at detecting mispronunciations in one-, two-, and three-syllable words embedded in a sentence when the mispronunciation occurred in the word-initial rather than the word-final position. Walley and Metsala (1990) observed similar effects for word-initial and word-final mispronunciations in children aged 5 and 8 years old. On the surface, the results of Walley and Metsala (1990) and Walley (1987) seem to contradict the findings of Cole and Jakimik (1980; see also Cole, 1973) described earlier and prove problematic for the claim that the initial portion of a word is psychologically important. Note, however, that Walley and colleagues used accuracy rates as their dependent measure, whereas Cole and colleagues used response latency as their dependent measure. Listeners may detect a mispronunciation in the word-initial position more accurately than they detect a mispronunciation in later word positions, but they may do so at the expense of response speed (i.e., a speed–accuracy trade-off). Thus, these findings are not incompatible with the hypothesis that the initial portion of a word is psychologically important.

Sensitivity to the initial portion of a word may develop as early as 9 months of age. Jusczyk, Goodman, and Baumann (1999) found that 9-month-old infants listened longer to lists of CVC (C = consonant, V = vowel) syllables when the items in the list shared the initial CVs, the initial Cs, or the same manner of articulation as the syllable onset than when they listened to lists of CVC syllables that were similar at the end. They argued that infants first develop sensitivity to similarities among words that occur in the initial portion rather than the final portion of syllables and words. These results further emphasize the psychological importance of the initial portion of a word.

The claim by Marslen-Wilson and Welsh (1978) regarding the special status of the initial portion of a word has received support from a number of different sources and tasks (e.g., Jusczyk et al., 1999; Marslen-Wilson & Zwitserlood, 1989; Treiman & Danis, 1988; Walley & Metsala, 1990; cf. Connine et al., 1993). Marslen-Wilson and Welsh claimed that the initial portion of the word activates a set of competitors that are gradually winnowed down to the single word that is to be recognized. Marslen-Wilson (1987) further claimed that this winnowing down of competitors in the cohort occurs in parallel. That is, the number of competitors in the cohort does not affect the speed and accuracy of spoken-word recognition. Or, in the words of Marslen-Wilson (1987), "the timing of word-recognition processes is not affected by the number

of alternatives that need to be considered" (p. 84). However, Luce and Pisoni (1998) and colleagues (e.g., Luce et al., 1990; Vitevitch & Luce, 1998) have demonstrated with a variety of tasks that the number of competitors activated in memory does affect the speed and accuracy of spoken-word recognition. Specifically, a word that activates few competitors (a word with a sparse neighborhood) is recognized more quickly and accurately than a word that activates many competitors (a word with a dense neighborhood).

The influence of neighborhood density on spoken-word recognition has been observed in a variety of tasks and participant populations. In a perceptual-identification task, in which participants hear a word mixed with noise and must type out the word they believe they heard, Luce and Pisoni (1998) demonstrated that words with sparse neighborhoods were identified more accurately than words with dense neighborhoods. Measures of online processing such as auditory naming (Luce et al., 1990; Luce & Pisoni, 1998; Vitevitch & Luce, 1998), speeded same–different decision (Vitevitch & Luce, 1999), and auditory lexical decision (Luce & Pisoni, 1998; Vitevitch & Luce, 1999) have also demonstrated that neighborhood density affects the speed with which spoken words are processed. In each of these tasks, it was observed that words with sparse neighborhoods were responded to more quickly than words with dense neighborhoods. The effects of neighborhood density on spoken-word recognition have been observed when monosyllabic (Luce & Pisoni, 1998) as well as bisyllabic words (Charles-Luce, Luce, & Cluff, 1990; Cluff & Luce, 1990) are used as stimuli in these tasks.

Neighborhood-density effects have also been observed in a variety of participant populations. Jusczyk, Luce, and Charles-Luce (1994) found that 9-month-old infants displayed significant listening preferences for nonwords that had high rather than low phonotactic probabilities. Phonotactic probability refers to the frequency with which segments and sequences of segments occur together in a word and is positively correlated with neighborhood density (Vitevitch, Luce, Pisoni, & Auer, 1999). Words with high phonotactic probability tend to have dense neighborhoods, whereas words with low phonotactic probability tend to have sparse neighborhoods. Thus, sensitivity to properties correlated with neighborhood density have been observed in 9-month-old infants. Sommers (1996; Sommers & Danielson, 1999) has demonstrated effects of neighborhood density among normal-hearing elderly adults. Finally, Kirk, Pisoni, and Miyamoto (1997) have demonstrated effects of neighborhood density in hearing-impaired adults who use cochlear implants.

Given the importance of the initial part of a word and the importance of the number of lexical competitors, the present set of experiments attempted to examine simultaneously the influence of both of these variables on the processing of spoken words. That is, the reported experiments tested whether varying the number of competitors activated by the initial part of the word affected spoken-word recognition.

Experiment 1: Auditory Shadowing

In many studies examining the influence of neighborhood density on lexical processing, a simple computational metric is used to assess the number of lexical competitors (e.g., Kirk et al., 1997; Sommers, 1996; Vitevitch & Luce, 1998, 1999). Note that Luce and Pisoni (1998) have discussed an alternative method of esti-

mating neighborhood density on the basis of perceptual estimates of similarity. Although the two methods of estimating neighborhood density are different, the outcomes of using both methods of estimation are equivalent. Using the computational metric, neighborhood density is estimated by counting the number of words formed by the substitution, addition, or deletion of a single phoneme into any position of a target word. If a real word is formed, that new word is considered a neighbor of the target word. For example, in the target word /sæd/ (sad), the substitution of a single phoneme will form the neighbors /bæd/ (bad), /sid/ (seed), and /sæk/ (sack). A word that has many neighbors formed by the substitution, addition, or deletion of a single phoneme is said to have a dense neighborhood, whereas a word that has few neighbors formed in this manner is said to have a sparse neighborhood.

In examining words with equivalent numbers of neighbors, one might notice that the neighbors formed by the computational metric are not equally distributed among the possible phoneme positions in the word. That is, if a three-phoneme word has nine neighbors, it may not necessarily be the case that three neighbors are formed by a substitution in the initial position, three neighbors are formed by a substitution in the medial position, and three neighbors are formed by a substitution in the final position. Rather, seven neighbors may be formed by a substitution in the initial position, and one neighbor formed by a substitution in each of the medial and final positions. Alternately, one neighbor may be formed by a substitution in the initial position, and four neighbors may be formed by a substitution in each of the medial and final positions. The present set of experiments attempted to determine whether this unequal distribution of neighbors in the initial phoneme position among words with equivalent-sized neighborhoods has any consequences on lexical processing. The work demonstrating the psychological importance of the initial part of a word (e.g., Jusczyk et al., 1999; Marslen-Wilson & Zwitserlood, 1989; Treiman & Danis, 1988; Walley & Metsala, 1990) has suggested that some effect on processing should be observed.

To further illustrate the difference in the distribution of neighbors in the initial position of a word, consider the words /mæs/ (mass) and /sæd/ (sad). The word /mæs/ (mass) has as neighbors—on the basis of the computational metric—the words /mɪs/ (miss), /mæd/ (mad), /mæn/ (man), and /pæs/ (pass). The word /sæd/ (sad) has as neighbors the words /bæd/ (bad), /fæd/ (fad), /læd/ (lad), and /sæk/ (sack). The words /mæs/ and /sæd/ have additional words as neighbors; however, for presentation purposes only these four neighbors have been listed. Note that three of the four neighbors of the word /mæs/ have the same initial phoneme as the target word /mæs/ (the phoneme /m/), whereas one of the four neighbors of the word /sæd/ has the same initial phoneme as the target word /sæd/ (the phoneme /s/). Despite an equal number of neighbors, the word /mæs/ has a greater proportion of neighbors that have the same initial phoneme as the target word (75%), whereas the word /sæd/ has a smaller proportion of neighbors that share the same initial phoneme as the target word (25%). In the present article I have used the term onset density to refer to the proportion of neighbors that share the same initial phoneme as the target word. Words with a high proportion of neighbors sharing the onset of the target word are referred to as having a dense onset, whereas words with a low proportion of neighbors sharing the onset of the target word are referred to as having a sparse onset. I selected the term onset density simply for ease of

exposition. It is not meant to imply or intimate the acceptance of any specific representational unit or theoretical construct (linguistic or otherwise).

In the present experiment, words with sparse and dense onsets were presented in an auditory shadowing task. Participants were instructed to repeat as quickly and as accurately as possible the word they heard presented (in the clear) over a set of headphones. Given that the two groups of words (dense vs. sparse onset) were equivalent in the number of overall neighbors (as well as a number of other variables), one would predict that there should be no difference in the accuracy or speed of responses to the two groups of words if only the overall number of lexical competitors or neighbors affects processing. In contrast, if the initial portion of a word does have some psychological importance, there should be a difference in the speed and accuracy of responses to words varying in onset density. Given the competitive influences of overall neighborhood density (e.g., Luce & Pisoni, 1998), I hypothesized that participants would respond to words with sparse onsets more quickly and accurately than to words with dense onsets.

Method

Participants. Eighteen native English speakers from the Indiana University pool of introductory psychology students participated in partial fulfillment of a course requirement. None of the participants reported a history of speech or hearing problems, and all were right-handed.

Stimuli. Ninety CVC words were used in the experiment and are listed in the Appendix. The stimuli were divided into two sets of 45 words each. In one set of stimuli, for each target word, the proportion of neighbors that shared the initial phoneme of that word was greater than 50%. This set, referred to as the *dense-onset condition*, had a mean proportion of 75.3% of the neighbors with the same initial phoneme as the target word. In the other set of stimuli, for each target word, the proportion of neighbors that shared the initial phoneme of that word was less than 50%. This set, referred to as the *sparse-onset condition*, had a mean proportion of 42.0% of the neighbors sharing the initial phoneme of the target word. The difference in the proportion of neighbors sharing the initial phoneme of the target word between the sparse- and dense-onset conditions was significantly different, $F(1, 88) = 704.17, p < .01$.

Both sets of words were equivalent in word familiarity, $F(1, 88) < 1$, as measured by a 7-point scale (Nusbaum, Pisoni, & Davis, 1984). Words in the dense-onset condition had a mean familiarity of 6.93. Words in the sparse-onset condition had a mean familiarity of 6.91, indicating that all the stimuli used were words highly familiar to native speakers of English.

The two sets of words were also equivalent in word frequency, $F(1, 88) < 1$, as measured by log transformations of Kučera and Francis's (1967) word counts. Words in the dense-onset condition had a mean log-frequency of 1.35. Words in the sparse-onset condition had a mean log-frequency of 1.23.

The number of neighbors (calculated using the computational metric) did not differ significantly between the two sets of words, $F(1, 88) = 3.78$, $p = .06$. Words in the dense-onset condition had a mean neighborhood density of 20.0 words. Words in the sparse-onset condition had a mean neighborhood density of 22.1 words. Note that the difference in overall neighborhood density was not statistically different at the traditional p value of .05 and that the slight difference in overall neighborhood density was in the opposite direction of the differences in onset density.

The two sets of stimuli did not differ in *neighborhood frequency*—the mean frequency of occurrence for the neighbors—as measured by log transformations of Kučera and Francis's (1967) word counts, $F(1, 88) < 1$. The mean neighborhood log-frequency for words in the dense-onset condition was 1.03. The mean neighborhood log-frequency for words in the sparse-onset condition was 1.05.

Recognition points, that is, the point in a word at which it diverges and becomes unique from all other words in the lexicon, were computed from a lexical database (approximately 20,000 words from *Webster's Pocket Dictionary*; see Luce & Pisoni, 1998, and Luce, 1986). This analysis showed that the recognition points did not differ between the two sets of words, $F(1, 88) < 1$. The mean recognition point for words in the dense-onset condition was 2.84 phonemes, and the mean recognition point for words in the sparse-onset condition was 2.78 phonemes.

The *phonological-P* (Vitevitch, 1998)—the number of phoneme positions that form at least one neighbor when a single phoneme substitution is made—did not differ between the two sets of stimuli, $F(1, 88) < 1$. Words in the dense-onset condition had a mean phonological-P of 2.91 phoneme positions forming neighbors. Words in the sparse-onset condition had a mean phonological-P of 2.87 phoneme positions forming neighbors. The phonological-P metric is analogous to "spread," or P, as described by Johnson and Pugh (1994). Johnson and Pugh defined P as the number of letter positions in a word that formed at least one neighbor after a single-letter substitution. To illustrate, the word *dog* has as neighbors the words *fog, bog, log, hog, cog, dig, dug,* and *dot*. Note that at least one neighbor is formed when a letter is substituted into the initial, medial, or final letter position of the word *dog*, giving it a P count of 3. A word like *kin* would have a P of only 2 because only two letter positions (the initial and final letter positions) form real words (*tin, win, pin, fin, sin, din, bin, kid*) when a single letter is substituted. The phonological-P metric is similar to the P metric described by Johnson and Pugh except that phonemes are substituted into phonological representations of words instead of letters being substituted into orthographic representations of words (Vitevitch, 1998).

Finally, equal numbers of words in each set contained the following phonemes in the initial position: /d, f, k, l, m, p, r, s, t, w/. Controlling the phonemes in the initial position of the words in each set ensured that possible differences in reaction time between the two sets were due to the manipulation of onset density and not to differences in the phonemes used in the initial position between the two sets of words. Furthermore, when using a voice key to measure response times in a shadowing task, it is important that the initial segments of the words found in each condition are comparable in their acoustic–phonetic properties so that the phonemes in one set do not differentially activate the voice key. By using the same initial phonemes in equal numbers of words in each condition, I was able to control these differences.

The words were spoken in isolation and recorded by a trained speech scientist in an Industrial Acoustics Company sound-attenuated booth with a high-quality microphone. The stimuli were low-pass filtered at 10.4 kHz and digitized at a sampling rate of 20 kHz with a 16-bit analog-to-digital converter. All words were edited into individual digital files, leveled at 70-db SPL, and stored on computer disk for later playback. Stimulus durations for both groups were equivalent, $F(1, 88) < 1$. The mean duration was 924 ms for words in the dense-onset condition and 925 ms for words in the sparse-onset condition.

Procedure. Participants were tested individually. Each participant was seated in a booth equipped with a computer terminal, a pair of Beyerdynamic DT-100 headphones, and a Shure Unidyne III dynamic microphone (Model 545) interfaced with a voice-activated response key. Presentation of stimuli and response collection was controlled by a 200-MHz Gateway 2000 Pentium computer.

A trial proceeded as follows: A prompt (the word *READY*) appeared on the computer screen for 500 ms. This was followed by one of the stimulus items randomly presented at 70-db SPL over the headphones. The participant then repeated the item as quickly and as accurately as possible into the microphone.

Reaction times were measured from the onset of the stimulus to the onset of the participant's vocal response. All responses were recorded on audiotape for accuracy analysis. Accuracy was assessed by listening to the participants' responses and comparing them with a written transcription of the words. A stimulus was scored as correct if there was an identical match

on all segments of the word. Prior to the experimental trials, each participant received 10 practice trials to become familiar with the task. The practice trials were not included in the final data analysis.

Results

Only correct naming responses were included in the analysis of the response latencies. Repeated measures analyses of variance (ANOVAs) were used to separately analyze response times and accuracy rates with participants as a random factor (F_1). Independent ANOVAs were also used to analyze response times and accuracy rates with the items as a random factor (F_2). Please note that the stringent selection process for the words used as stimuli in the present experiment—initial phoneme, syllable structure, familiarity, frequency, neighborhood density, neighborhood frequency, phonological-P, and recognition points were all stringently controlled in these words—made analyses with participants as the random factor the only appropriate statistic to use in this set of experiments (see Cohen, 1976; Hino & Lupker, 2000; Keppel, 1976; Raaijmakers, Schrijnemakers, & Gremmen, 1999; Smith, 1976; Wike & Church, 1976). That is, the items used in this experiment were not randomly selected, so the use of them as a random factor in an ANOVA would have been inappropriate. Treating nonrandomly selected stimuli as a random factor results in an increased probability of making a Type II error. However, the convention in psycholinguistic research dictates that analyses with participants and items as random factors be reported, so I have followed this practice in the analyses reported here.

The results from the shadowing task are presented in Table 1. A significant difference in the response latencies was found between the two conditions, $F_1(1, 17) = 7.24, p < .05; F_2(1, 88) = 4.44, p < .05$. Words in the sparse-onset condition were repeated more quickly $(M = 1,010$ ms$)$ than words in the dense-onset condition $(M = 1,021$ ms$)$. A difference of 11 ms between the means may appear to be a very small difference and may be construed as a small effect. However, a proper estimate of effect size such as *PV*, the proportion of variance in the dependent variable explained by the general linear model (i.e., the model that underlies analysis of variance), shows that the result of this experiment has a large effect size (Murphy & Myors, 1998). PV = .2986 for this experiment (calculated from Equation 7 of Murphy & Myors, 1998). This effect is comparable to the effect size (PV = .3167) obtained in Experiment 3 of Luce and Pisoni (1998), which used the same task

Table 1

Mean Reaction Times and Accuracy Rates for Words That Vary in Onset Density

Onset-density condition	Reaction time (ms)		Accuracy rate (%)	
	M	SD	M	SD
Experiment 1: Auditory shadowing				
Dense	1,021	76.4	94.8	0.04
Sparse	1,010	77.4	94.9	0.05
Experiment 2: Lexical decision				
Dense	1,029	97.2	95.7	0.03
Sparse	1,006	87.9	95.3	0.02

used in the present experiment to examine the influence of neighborhood density on spoken-word recognition.

Given that many cognitive psychologists are attempting to dissect the microstructure of cognitive processes, it should not be surprising that 11 ms can have important implications on the outcome of a cognitive process. Indeed, one might be more inclined to express concern when extremely large differences between means are obtained in tasks that examine the rapid and efficient processes characteristic of intact cognitive processes. No differences were found for the accuracy rates in the shadowing task (both $Fs < 1$), suggesting that participants did not adopt a response strategy trading speed of response for accuracy of response.

Although participants did not adopt a strategy of trading off speed for accuracy, one might ask whether differences in the initial portion of the word or in the final part of the word (the rime) were responsible for the observed difference in reaction time. To assess the influence of onset density and variability in the rime, I used a stepwise regression analysis to identify which variable significantly decreased the variance of the regression equation. The percentage of neighbors that shared the initial phoneme of each target word was used to measure onset density. Because neighborhood density was assessed with a single-phoneme substitution metric, there was an inverse relationship between onset density and what might be called *rime density*. That is, the proportion of neighbors that shared the initial phoneme (and had a different word ending than the target) was equal to one minus the proportion of neighbors that did not share the initial phoneme (and had the same word ending as the target). Thus, variability in the rime had to be assessed in another way.

Variability in the rime was measured by calculating the probability of the vowel and final consonant co-occurring in the target word. Recall that overall neighborhood density was equated between the two groups of words that varied in onset density. That is, the total number of neighbors was the same, but the number of neighbors sharing the initial phoneme of the target word varied. To maintain equivalent overall neighborhood size, I had to vary the rime of the word (the VC portion of the words in this experiment). Words with a dense onset had many neighbors that shared the same initial phoneme as the target. To keep the overall number of neighbors equal, there had to be many other rimes among the neighbors. That is, the rime of these target words had to have a low probability of occurring. In contrast, words with a sparse onset had few neighbors that shared the same initial phoneme as the target. To maintain the overall size of the neighborhood, there had to be many different initial phonemes among the neighbors, but the same rime was shared by many neighbors. That is, the rime of these words had a high probability of occurring. By using the probability of the vowel and the final consonant co-occurring together in the target word, the possible influence that the rime may have had on the results of this experiment could be assessed.

A stepwise regression analysis performed using VC co-occurrence probability and the percentage of onset density for each word as predictors of reaction time showed that onset density ($R = .219$) was the only variable that significantly predicted reaction time, $F(1, 88) = 4.42, p < .05$. Co-occurrence probability of the vowel and final consonant (VC) did not significantly predict reaction time ($R = -.157$) or significantly reduce the variance in the regression equation after the entry of onset density into the

regression equation. The results of this regression analysis suggest that variability in the rime did not significantly contribute to the results that were obtained. Rather, differences in the number of neighbors that shared the initial phoneme of the target word (i.e., onset density) significantly affected the speed of the responses in the auditory naming task.

Discussion

The results of Experiment 1 show that words with few neighbors that share the initial phoneme with the target word (i.e., sparse onset) are repeated more quickly than words with many neighbors that share the initial phoneme with the target word (i.e., dense onset). That is, the initial segment of a word may activate a varying number of potential lexical candidates in memory. These results are consistent with previous studies that have demonstrated the psychological importance of the initial portion of a word (e.g., Cole & Jakimik, 1980; Jusczyk et al., 1999; Marslen-Wilson & Zwitserlood, 1989; Treiman & Danis, 1988; Walley, 1987). Although the work of Marslen-Wilson and Zwitserlood (1989), for example, suggested that the initial portion of a word is important because it activates lexical candidates in memory, this work did not address the possible effects of different sizes of candidate sets, or cohorts, on spoken-word recognition.

The results of the present experiment are also consistent with studies that have shown that the number of lexical competitors activated in memory influences the recognition of spoken words (e.g., Charles-Luce et al., 1990; Cluff & Luce, 1990; Kirk et al., 1997; Luce et al., 1990; Luce & Pisoni, 1998; Sommers & Danielson, 1999; Vitevitch & Luce, 1998, 1999). The findings from these experiments and from the present experiment contrast with the claim by Marslen-Wilson (1987) that "the timing of word-recognition processes is not affected by the number of alternatives that need to be considered" (p. 84). Furthermore, the results of the present experiment suggest that it is not just the overall number of candidates activated in memory that affects lexical processing (e.g., Luce & Pisoni, 1998). Rather, the initial portion of a word may play an important role in determining the candidate set. In short, the results of the present experiment demonstrate the importance of the number of lexical competitors activated by the initial portion of a word.

Experiment 2: Auditory Lexical Decision

The present experiment used a speeded auditory lexical-decision task to further demonstrate that the number of neighbors that share the onset of the target word affects the online processing of spoken words. The auditory lexical-decision task was chosen because it—like the auditory naming task used in Experiment 1—uses reaction time as a dependent measure. If the dependent measure were switched to accuracy rates by using a task such as perceptual identification, possible confusion could arise in interpreting the results; recall the apparent contradiction between the findings of Walley (1987; see also Walley & Metsala, 1990) and Cole (1973; see also Cole & Jakimik, 1980). By using another task that measures reaction time, I was able to more closely replicate the findings of Experiment 1 and therefore place the influence of onset density on spoken-word recognition on firmer empirical grounds.

Method

Participants. Eighteen individuals from the same population sampled in Experiment 1 participated in this experiment. None of the individuals who took part in the present experiment participated in Experiment 1.

Stimuli. The same two sets of stimuli that were used in Experiment 1 were also used in this experiment. In addition, the last phoneme of 90 CVC words not found in the stimulus set was changed to form monosyllabic nonwords. The nonwords were recorded and treated in the same manner as the stimuli in Experiment 1.

Procedure. Participants were tested in groups of 4 or less. Each participant was seated in a booth equipped with a computer terminal, a pair of Beyerdynamic DT-100 headphones, and a two-button response box interfaced to a dedicated timing board in the computer. The left-hand button on the response box was labeled NONWORD, and the right hand button (i.e., the button pressed by the dominant hand of the participants) was labeled WORD. A 200-MHz Gateway 2000 Pentium computer controlled the presentation of stimuli and the collection of responses.

A trial proceeded as follows: A prompt (the word READY) appeared on the computer screen for 500 ms, and then one of the stimulus items was randomly presented at 70-db SPL over the headphones. The participant responded as quickly and as accurately as possible by pushing the appropriately labeled button. Reaction time was measured from the onset of the stimulus to the onset of the participant's 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

Only correct responses were included in the analysis of response latencies. Repeated measures ANOVAs were again used to separately analyze response times and accuracy rates with participants as a random factor (F_1). To maintain the conventions of psycholinguistic research, I again used independent ANOVAS to analyze response times and accuracy rates with the items as a random factor (F_2).

A significant difference in the response latencies was found between the two conditions, $F_1(1, 17) = 9.71$, $p < .01$; $F_2(1, 88) = 4.76$, $p < .05$. Words with sparse onsets were responded to more quickly ($M = 1,007$ ms) than words with dense onsets ($M = 1,029$ ms). These results are also displayed in Table 1. No differences were found for the accuracy scores in the lexical-decision task (both $Fs < 1$), suggesting that there was no strategic trade off between speed and accuracy in responding. As in Experiment 1, the results of the current experiment show that participants responded more quickly to words with sparse onsets than to words with dense onsets, further suggesting that onset density affects spoken-word recognition.

An estimate of effect size for the reaction-time result in the present experiment showed PV = .3635 (calculated from Equation 7 of Murphy & Myors, 1998). As in Experiment 1, the size of this effect is considered to be large (Murphy & Myors, 1998) and comparable to that obtained in other studies of neighborhood density (e.g., Experiment 3 of Luce & Pisoni, 1998).

A stepwise regression was again used to demonstrate that the observed effects were due to differences in onset density rather than differences in rime variability. As in Experiment 1, the percentage of neighbors that shared the initial phoneme of each target word was used to measure onset density, and the probability of VC co-occurrence was used to measure rime variability. The results of this stepwise regression analysis again showed that onset density ($R = .225$) was the only variable that significantly predicted reaction time, $F(1, 88) = 4.69$, $p < .05$. Co-occurrence probability of the vowel and final consonant (VC) did not significantly predict reaction time ($R = .09$) or significantly reduce the variance in the regression equation after the entry of onset density into the regression equation. As in Experiment 1, the results of this regression analysis suggest that variability in the rime did not significantly contribute to the results that were obtained. Rather, differences in the onset density significantly predicted the reaction times obtained in the present experiment.

Discussion

The results of Experiment 2 show that words with a sparse onset are responded to more quickly than words with a dense onset, replicating the results of Experiment 1, which used an auditory naming task. Taken together these experiments demonstrate that the initial portion of a word is psychologically important for the processing of spoken words (e.g., Cole & Jakimik, 1980; Jusczyk et al., 1999; Treiman & Danis, 1988; Walley, 1987), perhaps because it activates lexical candidates in memory (e.g., Marslen-Wilson & Zwitserlood, 1989). The results of the present set of experiments underscore the influence of the number of lexical candidates on the processing of spoken words (e.g., Luce & Pisoni, 1998). Moreover, these findings suggest that it is not just the total number of candidates activated in memory that affects spoken-word recognition. Rather, the location of the neighbors in the neighborhood may also influence lexical processing. Specifically, words with few neighbors sharing the initial phoneme of the target word are responded to more quickly than words with many neighbors sharing the initial phoneme of the target word.

General Discussion

The results of the present experiment show that target words with sparse onsets are responded to more quickly than target words with dense onsets. This finding suggests that the initial portion of a word (e.g., Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Zwitserlood, 1989) *and* the number of lexical competitors (e.g., Luce & Pisoni, 1998) influence the process of spoken-word recognition. These results are consistent with the findings of Sevald and Dell (1994), who found that participants named sequences of CVC words more quickly if the final part of the CVC was the same than if the initial part of the CVC was the same. That is, sequences like PICK TICK were produced more quickly than sequences like PICK PIN. Stated in terms of the present experiment, sequences with sparse onsets (PICK TICK) were repeated more quickly than sequences with dense onsets (PICK PIN). The results of Sevald and Dell, like the results of the present experiment, emphasized the psychological importance of the initial part of a word.

Sevald and Dell (1994) hypothesized that the location-specific effects of competition observed in their experiments could only be explained by representations (corresponding to phoneme nodes in their model) that existed between phonological word forms and acoustic–phonetic output in the speech production process. The results of the present experiment suggest that the same may hold true for speech perception: Intermediate representations may exist between acoustic–phonetic input and phonological word forms in the lexicon. The present findings are concordant with a number of

recent findings that suggest that lexical and sublexical representations may be required to adequately account for the process of spoken-word recognition (e.g., Auer & Luce, 1998; Luce, Goldinger, Auer, & Vitevitch, 2000; McClelland & Elman, 1986; Norris, 1994; Pitt & Samuel, 1995; Vitevitch et al., 1999; Vitevitch & Luce, 1998, 1999).

Models of spoken-word recognition that have intermediate representations, such as Shortlist (Norris, 1994), TRACE (McClelland & Elman, 1986), and PARSYN (Luce et al., 2000), may be able to account for the position-specific effects observed in the present data. In contrast, models of spoken-word recognition that do not have sublexical representations, such as cohort theory (Marslen-Wilson, 1987) and the neighborhood-activation model (Luce & Pisoni, 1998), may not be able to account for the present data without substantial modification of some kind. For example, in cohort theory, cohort size may need to be weighted in such a way that it will influence the recognition process. In the neighborhood-activation model, for example, similarity in the initial portion of a word may need to affect processing more than similarity in the final part of a word.

Although the results of the present set of experiments are consistent with several other findings that suggest the initial portion of a word is psychologically important (e.g., Marslen-Wilson & Welsh, 1978), these findings contrast with the results of Connine et al. (1993; see also Slowiaczek, Nusbaum, & Pisoni, 1987). Connine et al. found that target words were primed by nonwords that differed in a phonetic feature from the target word whether the different features occurred in phonemes that were in the initial or medial position in the nonword. They suggested that lexical access occurred on the basis of overall goodness of fit and not on the basis of word-initial information. In other words, there was no special status afforded to the initial portion of a word.

It should be noted, however, that Connine et al. (1993) used words that were at least two syllables long, whereas monosyllabic words were used in the present set of experiments. Wiener and Miller (1946) found that longer (i.e., multisyllabic) words were recognized more accurately than shorter (monosyllabic) words. In the case of multisyllabic words, additional information (e.g., phonotactic information) may be available in other parts of the word that may be redundant with the information provided by the initial portion of the word. Connine, Titone, Deelman, and Blasko (1997; see also Connine, 1994) described as "lexical extent" the later information found in longer words that can confirm possible hypotheses formed about the identity of a word on the basis of earlier information. In the case of monosyllabic words, the additional and redundant information that is found elsewhere in multisyllabic words may not be available as a result of the lack of other syllables in monosyllabic words. The lack of redundant information in monosyllabic words may make the initial portion of a monosyllabic word more important for word recognition than the initial portion of a multisyllabic word (because the same information can be found somewhere else in the word). The presence of lexical extent in longer words and lack of lexical extent in shorter words may account for the difference between the results of the present set of experiments and the results of Connine et al. (1993).

A computational analysis of recognition points carried out by Luce (1986) provided evidence to support the hypothesis that longer, multisyllabic words have greater lexical extent than shorter, monosyllabic words. Luce found that for most monosyl-

labic words, the *recognition point*—the point at which a word diverges or becomes unique from all other words in the lexicon—occurred after word offset. If the recognition point of a word is considered a measure of lexical extent (i.e., the information after the recognition point is redundant information), then the redundant information that may be available in later parts of multisyllabic words is not available in shorter words. Because the information needed to identify monosyllabic words starts with and may only be available in the initial segment of the word (cf. *tack* and *pack*), the initial part of a short word may therefore be very critical for the recognition of that word. The importance of the initial segment of a word demonstrated in the present experiments may, therefore, be a reflection of the availability (or the lack of availability) of redundant information in short, monosyllabic words.

In the discussion of lexical extent, word length was considered in terms of the number of syllables in the word rather than the temporal length, or duration, of the word. It should be noted that the monosyllabic words used in the present set of experiments had a mean duration of approximately 900 ms, which might be considered long for a monosyllabic word (i.e., spoken at a slow rate of speech). The number of syllables in a word and the speech rate are simply two of many sources of (potentially redundant) information present in the speech signal (see, e.g., Pisoni, 1996, for other types of information included in the speech signal). Under different listening conditions, different sources of information or cues in the speech signal may become more or less relevant or reliable for the process of recognizing the spoken word. Indeed, there is a long history of trading relations in speech research (e.g., Denes, 1955) and in psychology in general (e.g., trade offs between speed and accuracy, risks and gains). Thus, onset density may be a helpful piece of information for spoken-word recognition at a slow rate of speech but may be less helpful at a faster rate of speech (possibly because the next speech sound may be presented before processing of the initial sound has been completed). In any case, the results from the present set of experiments show that the number of competitors activated by the initial part of a (monosyllabic) word influences the speed with which a spoken word is recognized. These results suggest that intermediate representations (i.e., between acoustic–phonetic input and phonological word forms) may be required to recognize spoken words (cf. Marslen-Wilson & Warren, 1994).

References

Auer, E. T., & Luce, P. A. (1998). *PARSYN: A processing model of neighborhood activation and phonotactics in spoken word recognition.* Unpublished manuscript, University at Buffalo.

Charles-Luce, J., Luce, P. A., & Cluff, M. S. (1990). Retroactive influences of syllable neighborhoods. In G. T. M. Altmann (Ed.), *Cognitive models of speech perception: Psycholinguistic and computational perspectives* (pp. 173–184). Cambridge, MA: MIT Press.

Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken bisyllabic words. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 551–563.

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

Cole, R. A. (1973). Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics, 13,* 153–156.

Cole, R. A., & Jakimik, J. (1980). How are syllables used to recognize words? *Journal of the Acoustical Society of America, 67,* 965–970.

Connine, C. M. (1994). Vertical and horizontal similarity in spoken-word recognition. In C. Clifton Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 107–120). Hillsdale, NJ: Erlbaum.

Connine, C. M., Blasko, D. M., & Titone, D. A. (1993). Do the beginning of words have a special status in auditory word recognition? *Journal of Memory and Language, 32,* 193–210.

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language, 37,* 463–480.

Denes, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America, 27,* 761–764.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E.C.T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27–85). Hillsdale, NJ: Erlbaum.

Fudge, E. C. (1969). Syllables. *Journal of Linguistics, 5,* 253–286.

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 and Performance, 26,* 166–183.

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

Jusczyk, P. W., Goodman, M. B., & Baumann, A. (1999). Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language, 40,* 62–82.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33,* 630–645.

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

Kirk, K. I., Pisoni, D. B., & Miyamoto, R. C. (1997). Effects of stimulus variability on speech perception in listeners with hearing impairment. *Journal of Speech and Hearing Research, 40,* 1395–1405.

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

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., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics, 62,* 615–625.

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

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 142–147). Cambridge: MIT Press.

MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology, 3,* 210–227.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. In U. H. Frauenfelder & L. K. Tyler (Eds.), *Spoken word recognition* (pp. 71–102). Cambridge, MA: MIT Press.

Marslen-Wilson, W. D., & 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.

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 576–585.

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

Murphy, K. R., & Myors, B. (1998). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests.* Hillsdale, NJ: Erlbaum.

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 Rep. No. 10). Indiana University, Psychology Department, Speech Research Laboratory.

Pisoni, D. B. (1996). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullenix (Eds.), *Talker variability in speech processing* (pp. 9–32). San Diego, CA: Academic Press.

Pitt, M. A., & Samuel, A. G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology, 29,* 149–188.

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.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition, 53,* 91–127.

Slowiaczek, L. M., Nusbaum, H. C., & Pisoni, D. B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 64–75.

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

Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychology and Aging, 11,* 333–341.

Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging, 14,* 458–472.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition, 15,* 49–74.

Treiman, R. (1986). The division between onsets and rimes in English syllables. *Journal of Memory and Language, 25,* 476–491.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 145–152.

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

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., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language, 68,* 306–311.

Walley, A. C. (1987). Young children's detections of word-initial and -final mispronunciations in constrained and unconstrained contexts. *Cognitive Development, 2,* 145–167.

Walley, A. C., & Metsala, J. L. (1990). The growth of lexical constraints on spoken word recognition. *Perception & Psychophysics, 47,* 267–280.

Wiener, F. M., & Miller, G. A. (1946). Some characteristics of human speech. In *Transmission and reception of sounds under combat conditions* (Summary Tech. Rep.; pp. 58–68). Washington, DC: National Defense Research Committee, Division 17.

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.

(Appendix follows)

Appendix

Stimulus Words Used in the Present Experiments

Sparse-onset condition			Dense-onset condition		
deep	lob	roar	doll	lice	rim
dare	lip	rare	duke	leave	roam
dip	mare	cell	dud	mass	siege
dad	mug	sash	dies	mice	cease
dock	mop	sag	dime	mud	serve
dot	mob	sad	dull	moss	seam
dash	merge	tail	dim	mess	town
fad	map	tug	firm	miss	term
folk	mock	tag	foul	math	tool
fake	peer	tare	full	pitch	tern
keep	poor	top	calf	peach	type
kite	pair	ware	curl	pass	wish
cook	rock	womb	cape	rid	word
care	rear	wall	cove	rhyme	wife
lot	rob	weep	leaf	raid	wipe