

Competition in Spoken Word Recognition: Spotting Words in Other Words

James M. McQueen, Dennis Norris, and Anne Cutler

Although word boundaries are rarely clearly marked, listeners can rapidly recognize the individual words of spoken sentences. Some theories explain this in terms of competition between multiply activated lexical hypotheses; others invoke sensitivity to prosodic structure. We describe a connectionist model, SHORTLIST, in which recognition by activation and competition is successful with a realistically sized lexicon. Three experiments are then reported in which listeners detected real words embedded in nonsense strings, some of which were themselves the onsets of longer words. Effects both of competition between words and of prosodic structure were observed, suggesting that activation and competition alone are not sufficient to explain word recognition in continuous speech. However, the results can be accounted for by a version of SHORTLIST that is sensitive to prosodic structure.

To understand a linguistic message, a listener or reader must recognize the individual words in that message. In one respect, the reader has a distinct advantage over the listener in this task. The spaces between written words are clearly marked on the printed page, giving the reader unambiguous cues to the location of word boundaries. Spoken language, however, does not cue the listener in a similar way; word boundaries are not reliably marked (Lehiste, 1972; Nakatani & Dukes, 1977). However, the continuous speech signal is nevertheless perceived as a discontinuous string of words. How is this lexical parse obtained?

Three answers to this question have been suggested in the literature. One is that words are recognized in sequential order, as they occur in the stream of speech. Another is that there is some explicit mechanism that identifies, on the basis of sublexical information, points in a speech stream that are likely to be word boundaries. The third is that word recognition is achieved by a process of interword competition.

Sequential Recognition

Certain models of spoken word recognition propose that words are recognized in sequential order. The suggestion is that the onset of the next word can be accurately located when the current word has been recognized successfully (Cole &

Jakimik, 1978, 1980; Marslen-Wilson & Welsh, 1978). According to this argument, some words become unique before their offsets. Thus, if a listener hears *trespass*, for example, once the first five phonemes /trɛsp/ have been recognized, the only word that the string can become is *trespass*. The claim is that the onset of the next word must thus occur after /pəs/. Such models therefore do not need an explicit mechanism for locating word boundaries; boundaries emerge as a by-product of the recognition process.

The success of these sequential models depends on the proportion of words becoming unique before their offset. Distributional evidence does not, however, favor such models. Luce (1986a) computed the uniqueness point (the point at which a word diverges from all other words) of each word in a 20,000 word dictionary and found that 60% of these words diverged before their final phoneme. However, the most frequent words in the language are shorter words, which tend to become unique only after their offsets. When word frequency was taken into account, the probability of a word becoming unique on its last phoneme was .23, and the probability of a word becoming unique before its last phoneme was only .39. The consequence of these statistics is that more than a third of words encountered will only be recognizable after the listener has heard part of the following word. McQueen and Cutler (1992) have further shown that a majority of polysyllabic words in English have shorter words embedded within them and that—particularly problematic for sequential models—a majority of such embeddings occur at the onset of the polysyllabic words. Frauenfelder (1991) has also found a considerable degree of lexical embedding in a statistical analysis of the Dutch vocabulary.

Shillcock (1990) has pointed out another weakness of sequential models. Because of suffixation, words like *trespass* do not become unique before their offset: *trespassing*, *trespasses*, *trespassed*, and *trespasser* are all viable candidates. Recognition of *trespass*, or of one of these suffixed forms, cannot be achieved until the arrival, or nonarrival, of the affix. Where there is no affix, recognition must thus be delayed until the onset of the following word.

Experimental evidence also argues against sequential recog-

James M. McQueen, Dennis Norris, and Anne Cutler (now at Max-Planck-Institute for Psycholinguistics), MRC Applied Psychology Unit, Cambridge, United Kingdom.

This work was supported by the Joint Councils Initiative in Cognitive Science and HCI, Grant E304/148. We thank the Longman Group, United Kingdom Inc., for allowing us to use a machine-readable version of the *Longman Dictionary of Contemporary English* (Procter, 1975). We also thank Gerry Altmann and three anonymous reviewers, whose comments greatly improved this article, and Sally Butterfield and Brit van Ooyen, for assistance with subject testing.

Correspondence concerning this article should be addressed to James M. McQueen, who is now at Max-Planck-Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands. Electronic mail may be sent to james@mpi.nl.

dition. Grosjean (1985) and Bard, Shillcock, and Altmann (1988) have shown with the gating task in continuous speech (where subjects hear successively longer and longer parts of a sentence) that many words can only be recognized after the onset of the following word. Clearly, the strongest version of the sequential processing hypothesis is untenable: Continuous word recognition cannot *depend* on recognition of every word before its acoustic offset.

In response to criticisms such as these, Marslen-Wilson (1987) has argued that a solution can be provided by the use of higher level information in the word recognition process. The process of retrieval from lexical storage (the "access function"; Marslen-Wilson, 1987) may in fact occur very early (perhaps 200 ms after word onset; Marslen-Wilson, 1973). Retrieval will, however, provide multiple lexical candidates; contextual (syntactic, semantic, and pragmatic) information will act to speed up the process of lexical selection such that monosyllabic or suffixed words can be recognized before their uniqueness points. In Marslen-Wilson's cohort model (Marslen-Wilson & Welsh, 1978), as in other interactive models of word recognition, word recognition can be substantially affected by top-down feedback from syntactic and semantic processing levels. However, words that do not become unique until after their offset can occur in neutral contexts or can be contextually anomalous. For these reasons, it has been argued that contextual information cannot be *relied* on to provide a boost to the recognition process (Norris, 1982).

Explicit Segmentation

A very different approach to word recognition in continuous speech is provided by models that postulate an explicit process of lexical segmentation. The focus in this case is on where word boundaries are likely to be and, thus, on where in the signal it is appropriate to initiate lexical access. Such models make claims about how lexical access can be more efficient if possible word-boundary locations can be identified before lexical access. Cross-linguistic analyses (Cutler, Mehler, Norris, & Segui, 1986; Otake, Hatano, Cutler, & Mehler, 1993), particularly those with bilingual subjects (Cutler, Mehler, Norris, & Segui, 1992), have suggested that although these segmentation strategies vary across languages, they do have something in common: They are all based on prosodic structure.

Cutler and Norris (1988; see also Cutler & Carter, 1987) have proposed a Metrical Segmentation Strategy for stressed languages like English, which are characterized in terms of strong and weak syllables. Strong syllables are defined as those that contain a full vowel, and they are contrasted with weak syllables, which contain reduced vowels, usually schwa. A syllable is considered strong whether it carries the primary stress or only a secondary stress in a word. The Metrical Segmentation Strategy assumes that strong syllables trigger segmentation of speech. It claims that strong syllables are points in the speech signal at which lexical access is initiated because they are the most likely locations of content word onsets. The speech stream is thus segmented in the sense that the metrical segmentation strategy postulates word boundaries within it.

As Norris and Cutler (1985) have pointed out, segmentation does not require classification. A fully categorized prelexical

parse of the speech input into phonemes, syllables, or any other nonlexical unit is not necessary for the operation of the Metrical Segmentation Strategy. All that is required for the strategy is a prelexical mechanism that can detect strong syllables and initiate lexical access at these locations. The Metrical Segmentation Strategy is consistent both with models that classify the signal prelexically and with models in which lexical access is based on a raw acoustic representation (Cutler & Norris, 1988).

A word-spotting task has provided evidence supporting the Metrical Segmentation Strategy (Cutler & Norris, 1988). In that experiment, listeners were asked to listen to bisyllabic nonsense strings and to press a button if they heard a real word embedded at the beginning of these nonsense strings. It was found that monosyllabic words embedded as the first (strong) syllables of the strings were more difficult to detect when the second syllable was also strong (e.g., *mint* in /mɪntɛɪf/) than when the second syllable was weak (e.g., *mint* in /mɪntəf/). It was argued that the second syllable of /mɪntɛɪf/, being strong, was segmented from the first, and thus that *mint* had to be assembled across a segmentation boundary. Detection of *mint* in a strong-weak (SW) item was easier because the weak second syllable did not trigger segmentation.

Analyses of both natural and laboratory-induced misperceptions of speech (Cutler & Butterfield, 1992) indicate that listeners tend to assume that strong syllables are the onsets of words (the onsets of content words, to be more precise). The principal claim of the Metrical Segmentation Strategy, that content words are likely to begin at strong syllables, is thus supported by converging evidence from different tasks. Furthermore, such a strategy is appropriate for English: Cutler and Carter (1987) found that more than 90% of content words in a corpus of conversational English began with strong syllables.

Interword Competition

Another solution to the continuous speech problem has been provided by models of spoken-word recognition invoking mechanisms of competition. Like sequential recognition models, these models achieve early recognition when this is possible (i.e., with a word with an early uniqueness point); however, they avoid the strong claim that words must be recognized in strict sequential order. The TRACE model (McClelland & Elman, 1986) is of this type. In TRACE, continuous word recognition is achieved by a lexical competition process. Word nodes are activated on the basis of available bottom-up information. By a process of lateral inhibition, words beginning from the same point compete with each other. Competition between the cohort of words that share initial portions is also assumed in the (sequential) cohort model (Marslen-Wilson, 1987, 1990; Marslen-Wilson & Welsh, 1978). In TRACE, however, there is an additional type of competition. Words beginning at *different* points in time also compete for control of the same input segments. Thus, given the input /kætələg/, the candidates *cat* and *catalogue* will compete for the initial three phonemes, whereas *log* and *catalogue* will compete for the final three phonemes.

Competition between candidates straddling different parts of the input string in effect provides a mechanism for word

recognition (see Bard, 1990; Frauenfelder & Peeters, 1990, 1992). McClelland and Elman (1986) provided some worked out examples. Thus, if TRACE is given a string like /barti/, competition occurs between the word nodes for *bar*, *art*, and *tea*. The parse *bar tea* is chosen, and activation of *art* is suppressed, because only *bar tea* allocates all the input phonemes to words, with no phonemes left over (see Figure 28 of McClelland & Elman, 1986, p. 63).

However, the architecture of TRACE is implausible. The rationale for competition is that words can potentially begin anywhere in the input stream. In TRACE, the ubiquity of potential onset locations is dealt with by duplicating lexical networks so that there is a complete lexical network aligned with each point in the input where a word might begin. Thus, to recognize all possible words in an input stream 50 phonemes in length, TRACE would require 50 complete lexical networks. Furthermore, lexical nodes in nearby lexical networks would need to be connected by inhibitory links so that overlapping words inhibit each other. The total number of connections required is enormous.

In the SHORTLIST model (Norris, in press), the problem of duplicating lexical networks is avoided by separating the process of competition from the process of recognizing potential lexical candidates. SHORTLIST has two distinct stages. In the first stage, all potential lexical candidates beginning at every phoneme in the input are generated in a completely bottom-up fashion. This stage need not take account of whether or not candidates overlap. These shortlisted candidates are wired into a small interactive activation network that functions like the lexical level of TRACE. The candidate words in this small competitive network then compete for recognition. Because competition is limited to a small candidate set, the model copes well with a large lexicon of over 25,000 words.

The initial stage of the SHORTLIST model, which generates lexical candidates, clearly has a complex task to perform. In a current implementation of the model (Norris, in press) this first stage consists simply of an exhaustive dictionary search procedure that can produce all possible lexical candidates. However, this process of generating lexical candidates could also be performed by a system much like the simple recurrent network studied by Norris (1990, 1992). Norris demonstrated how a three-layer back-propagation network with recurrent connections could be trained to recognize words in continuous input and could simulate many of the characteristics of human spoken word recognition. However, although recurrent networks can identify words from continuous input without the duplication of lexical networks required by TRACE, they are unable to parse input reliably. For instance, these networks would detect *cat*, *log*, and *catalogue* given the input /kætələg/, but they would not be able to select from among such alternatives because there is not a mechanism for this. The lexical competition network in SHORTLIST provides exactly such a mechanism.

Finally, recent studies of spoken word recognition have provided evidence for competition effects in the human listener. Experiments with the cross-modal priming task (Swinney, 1979), for example, have suggested that multiple lexical entries may be activated when a word is presented (Marslen-

Wilson, 1987, 1990; Shillcock, 1990; Swinney, 1981; Zwitserlood, 1989). Further evidence for activation and competition comes from results indicating that the recognition of a spoken word depends on its frequency of occurrence and its similarity neighborhood, that is, on the number and frequency of occurrence of phonetic neighbors (see Luce, Pisoni, & Goldinger, 1990, for a review).

Reconciling Alternative Approaches

There is at least some experimental evidence in favor of all three approaches that we have outlined. Early recognition of words with early uniqueness points suggests support for sequential recognition. Prosodic effects on segmentation suggest that spoken word recognition is influenced by the prosodic structure of the input language through a possible explicit procedure of segmentation. Evidence for multiple activation of word candidates in spoken word recognition suggests that recognition involves a process of activation and competition.

Moreover, in some respects the three types of approaches are not incompatible with one another. First, competition models such as TRACE and SHORTLIST incorporate a version of sequential processing. When the evidence in the input strongly supports a unique lexical entry, the model will be highly likely to parse the string with a word boundary at the end of this word, and the effect of this will be that words beginning immediately after the juncture will be advantaged. Thus, on words with early uniqueness points, competition models will operate in a way analogous to sequential recognition models.

It is important to note, however, that competition models do not depend on a sequential mechanism. They can cope just as well with words that do not become unique before their offsets. Thus, the apparent sequential recognition produced by competition models with certain forms of input should perhaps be thought of merely as an aspect of the more general mechanism of lexical competition. As we have argued, sequential recognition will in fact not work with most words; it might therefore be argued that early recognition of words with early uniqueness points could just as well serve as evidence in favor of competition models.

Second, an activation-competition model such as SHORTLIST is clearly compatible with explicit segmentation strategies such as the Metrical Segmentation Strategy. The Metrical Segmentation Strategy stipulates where in the signal lexical access should occur; SHORTLIST assumes that lexical access should be considered as the activation of word hypotheses and that word recognition is based on competition between these candidates. In SHORTLIST terms, then, the Metrical Segmentation Strategy can be viewed as a factor that either determines which lexical candidates should be activated or influences activation levels so that competitors beginning at strong syllables enjoy an advantage. In the first of these alternatives, the Metrical Segmentation Strategy would act to initiate lexical access at strong syllables, and thus not every phoneme in the input would be treated as a possible word onset. In the second alternative, the access component of SHORTLIST would be unaffected, and the Metrical Segmentation Strategy would operate as a bias in the competition process. In either case, however, it is clear that the central claim of the Metrical

Segmentation Strategy, that lexical access is more likely to be successful at strong syllables, can easily be incorporated in the competition framework of SHORTLIST.

What remains unclear, however, is whether both of these approaches are necessary for an adequate account of continuous speech recognition. It is possible, for example, that some of the evidence taken as support for the Metrical Segmentation Strategy could be accounted for by mechanisms of lexical competition. Indeed, Cutler and Norris (1988, p. 120) have suggested that the difficulty of detecting *mint* in /mɪnteɪf/ may in part be due to competition of lexical hypotheses.

The present study was undertaken to address this question. First, we tested the mechanisms of activation and competition, as instantiated in SHORTLIST, by examining competition between words beginning both at the same and different points in the signal. Second, we manipulated prosodic factors to test for effects predicted by the Metrical Segmentation Strategy.

The task chosen was the word-spotting task developed by Cutler and Norris (1988), as just described. In Cutler and Norris's study, the embedded words always appeared at the beginning of the nonsense strings. In Experiment 1, however, target words could appear at either the beginnings or the ends of the strings. The crucial manipulation was whether or not the strings were themselves the beginnings of longer real words. Targets, such as *mess*, could appear as the second syllable of word onsets like /dəməs/ (the beginning of *domestic*) or matched items such as /nəməs/ (which does not begin a word). Competition between words beginning together was also tested by using SW pairs such as /sækrəf/ (the onset of *sacrifice*) and /sækrək/ (a nonword onset).

The competition prediction is straightforward: Activation of *domestic* should make it more difficult to detect the target *mess* in /dəməs/ than in /nəməs/, in which there is no long competitor word. In the weak-strong (WS) strings, the task thus tests for competition between words beginning at different parts of the string. Likewise, competition predicts that it should be more difficult to detect *sack* in /sækrəf/ than in /sækrək/. Here, the task tests for competition in words beginning in the same way.

The Metrical Segmentation Strategy makes no predictions about competition effects for items with the same stress pattern, but it does predict a main effect of stress pattern: that word-spotting should be easier for WS items than for SW items. The Metrical Segmentation Strategy predicts that lexical access should be initiated at strong syllable onsets. Thus, for a WS item, lexical access should occur at the second (strong) syllable, the target word, segmenting the item at the appropriate point for target detection. In contrast, lexical access should occur at the first syllable and not at the second (weak) syllable of an SW item. There will therefore be no internal segmentation of SW items, making target detection more difficult.

Experiment 1

Method

Subjects. Thirty-four student volunteers were paid for their participation. Most of the subjects were members of King's College, Cambridge, United Kingdom. There were 21 women and 13 men, all between 18 and 30 years of age.

Materials. The materials consisted of two sets of 36 yoked triplets of bisyllables. Items in the first of these sets had embedded targets: monosyllabic words that were either the first or the second syllable of the item. One item ("word onset") in each triplet was the first two syllables of a polysyllabic word. One half of the word onsets consisted of a weak syllable followed by a strong syllable (WS), where the target word appeared in the second syllable (e.g., /dəməs/, the onset of *domestic*, with the target *mess*). The other word onsets consisted of a strong followed by a weak syllable (SW), where the target word appeared in the first syllable (e.g., /sækrəf/, the onset of *sacrifice*, with the target *sack*). The other two items ("nonword onsets") in each triplet contained the same target word but could not be continued to form words. One nonword onset in each triplet had a WS stress pattern, the other nonword onset had a SW pattern. The 18 triplets of each type (i.e., on the basis of WS or SW word onsets) are listed in the Appendix.

The structure of the language made it impossible to use fully matched quadruplets; that is, where the same target word appeared as the strong syllable of two WS and two SW items, with one of each being the onset of a real word (only a few such sets can be constructed, e.g., a set for *come*: /ənkʌm/, *encumber*: /ɪnrʌkʌm/; /kʌmpən/, *company*; and /kʌmpəg/). The partially between-items design was therefore adopted, with both WS and SW word onsets yoked to nonword onsets of both stress patterns (e.g., WS word onset /dəməs/ with WS nonword onset /nəməs/ and SW nonword onset /məstəm/, and SW word onset /sækrəf/ with SW nonword onset /sækrək/ and WS nonword onset /kləsæk/). Thus, there was only one set of word onsets of each stress pattern, but two sets of nonword onsets of each stress pattern. Items in one of these sets of nonword onsets ("matched") contained the same target words as the word onsets with that stress pattern (e.g., WS /nəməs/ matched with /dəməs/, and SW /sækrək/ matched with /sækrəf/). In items from the other set ("unmatched"), the target words were not the same as those in the word onsets with that stress pattern (e.g., WS /kləsæk/ not matched to /dəməs/, and SW /məstəm/ not matched to /sækrəf/). In other words, the unmatched WS nonword onsets contained the targets used in the SW word onsets, and vice versa.

There were two important constraints on the choice of these materials. First, there had to be only one embedded word in each item. For example, the string /dərəɪz/ has the word *rise* as its second syllable, but the words *rye*, *eye*, and *eyes* are also possible word targets. Second, it was hoped that there would be only one embedding word, such that the word-onset materials could only be completed in one way. This constraint proved impossible to satisfy, so a weaker criterion was adopted. A word-onset string was accepted if it could be completed by only one set of morphologically related words. Thus, /fələs/ was accepted, although it could be completed as *philosopher*, *philosophize*, and *philosophy* (and further inflections). These constraints made it difficult to generate a fully balanced set of materials. For example, it was impossible to match the frequency of occurrence of either the target words or the embedding onsets (i.e., of *domestic* and *sacrifice*) across the two subsets of 18 target-bearing triplets. However, the frequencies listed in the Appendix (from Francis & Kučera, 1982) show that the words were at least chosen from overlapping frequency distributions.

Another constraint that proved difficult to satisfy involved the syllabification of the embedding words. We wanted to select items in which the embedded target exactly matched the syllabification of the embedding word. This was not problematic for the WS items because the medial consonants in a WS string clearly belong to the second, strong syllable (under the principle of maximal onset, Selkirk, 1982). There is, however, a problem with SW items because there is little consensus on the correct syllabification of such items. Different linguistic theories make different claims (cf., e.g., Pulgram, 1970; Selkirk, 1982; Kahn, 1976), and listeners, although showing certain sys-

tematic patterns, are far from being in agreement on syllabification either (Treiman, Gross, & Cwikiel-Glavin, 1992; Treiman & Zukowski, 1990). Some words are fairly unambiguous, however. Phonotactic constraints mean that certain phonemic sequences cannot be syllable-internal. Thus, the first syllable of *hypnotize* is the word *hip* because the sequence /pn/ cannot form a syllabic coda or onset. Items of this type were chosen if possible. Furthermore, monosyllabic English words with lax vowels have to be closed syllables, ending with at least one consonant (i.e., lax vowels do not occur in words of open consonant-vowel structure; Fudge, 1987). Thus, all target words we used (except *arm*) had lax vowels. The medial consonants in such items should therefore have been syllabified as part of the first syllable if that syllable was viewed in isolation (e.g., /hɛm.əs/ rather than /hɛ.məs/ because /hɛm/ is a possible syllable, whereas /hɛ/ is not). The SW items were therefore highly consistent, if somewhat less well constrained in their syllabification than the WS items.

The second set of 36 triplets is also listed in the Appendix. These were fillers in which there were no embedded word targets. They were structured in the same way as the target-bearing items such that the same monosyllabic nonword appeared in three embedding strings, one of these strings being the onset of a real word (e.g., /fəls/, the onset of *felicity*) and the other two strings not being the onsets of real words, but varying in stress pattern, consisting of one WS item (e.g., /kəls/) and one SW item (e.g., /lɪsəl/). Eighteen of the triplets had WS word onsets (/fəls/); the other 18 had SW word onsets (e.g., /dʒævəl/, the onset of *javelin*). As with the target-bearing items, the embedding word strings were selected so that the cohort of completions were all morphologically related.

An additional set of 18 items were selected. There were 9 WS items and 9 SW items; 9 were the onsets of words and 9 contained target words. These were used to construct a practice tape of 12 items, and as the first 6 items in the experimental run, as further warm-up materials.

Simulations. The experimental predictions of lexical competition were confirmed for the materials by simulation. The simulations were performed with SHORTLIST operating on a 26,450-word subset of the *Longman Dictionary of Contemporary English* (Procter, 1975). Note that although large-vocabulary simulations of this nature can easily be performed with SHORTLIST, they are beyond the scope of current implementations of TRACE. TRACE is normally limited to less than 1,000 words and uses only a subset of the English phoneme inventory. SHORTLIST has the additional advantage of being a completely bottom-up system in which there is no feedback from the lexical to the phonemic level.

For the simulations, each target-bearing item in the materials was transcribed and used as input to the model. Performance on the different classes of material was compared at "time slices" (i.e., for each additional phoneme) moving through the items, and for up to four time slices of silence after item offset. In four of the unmatched SW nonword onsets, -er suffixed words (*wrecker*, *robber*, *fisher*, and *ticker*) won out in the competition process, markedly reducing the activation of the target monosyllables (*wreck*, *rob*, *fish*, and *tick*). These four items were removed from the simulation (in fact, polysyllabic words were hardly ever spotted by subjects either in the present experiments or in previous studies with this task). The simulations were run with the eight default parameters specified in Norris (in press). The mean activation functions for the targets in WS items, and the embedding words in the appropriate subset of these items, are shown in Figure 1; those for the targets and embedding words in the SW items are shown in Figure 2.

The competition effects can be seen by comparing the activation levels of the target words. In WS items, targets in word onsets have a lower level of activation than those embedded in nonword onsets (e.g., *mess* in /dɛməs/ vs. /nɛməs/). This effect is only present near the offset of the target word (C in Figure 1). It is clear that it is due to the high activation of the embedding word (e.g., *domestic*) given the word-onset

items. The embedding word did not enter the candidate set for any of the WS matched nonword-onset items (e.g., *domestic* was not considered given the input /nɛməs/). The activation functions for targets in the matched and unmatched WS nonword onsets are thus indistinguishable (cf. *mess* in /nɛməs/ and *sack* in /klɔsæk/).

In SW items, however, targets have equivalent levels of activation in word onsets and matched nonword onsets (e.g., *sack* in /sækræf/ vs. /sækræk/) at the offset of the target (C in Figure 2). This is because the embedding words did enter the candidate set for the matched SW nonword onsets. The activation functions for targets in word onsets and matched nonword onsets only diverge later. Thus, SHORTLIST predicts differences in target detection in SW items between word onsets and matched nonword onsets (e.g., *sack* in /sækræf/ and /sækræk/), but that these should be smaller than, or emerge later than, those in WS items. SHORTLIST also predicts differences in target detection in SW matched and unmatched nonword onsets (e.g., *sack* in /sækræk/ and *mess* in /mɛstəm/). The matched nonword onsets tended to differ from the word onsets only in their final phoneme (see Appendix), and therefore became nonwords later than the unmatched nonword onsets. As a result, the activation levels of targets in unmatched nonword onsets are higher than those in matched word onsets. However, the model predicts that, over time, when the difference in activation between targets in SW word onsets (e.g., *sack* in /sækræf/) and SW matched nonword onsets (e.g., *sack* in /sækræk/) is large, the difference between SW matched and unmatched nonwords (e.g., *sack* in /sækræk/ and *mess* in /mɛstəm/) is small, and vice versa (see Figure 2).

Design and procedure. Three experimental tapes were constructed, each beginning with the 6 warm-up items and having the same running order of the 72 embedded strong syllables (36 word targets plus 36 nonword fillers). The strings in which these syllables were embedded were counterbalanced across the three tapes. Thus, for example, the first target in each tape was *dead*. On Tape 1, this target was embedded in /dɛdəl/, on Tape 2 in /dɛdæk/, and on Tape 3 in /gɛdɛd/. Each tape contained 12 target-bearing items that were word onsets (6 WS and 6 SW), 12 target-bearing WS items that were nonword onsets (half matched and half unmatched), and 12 target-bearing SW items that were nonword onsets (again, half from each subset). This counterbalancing was repeated for the set of nonword fillers. One half of the items on each tape contained targets, and targets were equally likely to occur in the first or second syllable. One half of the items that were word onsets contained embedded targets, and the other half did not.

The practice tape and the three experimental tapes were recorded by a male native speaker of British English in a sound-damped booth, onto the left channel of a digital audiotape. The items were spoken at the rate of one every 3 s. Timing pulses were placed on the right channel of this tape, aligned approximately with item onset. Before the data analysis, each item was digitized (sampling at 10 kHz with 12-bit A/D conversion) and examined using a speech editor. For each item, three measurements were made: the time between the timing pulse and target (strong syllable) onset, strong syllable length, and item length.

Subjects were tested separately in a quiet room. They were told that they would hear a list of nonsense words, presented individually and that they were to press the button in front of them as quickly as possible if the nonsense word began or ended with any real word and then say aloud, into a microphone, the word that they had spotted. These verbal responses were recorded onto audiotape. The target words were not given to the subjects in advance. The manual responses were made with a finger of the subject's preferred hand. These reaction times were collected by a microcomputer, with responses measured from the timing pulses (which were inaudible to the subjects). Subjects heard the items binaurally, over headphones. All subjects were given the same practice tape, followed by one of the three experimental tapes. The experiment took less than 10 min to run.

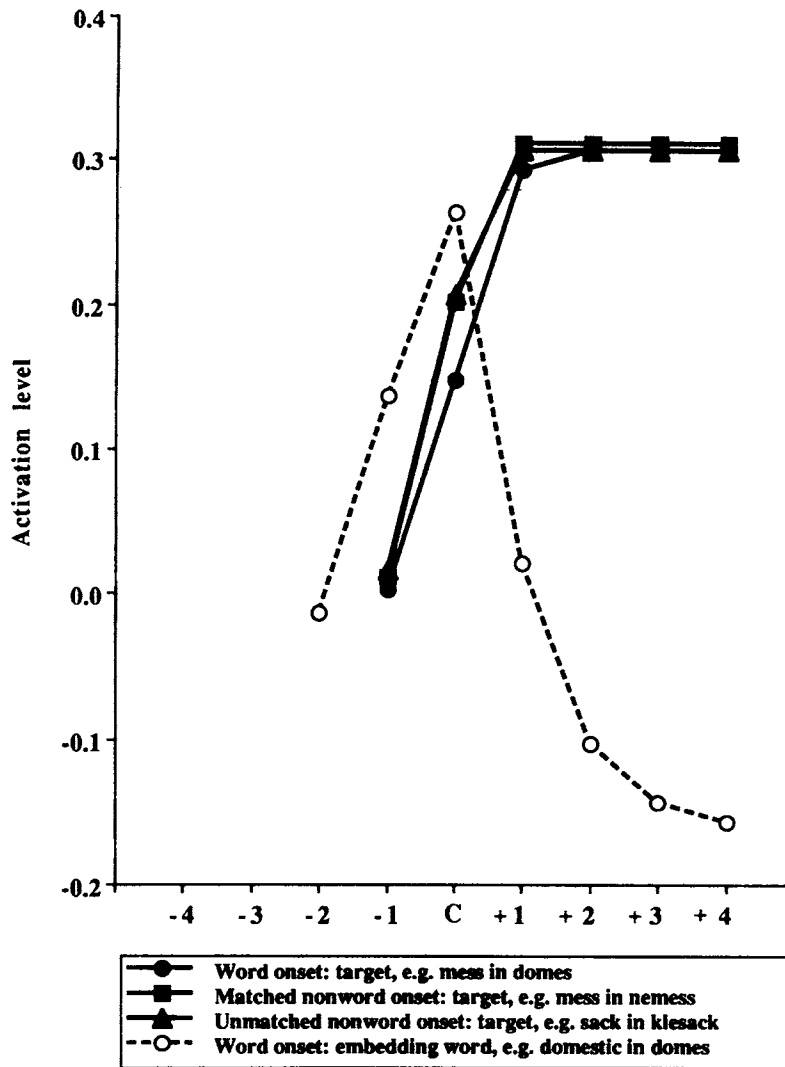


Figure 1. Mean activation levels of target words and embedding words given the weak-strong items, over time slices, in SHORTLIST. Filled symbols show activation of targets embedded in word onsets (circles; e.g., *mess* in /dɒməs/, the onset of *domestic*), in nonword onsets matched to word onsets (squares; e.g., *mess* in /nəməs/), and in unmatched nonword onsets (triangles; e.g., *sack* in /kləsæk/). Open circles show the activation of the embedding words in the word-onset items (e.g., *domestic* in /dɒməs/). The time slices are marked to indicate the alignment of the activation functions relative to the last consonant of the target word (C). Slices before C are for each phoneme working back through each item; slices after C contained silence markers.

Results and Discussion

The verbal responses were analyzed, and the number of missing responses for each subject was counted. Occasions on which subjects made a manual response to a target-bearing item but then either failed to give a verbal response or responded with a word other than the target were discounted: These responses were treated as missing data. Four subjects correctly detected less than 50% (18) of the targets and were therefore discarded, leaving three groups: 10 subjects for each tape. Response times (RTs) were adjusted so that they were measured from the offset of each embedded target word.

Responses of less than 200 ms or greater than 1,800 ms were also treated as missing data. For the RT analyses, missing data points for each subject were replaced with the mean of that subject's available responses for WS items and SW items, as appropriate, and missing data for each target were replaced with the mean of the available data, across subjects, for that target. The mean RTs, measured from target offset, and the mean error rates are given in Table 1.

Both the RT and the error rate data were subjected to two separate subanalyses, both with subjects (F_1) and items (F_2) as the repeated measure. In the first analysis, which we term the stress analysis, responses to nonword onsets (i.e., re-

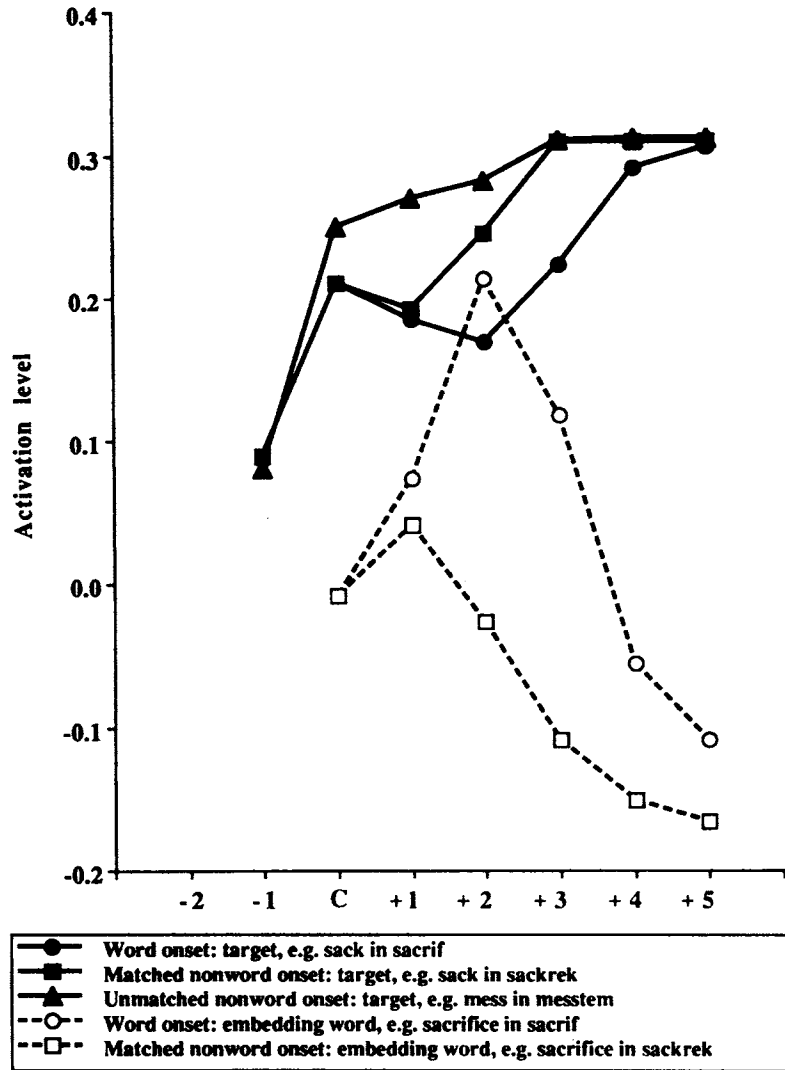


Figure 2. Mean activation levels of target words and embedding words given the strong-weak items, over time slices, in SHORTLIST. Filled symbols show activation of targets embedded in word onsets (circles; e.g., *sack* in /sækrəf/, the onset of *sacrifice*), in nonword onsets matched to word onsets (squares; e.g., *sack* in /sækrək/), and in unmatched nonword onsets (triangles; e.g., *mess* in /mestəm/). Open symbols show the activation of the embedding words in word onsets (circles; e.g., *sacrifice* in /sækrəf/) and in matched nonword onsets (squares; e.g., *sacrifice* in /sækrək/). The time slices are marked to indicate the alignment of the activation functions relative to the last consonant of the target word (C). Slices before C are for each phoneme working back through each item; slices after C contained the following phonemes in the bisyllable and then silence markers. Note from the Appendix that the bisyllables varied in the number of segments following the target, so the position of the offset of the bisyllable is variable. Note also that in one item pair (/hɪpnət/-/hɪpəʃ/), the matched nonword onset diverged from the word onset before the final phoneme. This is why the activation functions for the word onsets and matched nonword onsets begin to diverge on the first segment after the final consonant of the target.

sponses to /nəməs/, /kləsæk/, /sækrək/, and /mestəm/) were examined alone, ignoring the word-onset data, thus allowing a test of any effects of stress pattern in the absence of any competition effects. The comparison of matched and unmatched nonword onsets alone also provides a check on the reliability of any stress effects across all target words. In the second subanalysis, the competition analysis, the unmatched

nonword onsets were ignored, and the word-onset data were compared with the matched nonword-onset data (i.e., responses to items like /dəməs/, /nəməs/, /sækrəf/, and /sækrək/). This analysis allowed competition effects to be tested in a balanced design. The competition analysis also includes a test for effects of stress pattern to assess their interaction with effects of lexical competition.

Table 1
Mean Response Times (RTs; in Milliseconds) and Error Rates for Word Spotting of Targets in Weak-Strong (WS) and Strong-Weak (SW) Strings in Word Versus Nonword Onsets in Experiment 1

Stress pattern	Word onset	Nonword onset	
		Target matched	Target unmatched
WS			
RT	665	558	569
Error rate (%)	44	26	24
Example	/d̩m̩s/	/n̩m̩s/	/kl̩sæk/
SW			
RT	843	847	843
Error rate (%)	57	45	46
Example	/sækr̩f/	/sækr̩k/	/m̩st̩m/

Note. Mean RTs were measured from target offset.

In the stress analysis, there was a highly significant effect of stress pattern in the RT data, $F_1(1, 27) = 113.13, p < .001, MS_e = 20,978; F_2(1, 34) = 51.10, p < .001, MS_e = 13,184$, with responses to targets in WS items (e.g., /n̩m̩s/ and /kl̩sæk/) on average nearly 300 ms faster than responses to SW items (e.g., /sækr̩k/ and /m̩st̩m/). This pattern was repeated in the errors, $F_1(1, 27) = 37.17, p < .001, MS_e = 0.0341; F_2(1, 34) = 17.37, p < .001, MS_e = 0.0438$, with the error rate on WS items (25%) only half that on SW items (45%). There were no reliable differences (in either RTs or errors) between matched and unmatched nonword onsets and no interaction of this factor with stress pattern (i.e., responses to /n̩m̩s/, matched with /d̩m̩s/, and /kl̩sæk/, unmatched, were equivalent, as were responses to /sækr̩k/ and /m̩st̩m/). This finding makes it unlikely that any differences between the word-onset items can be attributed to differences between the target words.

In the competition analysis, there was again a highly reliable main effect of stress pattern. In RT, responses to WS items ($M = 611$ ms; e.g., /d̩m̩s/ and /n̩m̩s/) were about 230 ms faster than those to SW items ($M = 845$ ms; e.g., /sækr̩f/ and /sækr̩k/): $F_1(1, 27) = 89.78, p < .001, MS_e = 18,248; F_2(1, 34) = 20.01, p < .001, MS_e = 22,286$. In errors, the effect (WS, 35%; SW, 51%) was only significant by subjects: $F_1(1, 27) = 16.12, p < .001, MS_e = 0.0483; F_2(1, 34) = 2.88, .1 > p > .05, MS_e = 0.1623$.

Responses to targets in word onsets ($M = 754$ ms; e.g., /d̩m̩s/ and /sækr̩f/) were slower than responses to targets in nonword onsets ($M = 703$ ms; e.g., /n̩m̩s/ and /sækr̩k/): $F_1(1, 27) = 9.03, p < .01, MS_e = 8,734; F_2(1, 34) = 8.86, p < .01, MS_e = 4,393$. This competition effect was not equivalent across stress pattern, as revealed by a significant interaction, with a large effect for WS items and no effect for SW items: $F_1(1, 27) = 18.42, p < .001, MS_e = 4,925; F_2(1, 34) = 25.48, p < .001, MS_e = 4,393$. A different pattern was found in the error analyses. There were more errors to targets in word onsets (50%) than to targets in nonword onsets (35%), $F_1(1, 27) = 23.99, p < .001, MS_e = 0.0281; F_2(1, 34) = 13.05, p < .005, MS_e = 0.0310$, but, in contrast to the RT data, this word-nonword effect did not interact with the stress pattern effect (F_1 and F_2, ns).

Thus, according to the error data, subjects found it more difficult to spot targets in SW items than in WS items, and they

found it more difficult to spot targets in items that were word onsets than in items that were not word onsets. The same patterns were found in the RT data (WS faster than SW and nonword onsets faster than word onsets), but these two effects interacted. This interaction was examined by performing t tests. Subjects were slower to detect targets in the second syllables of WS items that were word onsets (*mess* in /d̩m̩s/) than in WS items that were not word onsets (*mess* in /n̩m̩s/): by subjects, $t_1(29) = 5.20, p < .001$; by items, $t_2(17) = 4.84, p < .001$. There was no significant effect for the SW items, by either subjects or items. The difference in target detection time between WS and SW word onset items (/d̩m̩s/ and /sækr̩f/) was significant both by subjects, $t_1(29) = 6.45, p < .001$, and by items, $t_2(34) = 2.29, p < .05$. The stress effect was also found within the matched nonword onsets. Targets were detected more rapidly in WS items than in SW items: by subjects, $t_1(29) = 10.38, p < .001$; by items, $t_2(34) = 5.59, p < .001$.

Most of these effects were replicated in RT analyses measuring from target onsets. These analyses do not control for word length, however. Separate analyses of variance (ANOVAs) of target and item lengths revealed that there were systematic length differences in the materials. In a comparison of target lengths in word onsets and matched nonword onsets, there was a significant stress pattern effect, $F(1, 34) = 115.76, p < .001, MS_e = 6,483$. Target words in WS items were, on average, 440 ms long; they were 236 ms long, on average, in SW items. There were no differences between word onsets and nonword onsets, and there was no significant interaction. The same pattern was found in a comparison of lengths of the complete items. On average, WS items were longer (566 ms) than SW items (529 ms), $F(1, 34) = 16.87, p < .001, MS_e = 8,294$. The large length differences acted to decrease the size of (but not remove) the RT effects when measuring from target onset.

A correlational analysis was also performed, testing for effects of word frequency. The frequencies of embedded and of embedding words for each item were separately correlated against both mean RT and mean error rate. The frequency of occurrence of the target (embedded) words did not predict RT or error performance in any condition. Nor did frequency of embedding words correlate with either RT or errors on the word-onset items.¹

The results support the following description. It is more difficult to spot target words in SW strings than in WS strings.

¹ Frequency effects have been measured in the word-spotting task (Freedman, 1992). Detection was faster and more accurate for higher frequency targets (> 130 counts per million) than for lower frequency targets (< 13 counts per million). However, there were no reliable differences comparing high-frequency targets with those of a medium range (30–97 counts per million). The frequency-word-spotting latency function thus appears to be nonlinear. Fourteen of the 36 target words in the current experiments were in the low-frequency (≤ 13 counts per million) range, so frequency effects ought to have been detected. In fact, in Experiment 1, the mean RT for the low-frequency words was 727 ms and that for the other words (> 13 counts per million) was 703 ms. However, a one-way ANOVA with unequal n indicated that this difference was not reliable, $F(1, 34) < 1, MS_e = 9,404$. The very large effects of competition and stress pattern, and the fact that we did not manipulate word frequency explicitly, may have prevented us from detecting a reliable frequency effect.

It is also more difficult to spot target words in strings that are themselves word onsets than in strings that cannot be continued to form words. The stress effect was reliable by both speed and accuracy measures for both word and nonword onsets. The word-onset effect appeared in both speed and error rates for WS items, but only in errors for SW items.

The stress effect was predicted by the Metrical Segmentation Strategy. Responses are faster and more accurate to targets in WS items because lexical access should be initiated at the second strong syllable, segmenting the item at the appropriate location. Responses are slower and less accurate in SW items because there is no segmentation at the second weak syllable.

The reliable difference in word-spotting responses between targets in word onsets and targets in nonword onsets was predicted by SHORTLIST, consistent with the hypothesis that multiple word candidates are considered for any stretch of speech input. It is harder to detect *mess* in /dəməs/ than in /nəməs/ because in the former case, the word *domestic* is being entertained as a hypothesis. The interference between *mess* and *domestic* (or between *sack* and *sacrifice*) is evidence not only that multiple words have been activated but that they are actively competing. SHORTLIST also predicted that the competition effect would be larger in WS than in SW items, as found. This is because, in WS items, the embedding words are exerting maximal inhibition at the offset of the target words, whereas in SW items, the embedding words do not have their largest effect until later, so competition effects are slower to emerge (cf. Figures 1 and 2).

Experiment 2

The results of Experiment 1 suggest that activation and competition, as instantiated in the SHORTLIST model, and sensitivity to prosodic structure, as captured by the Metrical Segmentation Strategy, are all components of spoken word recognition. However, there is a potential problem with the methodology of Experiment 1. Responses to WS items were reliably faster and more accurate than responses to SW items. Perhaps these differences can be attributed to an attentional strategy. Listeners were required to detect target words that could appear at either the beginnings (as in SW items) or the ends (as in WS items) of the stimuli. It is possible that subjects preferred to attend to item ends, yielding faster and more accurate performance on the WS items.

In Experiment 2, therefore, target location was blocked. Listeners were asked to listen only for words either at the beginnings of the items or at the ends of the items. If the stress pattern effect in Experiment 1 was due to an item end preference, then forcing subjects to attend to one target location should remove the WS–SW difference. This manipulation also allowed us to address another issue. Are the competition effects mandatory? If listeners can attend to one target location, they may be able to ignore the other information in each item. If the first syllables of /dəməs/ and /nəməs/ are not processed, or are only processed very shallowly, *domestic* may not be activated in the former case, and no competition effect will emerge. On the other hand, hearing /dəməs/ may activate *domestic* in spite of any attentional focus. Competition effects with blocked target location would thus indicate that activation

and competition are mandatory features of spoken word recognition.

Method

Subjects. Sixty-three student volunteers, mainly from Clare College, Cambridge, United Kingdom, were paid for participating. There were 42 men and 21 women, aged between 18 and 35 years.

Materials and procedure. The three experimental tapes from Experiment 1 were used. The only change in procedure was a change in the instructions. Thirty-three of the subjects were told to monitor for words in initial position and to ignore words in final position; 30 subjects received the opposite instructions; they were instructed to monitor for words in final position and ignore those in initial position. These instructions meant that only half of the original target-bearing items now contained words to be detected, reducing the proportion of targets from 0.5 to 0.25.

Results and Discussion

As before, verbal responses were analyzed and missing responses were tallied. Failures to give a verbal response and occasions in which a word other than the target was produced were again discounted and treated as missing data. Three subjects who monitored for words in initial position detected 50% (9) or less of the targets. They were excluded from the analysis, leaving three groups of 10 subjects who heard each tape in this condition. All subjects who monitored for item-final words detected more than 50% of the targets. There were three groups of 10 subjects on each tape with these instructions.

It is worth noting that subjects were able to follow the instructions. Those subjects who were asked to detect words in initial position never produced a false alarm in which they detected a word in final position. In word-final detection, there were five false alarms (out of a possible 540) in which subjects responded with a word beginning at the initial position. However, these responses were all based on both syllables of the input string; they were not detections of targets in the first syllable alone: One response was *verbose* given the string /vəbəs/, and the other four responses were *clip* given the string /kəlip/. Clearly, subjects were able to attend to the appropriate target location and ignore the words embedded in the other location.

RTs were again adjusted so as to measure from target offset. Responses of less than 200 ms or greater than 1,800 ms were also treated as missing data. The mean RTs, measured from target offset, and the mean error rates are given in Table 2.

Several ANOVAs were again performed. As in the first experiment, the data were split into two subanalyses, one set focusing on the nonword-onset data alone (stress analysis) and the other set testing for competition and stress effects in a balanced design comparing the word-onset and matched nonword-onset data (competition analysis).

In the stress analysis, as in Experiment 1, there was a highly reliable stress effect. Responses to targets in WS items (e.g., /nəməs/ and /kləsæk/) were faster and more accurate than responses to targets in SW items (e.g., /sækrək/ and /mɛstəm/): for RT, $F_1(1, 54) = 85.31, p < .001, MS_e = 26,902$, and $F_2(1, 34) = 107.68, p < .001, MS_e = 7,942$; for errors, $F_1(1, 54) =$

Table 2
Mean Response Times (RTs; in Milliseconds) and Error Rates for Word Spotting of Targets in Weak-Strong (WS) and Strong-Weak (SW) Strings in Word Versus Nonword Onsets in Experiment 2

Stress pattern	Word onset	Nonword onset	
		Target matched	Target unmatched
WS			
RT	595	470	492
Error rate (%)	24	21	21
Example	/dəmes/	/nəmes/	/kləsæk/
SW			
RT	772	759	756
Error rate (%)	39	37	27
Example	/sækrəf/	/sækrək/	/məstəm/

Note. Mean RTs were measured from target offset.

8.00, $p < .01$, $MS_e = 0.0440$, and $F_2(1, 34) = 5.64$, $p < .05$, $MS_e = 0.0375$. There was no significant difference in RT between the two types of nonword onset (/nəmes/ vs. /kləsæk/ and /sækrək/ vs. /məstəm/; F_1 and $F_2 < 1$). This effect did not interact with the stress effect. In errors, there were no reliable differences between the two types of nonword onset, although in the subject analysis the effect approached significance: $F_1(1, 54) = 3.39$, $.1 < p < .05$, $MS_e = 0.0197$; $F_2 < 1$, $MS_e = 0.0617$. This effect appeared to interact with the stress effect, but this was again marginally reliable by subjects only: $F_1(1, 54) = 3.39$, $.1 > p > .05$, $MS_e = 0.0197$; $F_2(1, 34) = 1.07$, $p > .1$, $MS_e = 0.0375$. There was clearly no effect in the WS items (0% difference, on average). The mean difference of 10% in the SW items (responses to /sækrək/ less accurate than responses to /məstəm/) was examined with t tests; it was marginally significant by subjects only: $t_1(29) = 1.83$, $.1 > p > .05$; $t_2(17) = 1.00$, $p > 0.1$.

Next, the competition analysis was performed. Again, there was a reliable stress pattern effect in RT, $F_1(1, 54) = 54.22$, $p < .001$, $MS_e = 30,048$, and $F_2(1, 34) = 18.24$, $p < .001$, $MS_e = 32,087$, and in errors, $F_1(1, 54) = 12.03$, $p < .005$, $MS_e = 0.0582$, and $F_2(1, 34) = 4.13$, $p < .05$, $MS_e = 0.1016$. Subjects were faster and more accurate at word spotting in WS items (e.g., /dəmes/ and /nəmes/) than in SW items (e.g., /sækrəf/ and /sækrək/). There were no other significant effects in the error analyses (no differences between word onsets and nonword onsets and no interaction of stress pattern with word-nonword onset; all $F_s < 1$). However, speed of response to targets in word onsets was slower ($M = 684$ ms) than to targets in items that could not be continued to form longer words ($M = 615$ ms): $F_1(1, 54) = 17.70$, $p < .001$, $MS_e = 8,125$; $F_2(1, 34) = 10.09$, $p < .005$, $MS_e = 8,777$. This effect was not equivalent across stress pattern: $F_1(1, 54) = 11.59$, $p < .005$, $MS_e = 8,125$; $F_2(1, 34) = 9.34$, $p < .005$, $MS_e = 8,777$. In the WS items, responses to targets in word onsets, like *mess* in /dəmes/, were reliably slower than responses to targets in nonword onsets (*mess* in /nəmes/): by subjects, $t_1(29) = 5.05$, $p < .001$; by items, $t_2(17) = 5.56$, $p < .001$. For the SW items, however, the difference in speed of detecting, for example, *sack* in /sækrəf/ and /sækrək/, was not significant.

Correlational analyses were also performed. As in Experiment 1, neither speed nor accuracy measures correlated either with the frequency of occurrence of the target words or with

the frequency of occurrence of the embedding (word-onset) words.

An explicit comparison of the results of Experiments 1 and 2 was undertaken (treating the WS-SW comparison in Experiment 1 as between subjects). In the stress analysis, there was a main effect of experiment both in RT, $F_1(1, 108) = 13.71$, $p < .001$, $MS_e = 31,536$, and $F_2(1, 34) = 33.58$, $p < .001$, $MS_e = 6,985$, and in errors, $F_1(1, 108) = 9.82$, $p < .005$, $MS_e = 0.0438$; $F_2(1, 34) = 13.24$, $p < .001$, $MS_e = 0.0195$. In RT, the experiment variable did not interact with the stress-pattern variable, suggesting that although subjects were faster with blocked target position, the size of the WS advantage was unaffected by this increase in speed. However, in the error data, the experiment variable did interact with the stress variable, although this was only significant by items: $F_1(1, 108) = 3.23$, $.1 > p > .05$, $MS_e = 0.0438$; $F_2(1, 34) = 6.86$, $p < .05$, $MS_e = 0.0124$. The difference in error rate between WS and SW items was somewhat larger in Experiment 1.

In the competition analysis, there was again a main effect of experiment, with responses faster and more accurate in Experiment 2 than in Experiment 1: for RT, $F_1(1, 108) = 11.60$, $p < .001$, $MS_e = 32,175$, and $F_2(1, 34) = 24.76$, $p < .001$, $MS_e = 9,041$; for errors, $F_1(1, 108) = 18.18$, $p < .001$, $MS_e = 0.0527$, and $F_2(1, 34) = 24.39$, $p < .001$, $MS_e = 0.0236$. In the RT analyses, the experiment variable did not interact with any other variable. In the error analyses, however, experiment interacted with the competition effect: $F_1(1, 108) = 10.27$, $p < .005$, $MS_e = 0.0228$; $F_2(1, 34) = 10.21$, $p < .005$, $MS_e = 0.0138$. The differences in accuracy of word spotting between targets in word onsets and in nonword onsets were present only in Experiment 1.

In this experiment, we have replicated the stress-pattern effect found in Experiment 1. Subjects were faster and more accurate spotting words in WS strings than in SW strings, even when they could attend to a prespecified target location. This result shows that the stress effect found in Experiment 1 cannot be attributed to a strategy of attending to the ends of items. Instead, the stress effects found in both experiments support the prediction of the Metrical Segmentation Strategy. Targets are detected more readily in WS strings because these strings are segmented at the onset of the target word (i.e., at the onset of the strong syllable). Targets are detected with greater difficulty in SW strings because these strings are not segmented at the offset of the target word (i.e., at the onset of the weak syllable).

The competition results of Experiment 2, however, differ in interesting and informative ways from those of Experiment 1. For WS strings, there were competition effects in both experiments. In Experiment 1, there were competition effects in speed and accuracy; in Experiment 2, there was only an RT effect. In contrast, for SW strings, there was evidence of competition only in Experiment 1, and then only in error rates. In Experiment 2, in which subjects could attend specifically to item onsets, no differences emerged between items like /sækrəf/ and /sækrək/. Although these results appear to be problematic for competition models, they are in fact predicted by SHORTLIST.

There are two findings in the competition results that need to be explained. The first is that the competition effects are larger in WS than in SW strings; the second is that they were

absent for SW strings in Experiment 2. As discussed in Experiment 1, SHORTLIST predicts the first finding because the inhibitory effect of the embedding words comes into play earlier, relative to target offset, in WS strings than in SW strings. The fact that the competition effect emerges slowly in SW strings provides an account of the second finding. The explanation of both rests on determining which part of the activation function should be used to predict performance in the word-spotting task. For WS strings, SHORTLIST only predicts a competition effect (e.g., greater activation for *mess* in /nəməs/ than /dəməs/) in a relatively narrow time window, centered around the final consonant of the target (C in Figure 1). At this point, activation of the embedding word (e.g., *domestic* given /dəməs/) is already high, inhibiting the activation of the target word (e.g., *mess*) and producing a large competition effect.

Near the final consonants of targets in SW strings (C in Figure 2), SHORTLIST predicts no difference in detection of targets in word onsets and matched nonword onsets (e.g., *sack* in /sækɹəf/ and /sækɹək/). Activation of *sacrifice* is equivalent in these two cases at this point. Differential activation of *sacrifice*, and the consequent changes in the activation level of *sack*, do not appear until later. If we note that subjects were faster and more accurate in Experiment 2, a simple explanation of the data pattern emerges. When subjects could attend to targets located in the first syllable, and respond rapidly, competition effects between the word-onset and matched nonword-onset strings were not found. In Experiment 1, in which targets could appear in either location, subjects responded more slowly: late enough for small differential competition effects to have emerged. Note that it is only much later (at C + 2 and C + 3) that SHORTLIST predicts that the competition effect (the difference in target activation between word onsets and matched nonword onsets) should be larger in SW than in WS strings.

In Experiment 1, performance on the matched SW nonword onsets (/sækɹək/) was equivalent to that on unmatched SW nonword onsets (/məstəm/), whereas in Experiment 2 there was a tendency for performance on the unmatched items to be more accurate (by 10%, on average). The unmatched items tended to become nonwords earlier than the final phoneme, whereas the matched items had nonword points on their final segments. As Figure 2 shows, in SHORTLIST, late in time, the activation of targets in matched and unmatched items is equivalent (after the final /k/ of /sækɹək/, e.g., has lowered the activation of *sacrifice*). Earlier, however, there is an advantage for targets in unmatched items. As predicted, when differences in target detection between SW word and matched nonword onsets were larger, those differences between matched and unmatched nonword onsets are smaller, and vice versa. The differences between SW nonword items across experiments is thus also consistent with SHORTLIST.

In summary then, SHORTLIST predicts the pattern of competition effects found over Experiments 1 and 2. In WS strings, the longer, embedding word (e.g., *domestic*) begins earlier in time than the embedded word (*mess*). The longer word will have established a degree of activation before there is any evidence for the target word. Near the final phoneme, target activation is strongly suppressed by the high activation of the longer word, producing a large competition effect. The

model thus predicts a large competition effect for targets in WS strings. Specification of item-final target location was not sufficient to remove this effect. In contrast, the activation of a target word in an SW string (e.g., *sack*) is not suppressed by that of the longer, embedding word (*sacrifice*) until after the final phoneme of the target, producing a small competition effect. SHORTLIST thus predicts smaller competition effects for targets in SW than in WS strings. With specification of item-initial target location, word spotting was fast enough for this competition effect to go undetected (Experiment 2). In the absence of a location cue, a small competition effect was found (Experiment 1).

Experiment 3

The results of Experiment 2 indicate that the main effect of stress pattern is resistant to attentional biases. Hence, the stress effect in Experiment 1 was not due to a strategy of attending preferentially to item-final targets. The advantage in both speed and accuracy of responses to targets in WS strings over responses to targets in SW strings, as predicted by the Metrical Segmentation Strategy, appears to be robust.

There is, however, another potential problem with the stress-pattern effect. As noted earlier, the targets in the WS strings were considerably longer than those in the SW strings. The differences in performance on these two types of string may therefore be due to this length confound. The targets in WS strings could have been detected more rapidly and more accurately because they were longer than those in SW strings.

Experiment 3 was designed to address this issue. We again used the word-spotting task, and in fact used exactly the same materials as those used in Experiments 1 and 2. We attempted to equate the target words for length with a speech compression algorithm. We compressed the WS strings and expanded the SW strings. If the differences we have reported are due to the length confound, there should be no differences in the speed or accuracy of detection of words of the same length in WS and SW strings. If, however, the differences are due to the Metrical Segmentation Strategy, they should emerge in spite of the length control.

Method

Subjects. Thirty-two student volunteers, mainly from St. Catharine's College, Cambridge, United Kingdom, were paid for participating. There were 22 men and 10 women, aged between 18 and 24 years. Two subjects had to be excluded as a result of experimenter error, leaving 10 subjects in each group.

Materials and construction. The materials from Experiments 1 and 2 were used. Each bisyllabic item was digitized, sampling at 22.05 kHz with 16-bit resolution. Items were then compressed or expanded with a compression algorithm (Charpentier, 1988).

This algorithm uses pitch-period extraction and averaging. It produces a much higher quality of speech than earlier techniques using deletion or addition of speech samples. For voiced portions of speech, individual pitch periods are averaged together. For unvoiced portions, individual digital sample points are averaged. Compression occurs when averaged pitch periods (or sample points) replace original periods (samples) and expansion occurs when the averages are added between the originals. Compression or expansion rate is determined by the frequency of averaging. The output is speech that is significantly

shorter or longer than the original but that still retains the major features of the uncompressed version, such as speaker identity and pitch. The signals sound as if they have been spoken either rather quickly or rather slowly, but there is little loss in quality or intelligibility (except at very high rates of compression; Altmann and Young, 1993).

WS strings were compressed and SW strings were expanded. The measured lengths of each target word were used to compute the compression and expansion rates for that word. Recall that each word appeared in three contexts, two with one stress pattern and one with the opposite stress pattern. For those targets that appeared twice in WS strings, an average target length halfway between the mean of the two WS measurements and the one SW measurement was calculated for each target. For those that appeared twice in SW strings, an average of the mean of the two SW measurements and the one WS measurement was calculated for each target. These mean lengths could then be used to compute compression or expansion ratios that would make each occurrence of a target the same mean length if the algorithm operated on absolute duration. However, because the algorithm uses pitch periods, the length resulting from a given compression depends on both the compression ratio and the pitch of the input speech. To guarantee that the lengths of the targets were at least equated, or, if anything, that the lengths of the targets in the SW strings were greater than those in the WS strings, we therefore compressed or expanded at seven-sixths the rate calculated on the basis of mean length.

Each complete item, not just each target word, was compressed or expanded. In 12 of the 54 SW items this resulted in an unnaturally long, weak second syllable. Because duration is a strong correlate of stress, these sounded more like strong-strong (SS) than SW items. For these items, we spliced the original weak second syllable (schwa plus final consonant) onto the expanded target word. Splices were made at the zero-crossing at the onset of the first pitch period of the schwa. Because the algorithm operates at the level of pitch periods, preserving transitional information, the splices were undetectable.

The mean rate of compression, for the targets in WS strings, was 73%, resulting in a mean target length of 294 ms (note that it was 440 ms in the original, uncompressed materials). For targets in SW strings, the mean expansion rate was correspondingly 127%, resulting in a mean length of 333 ms (compared with 236 ms in the originals). The length difference in the new materials was therefore reversed, with targets in the new SW strings 39 ms longer, on average, than those in the new WS strings, compared with a difference of 204 ms in the opposite direction in the original materials. Two ANOVAs tested these length differences. The first ANOVA compared the lengths of the targets in nonword onsets. There was a significant difference between targets in WS strings (290 ms, on average) and targets in SW strings (334 ms, on average), $F(1, 34) = 50.04, p < .001, MS_e = 711$. The compression and expansion, for these items, had therefore reversed the length difference. The second ANOVA compared target lengths in word onsets and matched nonword onsets. Here the difference was again reversed (targets in SW items longer than those in WS items), but was slightly smaller (WS, 301 ms; SW, 327 ms). The difference was not quite significant, $F(1, 34) = 2.95, .05 < p < .1, MS_e = 4,374$.

All of the fillers from the earlier experiments (including the practice items) were also compressed or expanded. All WS fillers were compressed at the average rate (73%) and all SW fillers were expanded at the average rate (127%).

After compression or expansion of each target-bearing item, timing pulses were aligned with the onset of each target word with a speech editor. At output, the items were upsampled from 22.05 kHz to 44.1 kHz and then recorded onto the left channel of a digital audiotape at a rate of one item every 3 s. The timing pulses were recorded onto the right channel.

As in Experiments 1 and 2, three experimental tapes were constructed. These tapes had exactly the same running order of items as

before. The experiment was therefore an exact analogue of Experiment 1, but for the length changes in the items.

Procedure. The procedure was identical to that in Experiment 1, except for a small change in the instructions given to subjects. They were warned that the items would appear to be said either rather quickly or rather slowly, and they were told to try to ignore these changes in speed.

Results and Discussion

Verbal responses were again analyzed and missing responses were counted. When a subject failed to give the target word as a verbal response but pressed the button, that manual response was discounted and treated as missing data. As shown in Table 3, subjects found this experiment much more difficult than the earlier two. Only 12 subjects managed to detect more than 50% (18) of the targets. It therefore seemed inappropriate to use the previous rejection criterion, which required subjects to detect more than half of the targets. RTs were adjusted for measurement from target offset, with those falling outside a window of 200–1,800 ms treated as missing data. The mean RTs, measured from target offset and the mean error rates are given in Table 3.

As in Experiments 1 and 2, stress and competition analyses were carried out separately. In the stress analysis, there was a main effect of stress pattern: for RT by subjects, $F_1(1, 27) = 14.97, p < .001, MS_e = 5,278$; for RT by items, $F_2(1, 34) = 11.05, p < .005, MS_e = 2,743$; and for errors, $F_1(1, 27) = 122.43, p < .001, MS_e = 0.0360$, and $F_2(1, 34) = 46.96, p < .001, MS_e = 0.0563$. Word spotting in WS items (e.g., /nəməs/ and /kləsæk/) was, on average, 51 ms faster and 39% more accurate than in SW items (e.g., /sækræk/ and /mɛstəm/). There were no other significant effects in these analyses.

In the competition analysis, there was no stress effect in the RT data (F_1 and $F_2 < 1$), but a highly significant effect in the errors. Target detection was an average of 33% more accurate in WS items (e.g., /dəməs/ and /nəməs/) than in SW items (e.g., /sækræk/ and /sækræk/): by subjects, $F_1(1, 27) = 77.43, p < .001, MS_e = 0.0445$; by items, $F_2(1, 34) = 19.36, p < .001, MS_e = 0.1068$. On the other hand, there was no competition effect in the errors (F_1 and $F_2 < 1$), but there was one, significant by subjects only, in the RTs: $F_1(1, 27) = 8.13, p < .01, MS_e = 4,934$; $F_2(1, 34) = 2.36, p > .1, MS_e = 6,332$.

Table 3
Mean Response Times (RTs; in Milliseconds) and Error Rates for Word Spotting of Targets in Weak-Strong (WS) and Strong-Weak (SW) Strings in Word Versus Nonword Onsets in Experiment 3

Stress pattern	Word onset	Nonword onset	
		Target matched	Target unmatched
WS			
RT	759	685	672
Error rate (%)	46	39	38
Example	/dəməs/	/nəməs/	/kləsæk/
SW			
RT	727	728	733
Error rate (%)	73	79	74
Example	/sækræk/	/sækræk/	/mɛstəm/

Note. Mean RTs were measured from target offset.

There was a significant interaction of the stress and competition effects for RT by subjects, $F_1(1, 27) = 7.34, p < .05, MS_e = 5,538$; for RT by items, $F_2(1, 34) = 5.88, p < .05, MS_e = 6332$; and for errors, $F_1(1, 27) = 6.99, p < .05, MS_e = 0.0191$, and $F_2(1, 34) = 2.84, p = .10, MS_e = 0.0282$. These interactions were examined with t tests.

In RT, the competition effect was due entirely to the WS items (74 ms difference, on average): by subjects, $t_1(29) = 2.96, p < .01$; by items, $t_2(17) = 2.21, p < .05$. Responses to targets in WS word onsets (/dæməs/) were reliably slower than those to the same targets in WS nonword onsets (/nəməs/). The average difference of 1 ms between responses to targets in SW items that were or were not word onsets (/sækɹəf/ and /sækɹək/) was not significant. Responses to targets in WS nonword onsets (/nəməs/) were reliably faster than those to targets in SW nonword onsets (/sækɹək/) only in the subject analysis (mean difference of 43 ms): By subjects, $t_1(29) = 2.43, p < .05$; by items, $t_2(34) = 1.36, p > .1$. The inverse effect (responses to targets in WS word onsets such as /dæməs/ were 32 ms slower, on average, than those to targets in SW word onsets such as /sækɹəf/) was not significant by subjects or items.

In errors, t tests showed that word spotting was reliably more accurate (27%, on average) in WS word onsets (/dæməs/) than in SW word onsets (/sækɹəf/): by subjects, $t_1(29) = 4.80, p < .001$; by items, $t_2(34) = 3.11, p < .005$. Similarly, targets in WS nonword onsets (/nəməs/) were detected more accurately (40%, on average) than those in SW nonword onsets (/sækɹək/); by subjects, $t(29) = 9.02, p < .001$; by items, $t(34) = 4.73, p < .001$. There were no significant differences in error rates in the pairwise comparisons of word and nonword onsets with the same stress patterns.

Finally, the stress effects in Experiments 1 and 3 were compared. In the stress analysis, there was a highly significant stress effect in both RT and errors: for RT, $F_1(1, 54) = 117.71, p < .001, MS_e = 6,847$, and $F_2(1, 34) = 47.71, p < .001, MS_e = 11,224$; for errors, $F_1(1, 54) = 148.38, p < .001, MS_e = 0.0351$, and $F_2(1, 34) = 49.47, p < .001, MS_e = 0.0631$. There were also interactions of this effect with the experiment variable in both RT and errors. In RT, this interaction indicated that the stress effect was smaller in Experiment 3 than in Experiment 1; $F_1(1, 54) = 36.54, p < .001, MS_e = 6,847$; $F_2(1, 34) = 48.03, p < .001, MS_e = 4,669$. In errors, the interaction indicated that the stress effect was larger in Experiment 3 than in Experiment 1: $F_1(1, 54) = 13.52, p < .001, MS_e = 0.0351$; $F_2(1, 34) = 7.68, p < .005, MS_e = 0.0370$.

A very similar pattern was obtained in the competition analysis. There was again a significant stress effect in both RT and errors: for RT, $F_1(1, 54) = 43.74, p < .001, MS_e = 4,979$, and $F_2(1, 34) = 7.26, p < .05, MS_e = 44,475$; for errors, $F_1(1, 54) = 80.82, p < .001, MS_e = 0.0464$, and $F_2(1, 34) = 11.55, p < .005, MS_e = 0.1948$. The interactions again indicated that the RT stress effect was smaller in Experiment 3 than in Experiment 1— $F_1(1, 54) = 36.26, p < .001, MS_e = 4,979$; $F_2(1, 34) = 30.24, p < .001, MS_e = 5,189$ —and that the error stress effect was larger in Experiment 3 than in Experiment 1, although this was not quite significant by items, $F_1(1, 54) = 10.22, p < .005, MS_e = 0.0464$; $F_2(1, 34) = 3.83, .1 > p > .05, MS_e = 0.0743$.

The results of Experiment 3 can now be summarized. As in Experiments 1 and 2, there was a significant competition effect in the RTs to spot words in WS strings. Subjects were slower to detect targets in word onsets (e.g., *mess* in /dæməs/) than in nonword onsets (e.g., *mess* in /nəməs/ and *sack* in /kləsæk/). Subjects were also less accurate in word onsets than in nonword onsets, but this small effect was not significant. Thus, for WS strings, the data replicate those of Experiment 2. The same is true for the SW strings: There were no competition effects in speed or accuracy, just as in Experiment 2. The absence of any competition effects in SW items in that experiment appeared to be due to the increased speed relative to Experiment 1 (it was only when subjects were going slowly that the differential effect of the final phonemes [e.g., of /sækɹəf/ and /sækɹək/] could be detected). Responses to the SW items are about 30 ms faster here than in Experiment 2. It thus appears that on the relatively few occasions that subjects were able to detect targets in SW strings, they were able to spot them before the following context (continuing or not continuing as a possible word) could act to produce a differential competition effect. On the other hand, the competition effect in WS strings is very robust, resisting attentional focus on target location (Experiment 2) and a 73% compression rate (Experiment 3).

Most important, the stress effect survived the compression–expansion manipulation. It was highly reliable in the error data for both word onsets and nonword onsets and reliable in the RT data for nonword onsets. Spotting words in SW strings was more difficult than spotting words in WS strings even when the targets in the SW items were longer than those in the WS items. The results of Experiments 1 and 2 cannot therefore be attributed solely to a length confound. Only the added size of the stress effect in Experiment 1 as compared with Experiment 3 can be the result of length differences.

The stress-pattern differences found in all three experiments thus appear to be due, primarily, to the operation of the Metrical Segmentation Strategy. A WS bisyllable is segmented by the Metrical Segmentation Strategy at the onset of the strong syllable, hence at the onset of the target word, making it easier to detect the target. An SW bisyllable is not segmented by the Metrical Segmentation Strategy. It is harder to detect a target in an SW string because no segmentation position is postulated in such strings.

In summary, the results of Experiment 3 indicate that the stress-pattern effects found in Experiments 1 and 2 are not due entirely to effects of target length. Even when targets in SW strings are somewhat longer than those in WS strings, word spotting is slower and less accurate for the SW strings. This is as predicted by the Metrical Segmentation Strategy.

General Discussion

The three experiments have separately tested the effects in spoken word recognition of lexical competition (as instantiated in SHORTLIST) and prosodically guided explicit segmentation (as instantiated in the Metrical Segmentation Strategy). Clear support is provided for both.

The results of all the experiments strongly support the competition predictions of SHORTLIST. Multiple lexical

candidates appear to be considered as hypotheses for what words a piece of speech contains. Target words embedded in bisyllabic strings are more difficult to spot when the strings are themselves the beginning of longer words. In WS strings, this effect emerged even when listeners knew the location of the target word (Experiment 2), and after compression (Experiment 3). In SW strings, however, the effect was only present in Experiment 1. When subjects responded more rapidly to SW strings (Experiments 2 and 3), a competition effect was not detected. This pattern of results is in line with the predictions of the SHORTLIST model: Competition effects are larger for WS than SW strings at the offset of the target word (see Figures 1 and 2) because the longer, embedding word has the advantage of starting earlier than the shorter target word in the WS case. In the SW case, the two words begin at the same time. Earlier activation produces more inhibition sooner and hence larger competition effects at the offsets of target words in WS strings.

The results of all the experiments are also exactly as predicted by the Metrical Segmentation Strategy. Word detection responses are both faster and more accurate in WS than in SW strings. This is true even when subjects can focus on target location (Experiment 2) and when the length of the target words has been controlled (Experiment 3). As predicted by the Metrical Segmentation Strategy, WS items are at an advantage because they are segmented at the onset of their second (strong) syllables, unlike SW items, which are not segmented at the onset of their second (weak) syllables. Previous evidence in favor of prosodically based segmentation has been explained in terms of the efficiency of this strategy: The distribution of words in English speech is such that assuming strong syllables to be word initial will be a very good bet. The present new variant of the word-spotting task thus bolsters the previous evidence from initial word-spotting and juncture misperceptions.

The evidence for lexical competition provided by the word-spotting task, in turn, joins that which is obtained from other tasks, such as cross-modal priming and perceptual identification. Using cross-modal semantic priming, Zwitserlood (1989) has shown that words that begin in the same way are activated when the input is consistent with those candidates, but that there is rapid selection of the appropriate candidate when the input becomes consistent with only that word. For example, with the pair *kapitein* and *kapitaal*, lexical decision was speeded on semantic relatives of both words for visual probes presented up to the /t/ of either word. Priming was obtained only for relatives of the appropriate word for probes aligned with the following vowel. Similar evidence for the activation of multiple entries that begin in the same way, followed by rapid selection of the appropriate candidate, was reported by Marslen-Wilson (1987, 1990).

These results are all for words that begin in the same way, sharing the same initial phonemes. The previously existing evidence was weaker for activation of words that begin at different points in time. Swinney (1981) reported that when subjects heard *boycott*, they were facilitated on a semantic relative of *boy* (presented halfway through *boycott*) but not on a relative of *cot* (presented at the offset of *boycott*). Shillcock (1990), however, has demonstrated a priming effect of embed-

ded words, such as *bone* in *trombone*: Lexical decisions to probes like *rib*, presented at the offset of *trombone*, were speeded relative to those to unrelated control probes.

Other evidence for activation and competition has appeared in a number of other tasks, including perceptual identification, auditory repetition (naming), and lexical decision (Cluff & Luce, 1990; Goldinger, Luce, & Pisoni, 1989; Goldinger, Luce, Pisoni, & Marcario, 1992; Luce, 1986b; Luce et al., 1990; Taft, 1986). The effects of frequency and of similarity neighborhoods, for instance, are interdependent. High-frequency words in sparse neighborhoods of low-frequency words are recognized rapidly and accurately, whereas low-frequency words with many high-frequency neighbors are recognized with difficulty. Luce et al. (1990) have captured these results in a model of spoken word recognition (the Neighborhood Activation Model, or NAM), which assumes that multiple candidate words are activated on the basis of the sensory input. Word recognition is based on frequency-weighted neighborhood probability values computed for each activated word. Recognition occurs when a word decision unit reaches a criterion. Thus, although the candidate words do not directly compete through inhibition, as in SHORTLIST, selection is nevertheless based on competition because it takes into account the evidence supporting each candidate.

The research that has supported the NAM is therefore also consistent with competition models such as SHORTLIST. Most of this work, however, like that using the cross-modal priming task, has focused on monosyllabic words, and thus on predictions of activation and competition in which candidates share initial portions. One exception is a perceptual identification study reported in Cluff and Luce (1990). Here, listeners were asked to identify spondees (bisyllables with two strong syllables), such as *bucksaw*, which were presented in white noise. Some syllables were designated as easy (high-frequency words with sparse, low-frequency neighborhoods) and others as hard (low-frequency words with dense, high-frequency neighborhoods). Several neighborhood effects were detected. For example, listeners found it easier to identify a spondee with a hard first syllable when the second syllable was easy than when it too was hard. The results suggest that shorter words embedded in longer words (and the short words' neighbors) are all accessed during the recognition of longer words.

In summary, there now appears to be a substantial body of evidence in favor of interword competition in spoken word recognition. Thus, the present findings complement those in the previous literature. However, our findings also extend the previous evidence for competition. First, most of the previous evidence addressed competition between words beginning in the same way. It appears now to be strongly attested that when the onset of a spoken word is heard, all the words that begin in that way are treated as hypotheses by the recognition system. The previous evidence was rather less forthcoming, however, on the question of words embedded in other words or, more generally, of all possible words beginning at all possible locations in a string of speech sounds. The present findings provide clear-cut evidence of this type of lexical competition. It now appears certain that all words beginning at any point in an incoming speech stream are, at least momentarily, considered as potential candidates for word recognition.

Second, our results indicate that word recognition entails both activation and competition. They show not only that multiple candidates are activated but that these candidates then compete for recognition. Many of the previous results, such as those from the cross-modal priming task, have indicated only that multiple words are activated during recognition. The interference found here between, for example, *mess* and *domestic* indicates that these lexical hypotheses actually compete with each other.

Third, the present findings also extend the methodological range of the evidence for lexical competition. There are problems inherent in all laboratory tasks, including those that have been used to support claims for competition. The cross-modal priming task, for instance, rests on assumptions about the mechanism of priming—that activation of a candidate word will be sufficient to induce a priming effect and that the emergence and disappearance of priming directly reflects the rising and falling of the activation level of the candidate word. Neither assumption has been clearly supported by experimental evidence. The perceptual identification task, in turn, allows the possibility of responses based on guessing in cases in which a degraded signal has in fact been misperceived. Finally, the lexical-decision task may also involve postperceptual decision processes, so that here too competition effects in lexical decision could be the result of postperceptual strategies rather than the normal mechanisms of perception.

The word-spotting task as used in the present study is not subject to these problems. The predictions in word spotting emerge directly from the assumptions of the models under test and involve no assumptions about mechanisms of priming. The competition effects found in the word-spotting task likewise appear to be free of postperceptual influence and thus due to mandatory perceptual processing alone; they emerge even when subjects can attend to target locations. Furthermore, word spotting has some ecological validity in that it requires subjects to identify real words in unsegmented speech, just as in normal recognition.

We conclude that activation and competition of lexical candidates are genuine properties of the speech recognition system. In continuous speech recognition, a parse of the speech stream into a stream of words is achieved through a process of competition between word candidates. We conclude also that the recognition system is accorded added efficiency by sensitivity to the prosodic structure of the language. The stress-pattern effects found in the present study suggest that the competition process operates in conjunction with an explicit segmentation process based on prosodic structure.

There is no conflict between the predictions from the two models we have tested. Although the Metrical Segmentation Strategy suggests where word boundaries are likely to be, it entails no claims about the mechanism of parsing, that is, about ways in which lexical candidates are selected after they have been accessed. In contrast, SHORTLIST makes no assumptions about where word boundaries might occur, but does specify that a parse is achieved by competition. The two models are therefore compatible and can be jointly incorporated into a single overall model of spoken word recognition.

One way in which this could be achieved is for the Metrical Segmentation Strategy to be incorporated in SHORTLIST as a

bias in the competition process. In this instantiation, the Metrical Segmentation Strategy would not determine where lexical access would occur. Instead, it would act to make words beginning at probable onset positions (strong syllables) more likely to be recognized. Given a strong syllable in the input, candidate content words beginning at that syllable could receive a higher level of activation than candidates not beginning at that location. Success in the competition process would thus be more likely for words beginning at strong syllables than for words straddling these points. (Lexical entries in the model would of course have to be more sophisticated: They would have to include information about syllabification and stress patterns and about grammatical class.) A version of SHORTLIST incorporating the Metrical Segmentation Strategy would predict the main effect of stress pattern found in both experiments. Targets beginning at the strong second syllable in WS strings would have their activation levels boosted, making a word boundary more plausible at this location. In SW strings, words beginning at the second weak syllable would not be advantaged. A juncture between a strong syllable and a following weak syllable would therefore be less likely than one between a weak syllable and a following strong syllable. Aside from any competition effects, then, target detection would be predicted to be easier in WS than SW strings—exactly as observed.

SHORTLIST and the Metrical Segmentation Strategy together provide a powerful account of spoken word recognition, realistically adjusted to the structure of the vocabulary and the distributional occurrence of words in speech.

References

- Altmann, G. T. M., & Young, D. (1993). Factors affecting adaptation to time-compressed speech. In *Proceedings of EUROSPEECH '93* (Vol. 1, pp. 333–336). Berlin, Germany.
- Bard, E. G. (1990). Competition, lateral inhibition, and frequency: Comments on the chapters of Frauenfelder and Peeters, Marslen-Wilson, and others. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 185–210). Cambridge, MA: MIT Press.
- Bard, E. G., Shillcock, R. C., & Altmann, G. T. M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception and Psychophysics*, *44*, 395–408.
- Charpentier, F. (1988). *Traitement de la parole par analyse-synthese de Fourier application à la synthèse par diphtones*. [Speech processing by analysis/synthesis of the Fourier transformation of diphone synthesis]. Lannion, France: Centre National d'Etude des Télécommunications.
- Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 551–563.
- Cole, R. A., & Jakimik, J. (1978). Understanding speech: How words are heard. In G. Underwood (Ed.), *Strategies of information processing* (pp. 67–116). San Diego, CA: Academic Press.
- Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–163). Hillsdale, NJ: Erlbaum.
- Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*, 218–236.

- Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133–142.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25, 385–400.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, 24, 381–410.
- Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113–121.
- Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- Frauenfelder, U. H. (1991). Lexical alignment and activation in spoken word recognition. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech and brain* (Wenner-Gren International Symposium Series; pp. 294–303). London: Macmillan.
- Frauenfelder, U. H., & Peeters, G. (1990). Lexical segmentation in TRACE: An exercise in simulation. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 50–86). Cambridge, MA: MIT Press.
- Frauenfelder, U. H., & Peeters, G. (1992). *Simulating the time-course of word recognition: An analysis of lexical competition in TRACE*. Unpublished manuscript.
- Freedman, D. (1992). *Frequency effects in auditory word recognition using the word-spotting task*. Unpublished undergraduate project, Department of Experimental Psychology, University of Cambridge, Cambridge, United Kingdom.
- Fudge, E. C. (1987). Branching structure within the syllable. *Journal of Linguistics*, 23, 359–377.
- Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, 28, 501–518.
- Goldinger, S. D., Luce, P. A., Pisoni, D. B., & Marcario, J. K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1211–1238.
- Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception and Psychophysics*, 38, 299–310.
- Kahn, D. (1976). *Syllable-based generalizations in English phonology*. Bloomington, IN: Indiana University Linguistics Club.
- Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51, 2018–2024.
- Luce, P. A. (1986a). A computational analysis of uniqueness points in auditory word recognition. *Perception and Psychophysics*, 39, 155–158.
- Luce, P. A. (1986b). Neighborhoods of words in the mental lexicon. In *Research on speech perception* (Technical Rep. No. 6). Indiana, IN: Indiana University, Speech Research Laboratory, Department of Psychology.
- 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. 122–147). Cambridge, MA: MIT Press.
- Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, 244, 522–523.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25, 71–102.
- Marslen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 148–172). Cambridge, MA: MIT Press.
- 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.
- McQueen, J. M., & Cutler, A. (1992). Words within words: Lexical statistics and lexical access. In *Proceedings of the International Conference on Spoken Language Processing* (Vol. 1, pp. 221–224). Banff, Alberta, Canada: University of Alberta.
- Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, 62, 714–719.
- Norris, D. G. (1982). Autonomous processes in comprehension: A reply to Marslen-Wilson and Tyler. *Cognition*, 11, 97–101.
- Norris, D. G. (1990). A dynamic-net model of human speech recognition. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 87–104). Cambridge, MA: MIT Press.
- Norris, D. G. (1992). Connectionism: A new breed of bottom-up model? In R. G. Reilly & N. E. Sharkey (Eds.), *Connectionist approaches to natural language processing* (pp. 351–371). Hillsdale, NJ: Erlbaum.
- Norris, D. G. (in press). SHORTLIST: A connectionist model of continuous speech recognition. *Cognition*.
- Norris, D. G., & Cutler, A. (1985). Juncture detection. *Linguistics*, 23, 689–705.
- Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32, 258–278.
- Procter, P. (Ed.). (1975). *Longman dictionary of contemporary English*. London: Longman.
- Pulgram, E. (1970). *Syllable, word, nexus, cursus*. The Hague, The Netherlands: Mouton.
- Selkirk, E. O. (1982). The syllable. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations—Part II* (pp. 337–383). Dordrecht, the Netherlands: Foris.
- Shillcock, R. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 24–49). Cambridge, MA: MIT Press.
- Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645–659.
- Swinney, D. (1981). Lexical processing during sentence comprehension: Effects of higher order constraints and implications for representation. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech* (pp. 201–209). Amsterdam, the Netherlands, North-Holland.
- Taft, M. (1986). Lexical access codes in visual and auditory word recognition. *Language and Cognitive Processes*, 1, 297–308.
- Treiman, R., Gross, J., & Cwikiel-Glavin, A. (1992). The syllabification of /s/ clusters in English. *Journal of Phonetics*, 20, 383–402.
- Treiman, R., & Zukowski, A. (1990). Toward an understanding of English syllabification. *Journal of Memory and Language*, 29, 66–85.
- Zwitserslood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, 32, 25–64.

Appendix

Experimental Materials

The completions of the word onsets, together with their word frequency, taken from Francis and Kučera (1982), are shown in parentheses. The frequency counts for the target words are also shown in parentheses.

1. Target-Bearing Items

Words Embedded as Second Syllable of Weak-Strong (WS) Word Onsets, WS Nonword Onsets, and Strong-Weak (SW) Nonword Onsets

Words Embedded as First Syllable of SW Word Onsets, SW Nonword Onsets, and WS Nonword Onsets

WS Word Onset	WS Nonword Onset	SW Nonword Onset	Target Word	SW Word Onset	SW Nonword Onset	WS Nonword Onset	Target Word
doMES (tic63/tically 1)	neMESS	MESStem	mess (26)	SACrif (ice40)	SACKrek	kleSACK	sack (11)
peRIM (eter1)	seRIM	RIMent	rim (8)	BIGam (ist0)	BIGef	sheBIG	big (359)
caLYP (so 1)	baLIP	LIPnel	lip (87)	CELeb (rate25)	CELLeg	derCELL	cell (146)
coNUN (drum 0)	geNUN	NUNtek	nun (6) none (108)	DIGNif (y8)	DIGNep	feDIG	dig (33)
phiLOS (opher25/ophize3/ophy88)	meLOSS	LOSSkem	loss (132)	HYPnot (ise1)	HIPesh	geHIP	hip (18)
coRREC (tion7)	beWRECK	WRECKeb	wreck (19)	COMpan (y453)	COMEpeg	treCOME	come (1561)
traDIT (ion115)	keDISH	DISHek	dish (38)	HEMIs (phere15/pherical1)	HEMep	veHEM	hem (8)
coRROB (orate5)	peROB	ROBeg	rob (15)	ARMIs (tice4)	ARMek	fongARM	arm (278)
heRED (itary2/ity3)	veRED	REDle	red (180) read (83)	PALpit (ate0)	PALpent	vePAL	pal (3)
verNAC (ular2)	derKNACK	KNACKseth	knack (4)	DOCum (ent38)	DOCKyeb	preDOCK	dock (10) doc (19)
verBOS (ity0)	sherBOSS	BOSSet	boss (29)	LIMous (ine5)	LIMetch	kreLIM	limb (10)
seDUC (tion3/tive2)	freDUCK	DUCKrel	duck (21)	DEDic (ate25/ation21)	DEADel	geDEAD	dead (174)
coNNEC (tion86/tive3)	geNECK	NECKshef	neck (83)	JUGger (naut0)	JUGek	meJUG	jug (6)
proTAG (onist2)	teTAG	TAGelt	tag (13)	DECad (ent2/ence2)	DECKep	weDECK	deck (30)
reLUC (tant15/tance5/tantly7)	veLUCK	LUCKem	luck (49)	DIPlom (at12/atic28)	DIPlen	shenDIP	dip (11)
suFFIC (ient63/iency1/iently42)	reFISH	FISHep	fish (63)	RENeg (ade0)	WRENenk	kewWREN	wren (0)
rePUB (lic49)	lePUB	PUBetch	pub (2)	DENig (rate0)	DENedge	beDEN	den (3)
parTIC (ular184/ularity1/ularly146/ulate1)	sheTICK	TICKedge	tick (8)	GELat (ine0)	GELeb	deGEL	gel (2)
Mean frequency: 52 (sd81)			Mean frequency: 54 (sd63)	Mean frequency: 38 (sd102)			Mean frequency: 156 (sd359)

(Appendix continues on next page)

2. Filler Items

Nonwords Embedded as Second Syllable of WS Word Onsets,
WS Nonword Onsets, and SW Nonword Onsets

WS Word Onset	WS Nonword Onset	SW Nonword Onset	Embedded Nonword
feLIC (ity5/itous1)	keLISS	LISSle	liss
linGUIS (tic10/tically1/tics5)	monGWISS	GWISSek	gwiss
maNEUV (er19/erability1/ering1)	paNOOVE	NOOVEsh	noove
peTIT (ion27/ioner31)	greTISH	TISHet	tish
moNOG (amous1/amy1)	beNOG	NOGleb	nog
caTHED (ral11)	laTHEED	THEEDsen	theed
muNIC (ipal29/ipality11/ipally1)	shuNISS	NISStep	niss
noVEM (ber95)	geVEM	VEMshe	vem
gaZEEB (o0)	reZEEB	ZEEBeg	zeeb
umBRELL (a11)	udBRELL	BRELLev	brell
kiLOM (etre11)	meLOM	LOMedge	lom
ocTOB (er79)	epTOBE	TOBEk	tobe
seLEC (tion54/tive19/tively2)	feLECK	LECKeb	leck
traJEC (tory2)	kreJECK	JECKrek	jeck
beLLIG (erent5/erence2/erently1)	meLIDGE	LIDGElt	lidge
coMMEM (orate6)	teMEM	MEMus	mem
riSOTT (o0)	kiZOT	ZOTem	zot
orCHES (tral4)	iKESS	KESSetch	kess

Mean frequency:
24 (sd30)

Nonwords Embedded as First Syllable of SW Word Onsets,
SW Nonword Onsets, and WS Nonword Onsets

SW Word Onset	SW Nonword Onset	WS Nonword Onset	Embedded Nonword
RASperr (y1)	RAZble	weRAZ	raz
GOSSam (er1)	GOSSedge	heGOSS	goss
DROMed (ary0)	DROMEb	leDROM	drom
FACul (ty78)	FACKus	treFACK	fack
FRIVol (ous6)	FRIVek	keFRIV	friv
HABer (dashery2)	HABev	seHAB	hab
FOLLic (le0)	FOLLesh	pleFOLL	fol
PEDes (tal5)	PEDem	gePED	ped
Oblig (ation37/ate4/ational1)	OBlep	fingOB	ob
MESmer (ise1)	MEZmed	sheMEZ	mez
BALcon (y7)	BALKetch	terBAL	bal
DAFFod (i11)	DAFFent	velDAFF	daff
PRIVil (ege38/eged10)	PRIVex	mePRIV	priv
DEStit (ute2)	DEStlet	steDES	dess
SOLit (ude3/ary14)	SOLeth	beSOL	sol
VOLun (tary22/tarily9)	VOLeg	teVOL	vol
JAVel (in0)	JAVez	zeJAV	jav
RELeg (ate6)	RELeng	neREL	rel

Mean frequency:
14 (sd21)

Received October 5, 1992
Revision received June 29, 1993
Accepted June 30, 1993 ■