The recognition of words after their acoustic offset: Evidence and implications

FRANÇOIS GROSJEAN
Northeastern University, Boston, Massachusetts

The aim of this study is to show that continuous spoken word recognition cannot always be a strictly sequential, left-to-right, word-by-word process. The gating paradigm was used to present monosyllabic and polysyllabic words in context and to follow the responses given to them through the next three words of the sentence. The isolation points and total acceptance points were calculated with respect to the acoustic offsets of the words. It was found that more than half of the monosyllabic words were isolated *after* their offset, and that *all* these words, with one exception, received a perfect confidence rating during the next word or words. Reasons for this delay in recognition are examined, and the implications for sequential, left-to-right models of word recognition are discussed.

It is only in the last 10 years or so that the study of spoken (as opposed to written) word recognition has assumed a place of importance in the psycholinguistics of spoken language. The results of numerous experiments and the outcome of much theorizing have substantially increased our knowledge of how spoken words are recognized during on-line processing. We know, for example, that such properties of words as their frequency of use, their length, their phonotactic configuration, and their gender (if they have one) will affect their recognition: low-frequency words take more time to recognize than high-frequency words (Foss, 1969; Howes, 1957; Rubenstein & Pollack, 1963); long words are accessed more slowly than short words (Grosjean, 1980; Mehler, Ségui, & Carey, 1978); words with an early uniqueness point (the point in the left-to-right phonotactic configuration of a word at which it diverges from other words) are recognized earlier than words with a late uniqueness point (Marslen-Wilson, 1984); and words marked for gender (as, for instance, French words preceded by an article) are accessed more rapidly than words not marked in this way (Grosjean & Cornu, 1985). We also know that when words are presented in context, their lexical properties interact with various sources of knowledge to speed up or slow down the recognition process (Cole & Jakimik, 1978; Grosjean, 1980; Marslen-Wilson & Welsh, 1978;

Morton & Long, 1976; Salasoo & Pisoni, 1985; Tyler & Wessels, 1983). Among these sources, we find the listener's knowledge of the world and of the rules of the language, the linguistic information that has already been given by the speaker, the situation, and the topic. We should note at this point that the exact nature of the interaction between the properties of the words (including their acoustic-phonetic characteristics) and these sources of knowledge remains to be described adequately, and that the controversy over the moment at which "top-down" information enters the recognition process has yet to be resolved (Forster, 1976; Marslen-Wilson & Welsh, 1978; Swinney, 1982).

One conclusion that emerges from this research is that spoken word recognition may be a much more complex process than was at first thought. Instead of being a simple mapping between the acoustic-phonetic characteristics of the word and its entry in the mental lexicon (although see Klatt, 1979), it is an operation that probably involves various narrowing-in and monitoring stages, correcting strategies, postaccess decision stages, as well as look-ahead and look-back operations. Existing models of word recognition emphasize some of these aspects: the interaction of "top-down" and "bottom-up" information (Cole & Jakimik, 1978; Marslen-Wilson & Welsh, 1978; Morton, 1969; Tyler, 1984), the narrowing-in process (Forster, 1976; Grosjean, 1980; Marslen-Wilson & Welsh, 1978; Tyler, 1984), and the role of various word properties (Forster, 1976; Morton, 1969). Other views have stressed the autonomy of lexical access (Forster, 1976, 1979; Swinney, 1982) or the continuous word recognition process that occurs during running speech (Grosjean & Gee, 1984; Marcus, 1984; McClelland & Elman, 1985).

One assumption that is made by many models of word recognition is that words are recognized sequentially, left to right, one word at a time. Cole and Jakimik (1979) stated this explicitly when they wrote:

Speech is processed sequentially, word by word.... the words in an utterance are recognized one after another.... Listeners know where words begin and end by recognizing them in order (pp. 133-134).

According to them, this sequential processing enables the listener to locate the onset of the immediately following word and provides him or her with syntactic and semantic constraints to help narrow-in on the next word. Other researchers also make this word-by-word assumption, but in a more implicit manner. Forster (1976) states that the first step in accessing the master lexicon is accomplished by finding an entry for the word in question; Morton (1969) proposes that each word is represented by a logogen; and Marslen-Wilson and Welsh (1978) state that the first segment of a word has the critical role of activating a cohort of candidates. Of course, a word may often be recognized before its acoustic offset, thus allowing the next word to be recognized as it, in turn, is being said (Grosjean, 1980; Marslen-Wilson & Tyler, 1980; Tyler & Wessels, 1983), but, as we will attempt to show in this paper, some words may not be identified before the next ones appear, and word-by-word recognition models do not appear to allow for this.

It is interesting to note that other assumptions are related to the word-by-word approach. One is that the beginning of the word is critical in its recognition (but what happens if the system does not know it is dealing with the beginning of a word? See Cole & Jakimik, 1979, and Marslen-Wilson & Welsh, 1978, for a discussion of this assumption), and another is that words are recognized at their uniqueness point—the point in their phonotactic configuration at which they distinguish themselves from every other word (again proposed by Cole & Jakimik, 1978, and Marslen-Wilson & Welsh, 1978). The above assumptions are certainly correct for many content words, especially if these words occur in an appropriate context, but they probably do not apply to all words, especially if they belong to categories such as: monosyllabic words, low-frequency words, unstressed function words, and words with prefixes and suffixes.

Several reasons explain why researchers have been led to propose a word-by-word approach to word recognition. First, spoken word recognition research has always lagged behind the research on written word recognition and has borrowed from it the unit of analysis—the written word in its spoken form. Because written language is ever present in our everyday life, it is all too easy to think of spoken language as a concatenation of individual words, even though the actual acoustic-phonetic stream of spoken language does not reflect this. Many researchers have assumed that the written word has an acoustic-phonetic correlate in the speech stream, and that it is this "unit" that is involved in the recognition process. Whether there is direct access from the spectra (as in Klatt, 1979) or intermediary stages of representation that involve linguistic units, such as the phoneme or the syllable (as in other models; see Pisoni, 1984), the assumption has

long been that the domain over which processing takes place is the spoken analog of the written word (what Grosjean & Gee, in press, have referred to as the written dictionary word, or WD word). The final outcome of spoken word recognition is the stored lexical item in the internal lexicon (an item that shares many characteristics with the WD word), but it is questionable whether, in the earlier stages of word recognition, the WD word plays the important role many have assumed, either as a "unit of processing" or as the domain over which word recognition takes place.

A second reason why word-by-word models have had such an impact is that the research they are based on has often dealt with the recognition of single words—presented in their canonical form or excised from the speech stream and presented in isolation—and not of continuous strings of words. Third, the experimental tasks that have been used are often biased toward single WD-like words: word monitoring and lexical decision tasks, among others, force the listener to focus on single words; this, in turn, encourages us to think that the human processing system may, in fact, use the WD word as a unit or domain of processing. Finally, many of the words used in experiments have been content words (as opposed to function words) and have had few, if any, inflections. One consequence of this is that most of the stimulus words used have been recognized before their acoustic-phonetic offset, thus leading researchers to conclude that word recognition is a strictly sequential, word-by-word, process.

The question now becomes: What theoretical arguments and what empirical evidence can be put forward against this strictly sequential, left-to-right, word-by-word view of word recognition? One need only examine the spectrogram or waveform of an utterance to be reminded of the continuous nature of the speech stream and the ever present segmentation problem faced by the speech perception and word recognition systems. Word boundaries quite often fail to be marked segmentally or prosodically, unstressed function words are shortened and cliticized, and elisions and reductions occur throughout the speech stream. All this impacts on the segmentation operation, as shown in slips of the ear. Bond and Garnes (1979) report, for example, that 70% of multiple word slips of the ear involve problems with word boundaries: word boundary shifts (an ice bucket → a nice bucket), word boundary deletions (ten year party → tenure party), and word boundary additions (descriptive linguistics → the script of linguistics). Problems of segmentation can also be found in the experimental literature. In a set of classic studies, Pollack and Pickett (1963, 1964) found that only 55% of the words extracted from the speech stream were identified correctly when presented in isolation, and that this percentage reached only 70% to 80% accuracy when the words were presented in two- to three-word samples. Grosjean (1980) obtained similar results when he presented subjects with gated words, excised from context, in segments of increasing duration: only 50% of the one-syllable low-frequency words (and 75% of the high-

frequency words) were guessed correctly by five or more of the eight subjects used in the experiment.

The aim of the present study was to produce further evidence that word recognition is not a strictly sequential, left-to-right, word-by-word process. Although some words may be recognized as they are being said, others will be recognized *after* their acoustic offset, that is, during the following word or words. When this occurs, listeners will be faced with the problem of recognizing two or more words simultaneously and not sequentially, as most models would propose. To show that the results reported by Pollack and Pickett (1963, 1964) and Grosjean (1980) were not a consequence of extracting a word from its normal acoustic-phonetic environment and presenting it in isolation, we used the gating technique (see Grosjean, 1980, Ohman, 1966, and Pollack & Pickett, 1963, for earlier versions of the paradigm) to present a word in context and followed the responses given to it through the next three words of the sentence. Thus, in the sentence, "I saw the bun in the store," we presented "I saw the" and started gating at the beginning of "bun." We continued doing so, in segments of increasing duration, until we reached the end of the sentence. After each gate, subjects were asked to guess the noun being presented ("bun"), to indicate how confident they were in their guess, and to finish off the sentence.

We examined two "recognition points" produced by the gating paradigm: the isolation point, that is, that point at which the listener has isolated a candidate but may still feel unsure about it, and the total acceptance point, that is, that point at which a perfect confidence rating is given by the listener to the candidate. In terms of actual word recognition, we propose, quite tentatively, that the isolation point reflects the moment in time at which the listener has a strong candidate in mind but has not yet finally decided to use it in the construction of the interpretative representation of the on-going message. This point is very close to the word's uniqueness point as defined in Marslen-Wilson and Welsh's (1978) cohort model—some 19 to 71 msec according to Tyler and Wessels (1983, 1984)—and corresponds quite closely to what Bradley and Forster (1984) mean when they say that a word has been accessed. The total acceptance point, on the other hand, is that moment in time at which the word starts being used in the construction of the interpretative representation. This point occurs later than the separation point and appears to correspond quite closely to what Bradley and Forster mean by word recognition, that is, the listener's fixation of belief that he or she has indeed heard word X. Because we had no direct way of knowing exactly when actual word recognition takes place (what we have just proposed is based in large part on conjecture), we examined the responses at both the isolation and the total acceptance points, keeping in mind all along that these might actually be minimum and maximum points in the recognition process (see Grosjean, 1980, Tyler, 1984, and Tyler & Wessels, 1983, 1984, for a discussion of how various

measures produced by the gating paradigm may or may not reflect the actual on-line processing of a word).[1]

We had found, in our first gating study (Grosjean, 1980), that subjects had great difficulties with low-frequency, monosyllabic nouns, many of which they failed to isolate by the last gate, the gate at which the whole word was presented but not the following context. In the present study, we examined the isolation point and total acceptance point of an extended set of these words presented in context. Our intention was to attempt to show that some of these nouns were isolated *after* their acoustic offset, as information about the next word or words was being perceived, and fully accepted even later. We contrasted these words with those of a comparison (filler) group that were two or three syllables long, and of varying frequency, in an attempt to show that these longer words were often isolated and fully accepted *before* their acoustic offset. The latter result, if obtained, would be a replication of results found in other studies (Grosjean, 1980; Marslen-Wilson, 1984; Marslen-Wilson & Tyler, 1980); what would be new in the current study would be a demonstration that there *are* words that are actually recognized *after* their acoustic offset.

We end the report by discussing the reasons that can account for late recognition and by examining the implications that this must have on models of word recognition that function in a strictly sequential, word-by-word manner. We then outline a view of word recognition (Grosjean & Gee, in press) in which the acoustic-phonetic counterpart of the written dictionary word is no longer the unit or domain of processing.

METHOD

Subjects

Twelve undergraduate students, with no reported speech or hearing defects, served individually in three separate sessions lasting 1 h each.

Materials

Twenty monosyllabic nouns with a frequency of 1 were chosen from the Kučera and Francis (1967) list. In order not to bias subjects toward a particular word length or word frequency, another 20 "filler" nouns (11 two-syllable and 9 three-syllable words) with varying frequencies (geometric mean of 14.6, range extending from 1 to 382) were also chosen from the same list.

Each noun was embedded in a sentence that began with "I saw the" and ended with a prepositional phrase (preposition, determiner, noun) specific to each word. Fifteen judges were asked to estimate how well the prepositional phrases fitted the preceding context ("I saw the" plus the stimulus word). To do this, they put a slash on a 90-mm rating scale labeled "Poor fit" at one end and "Good fit" at the other. Only phrases with average ratings of 60 mm or above were retained for the experiment. Examples of the complete sentences containing the monosyllabic nouns (in capitals here) are:

I saw the DOE in the woods
I saw the BLIMP in the sky
I saw the BUN in the store
I saw the BOAR in the woods

Examples of the sentences containing two- and three-syllable nouns are:

I saw the DONKEY in the stable
I saw the TRAWLER out at sea
I saw the PRESIDENT at the convention
I saw the CAPTAIN in the cockpit

A complete list of the stimulus words and their context is given in the appendix.

The 40 sentences were randomized and recorded on audiotape by a female speaker at normal rate and with regular prosody. Each sentence was then digitized at a sampling rate of 10 kHz and gated following the procedure described by Grosjean (1980). For each sentence the "onsets" of the stimulus noun, the preposition, the determiner, and the last noun were located, as best as possible, by inspecting the speech wave and by using auditory feedback. The stimulus nouns always began with a plosive consonant or fricative, and their onset corresponded therefore to the release of the word initial plosive or the start of the frication. The onsets of the other three words (preposition, determiner, final noun) were harder to locate and judges were used to confirm these (see below).

The presentation set of each gated sentence was prepared in the following way. The first gate contained the sentence "I saw the" up to, but not including, the onset of the stimulus word. The second gate contained the same information plus the first 50 msec of the word. From then on, gates were incremented by 50 msec. When the duration of the stimulus word was not an exact multiple of 50, the gate containing the full stimulus word was incremented by an appropriate amount. The gate following the carrier sentence plus the stimulus word was once again incremented by 50 msec, presenting thus the beginning of the following preposition. The sequencing of gates was continued in this way (with certain end-of-word gates not always incremented by exactly 50 msec) until the presentation set was complete. Thus, each presentation set ranged from "I saw the" all the way to the end of the sentence (for example, "store" in "I saw the bun in the store"). Depending on the length of the stimulus noun and the last noun in the sentence, a presentation set contained between 15 and 35 gates.

To confirm the location of the onsets of the words in the prepositional phrase, the 40 presentation sets were played individually to 10 judges who were asked to indicate at which gate in the presentation set the preposition, the article, and the last noun began. To do this, the judges were given a response sheet for each set. At the top, the full sentence was presented, and to the right of the gate numbers (listed vertically), there were three columns entitled "Prep," "Art," and "Noun." The judges were asked to work down the Prep column as they started listening to the first gates (those containing the stimulus word) and to mark the sheet with a minus (−) sign for as long as they kept hearing parts of the stimulus word. As soon as they heard the beginning of the preposition, they were to put a plus (+) in the column and move over horizontally to the "Art" column. Here the procedure was to be repeated: a minus for each successive gate until they heard the beginning of the article (for which they would put a plus); they would then move over to the "Noun" column, and repeat the operation.

The gate numbers at which judges heard the beginning of each of the three words in each sentence were averaged and rounded to the nearest whole number. The gates corresponding to these numbers were then adopted to reflect the location of the onset of the three words in question: the preposition, the determiner, and the final noun. Whenever an onset gate did not correspond to the experimenter's original onset gate, the gate chosen by the judges took precedence. (It should be noted that the percent agreement between the experimenter and the judges was never lower than 80%.)

For the purpose of data analysis, word onset and word offset values were then computed. Although phonetic coarticulation does not allow one to pinpoint exactly the beginnings and ends of words, it was decided nevertheless to find some "tentative" beginning and

end values. The onset value of a word was taken to be the duration of the gate preceding the gate containing the onset of the word plus 1 msec. Thus, for example, if a gate with a duration of 270 msec was chosen as the gate containing the onset of the preposition, the actual onset value of the preposition was taken to be 221 msec: the duration of the preceding gate, 220 msec, plus 1 msec. The offset value of a particular word was taken to be the duration of the last gate in which only that particular word, and no elements of the following word, had been heard (220 msec for the noun in the example above). Subtracting a word's onset value from its offset value gave the word's duration in milliseconds.

Procedure

The 12 subjects were run individually on each of the 40 presentation sets. The sets were recorded on three tapes, and the subjects were cycled through the tapes in such a way that each new subject started with a different tape. A half hour rest (at the minimum) was given between each tape.

The subjects were instructed to listen to presentations within each set, and after each presentation, to write down the word they thought was being presented after "I saw the," to indicate how confident they were about their guess by circling a number on a 1-to-10 "very unsure-very sure" scale, and to write down an ending to the sentence. They were given 8 sec between each presentation to accomplish these three tasks. It should be noted that the answer sheet was arranged in such a way that the focus was put on the stimulus noun to be guessed. The left-hand column was entitled "Word," the middle column contained the confidence-scale numbers ranging from 1 to 10, and the right-hand column was entitled "Rest of sentence." This forced subjects to first propose a word, indicate their level of confidence concerning the word, and finally finish off the sentence. The subjects were asked to give a response after every presentation, however unsure they might feel about the stimulus word and the rest of the sentence.

Data Analysis

Individual response sheets were examined to determine the isolation point and the total acceptance point of each stimulus word. The isolation point was defined as the gate duration at which the subject correctly guessed the stimulus word and did not subsequently change his or her guess; the total acceptance point was defined as the gate duration at which the subject gave the stimulus word a perfect confidence rating (10) and did not subsequently change the rating.

Two values were computed for each of these two points. For the isolation point, one value was the isolation point of the word (in msec) with respect to its acoustic offset. This gave values that could range from a negative number, indicating that a word was isolated before it ended, to a positive number, indicating that the word was isolated after its offset. The second value was the isolation point with respect to its position in the sentence. Thus, if the stimulus word was isolated before its offset, it was given a value that corresponded to its isolation position within itself (expressed as a percent of the way through the word). If the stimulus word was isolated during the next word (the preposition), it was given a value that corresponded to its isolation position within that word (a percentage of the way through the preposition), and so on. In short, the word during which the stimulus word was isolated and the point of isolation within that word were tabulated for each stimulus word and each subject.

The total acceptance point was likewise represented by two values: the total acceptance point with respect to the offset of the stimulus word and the same point with respect to its position in the sentence.

Means were calculated for the different values obtained for each word. These means were based on the number of subjects who produced a value, and not on the total number of subjects run. Some subjects never isolated a word—they did not know it—or never gave it a perfect confidence rating and hence were not included in the computation of the mean. Simple t tests were then used to test the

difference between the means of the monosyllabic words of the experimental group and the polysyllabic words of the comparison group.

Finally, the erroneous candidates proposed before the isolation point were analyzed with respect to the acoustic-phonetic information that was available to the listener.

RESULTS AND DISCUSSION

The Isolation Point

Table 1 presents the mean, standard deviation, and range of the isolation points (in msec) of the monosyllabic and polysyllabic words with respect to their acoustic-phonetic offset. These measures are based on 231 obtained values out of a possible 240 for the monosyllabic nouns (15 words were isolated by all 12 subjects, 4 words were isolated by 11 subjects, and 1 word was isolated by only 7 subjects). For the polysyllabic words, the measures are based on 239 values out of a possible 240 (1 word was isolated by 11 subjects rather than 12). In all, 97.9% of all word tokens were finally isolated, indicating that the stimulus items were well known to the subjects. Monosyllabic words were isolated on the average 1.7 msec *after* their offset, whereas polysyllabic words were isolated on average 206.2 msec *before* their offset (t = 7.14, p < .01). This highly significant difference is in large part due to the earlier uniqueness points of the polysyllabic words (see "Reasons for delayed recognition" below, as well as Marslen-Wilson, 1984; Tyler & Wessels, 1983) and not to the higher frequency of occurrence of these words: as can be seen in Table 1, the eight low-frequency words in this group (frequency of 1 on the Kučera and Francis, 1967, list) were also isolated much before their acoustic offset: 194.5 msec on the average.

What is much more revealing, and of direct importance for our case, is the distribution of values for the two sets of words. Although all the polysyllabic words were isolated before their offset [the values range from −459 msec ("pharmacy") to −37 msec ("trawler")], only 9 of the 20 monosyllabic words achieved this: words like "flute" (−123 msec), "blimp" (−103 msec), "stunt" (−79 msec), "spout" (−78 msec), and "gash" (−64 msec), for example. The other 11 words were isolated *after* their offset, while subjects were already listening to the next word (or words). In this category, we find

words like "brawl" (+22 msec), "plum" (+43 msec), "pram" (+81 msec), "pout" (+98 msec), "boar" (+95 msec), and "doe" (+189 msec). It should be noted that these positive isolation points are not the result of a small number of subjects who isolated the words relatively late in the sentence and hence skewed the distribution of responses toward the positive values: for the 11 words isolated after their offset, 50.5% of the subjects, on the average, obtained positive isolation values.

We can conclude from this that words are not always recognized before their acoustic offset (that is, as they are being said), especially if they are short. We should stress at this point that the isolation times presented in Table 1 probably do not reflect the actual recognition times of the words. Isolation times not only underestimate the uniqueness points of the words (Tyler & Wessels, 1983, 1984, estimate this underestimation to be in the order of 19 to 71 msec, depending on the response requested of subjects), but they underestimate even more the actual recognition points of the words, which probably do not occur until there is a high level of perceptual belief for a particular word on the part of the listener. We can conclude from this that the majority of the monosyllabic words we used were probably recognized after their offset, and that, in general, many monosyllabic words are recognized after they have ended.

The isolation point of a word can also be studied with respect to its position in the sentence: during the word itself or during the next word or words. Figure 1 presents the percentage of stimulus words isolated as a function of where in the sentence they were isolated. We first note that all polysyllabic words are isolated before their ending (as we saw in Table 1), and that most fall in the interval corresponding to 51% to 75% of the way through the word. In contrast, we note that the isolation points of the monosyllabic words occur at various locations in the sentence: during the stimulus word itself (n=9), during the following preposition (n=10), and even during the following determiner (n=1). In the first location, we find words like "blimp," "spout," and "tramp"; in the second location, we note words like "plum" (isolated 32% of the way through the preposition), "pout" (66%), "pram" (82%), and "boar" (96%); and in the third location, we find "bun," located 6% of the way through the determiner.

Table 1
Mean, Standard Deviation, and Range of the Isolation Points (in msec) of the
Monosyllabic Words and Polysyllabic Words, All and the Low-Frequency
Subset, with Respect to Their Acoustic-Phonetic Offset

Monosyllabic Words (n=20)			Polysyllabic Words (n=20)			t Test		
Mean	SD	Range	Mean	SD	Range	t	df	Level (2-tail)
			All Frequencies (n=20)					
			−206.2	91.5	−459 to −37	7.14	38	.01
+1.7	87.9	−123 to +189						
			Low Frequencies (n=8)					
			−194.5	113.2	−459 to −37	4.71	26	.01

Note —t tests compare monosyllabic words with all polysyllabic words as well as with the low-frequency subset.
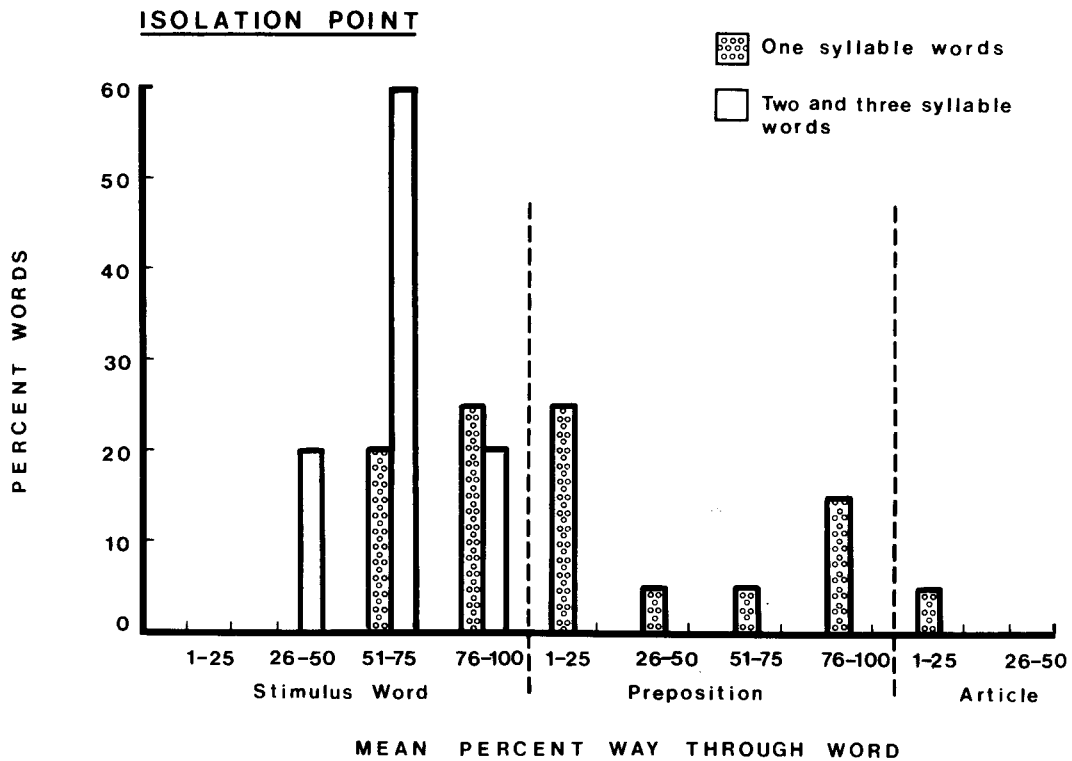
ISOLATION POINT

Figure 1. Percentage of the stimulus words isolated as a function of isolation position in the sentence: during the word itself, during the next word (preposition), or during the word after that (article). Percentages are based on 20 one-syllable words and 20 two- and three-syllable words.

This second analysis of the isolation point not only shows, once again, that some words are isolated after their acoustic offset, but also suggests that the perception of the following syllable is often sufficient to isolate the word that was not identified while it was being heard. Whether this next syllable is sufficient for the listener to totally accept the word will now be examined.

The Total Acceptance Point

Table 2 presents the mean, standard deviation, and range of the total acceptance points (in msec) of the monosyllabic and polysyllabic words with respect to their acoustic-phonetic offset. These measures are based on 228 obtained values out of a possible 240 for the monosyllabic

words (13 words were given a perfect confidence rating by all 12 subjects, 6 words were given this rating by 11 subjects, and 1 word was given the rating by only 6 subjects). For the polysyllabic words, the measures are based on 238 values out of a possible 240 (2 words were given a perfect rating by 11 subjects rather than 12). In all, 97.1% of all word tokens were given a perfect confidence rating, indicating that subjects felt very confident about the vast majority of words. Although two- and three-syllable words usually received a perfect confidence rating before their offset (56.4 msec before, on the average), the monosyllabic words were given a perfect rating only after their offset (143.2 msec after, on the average). The difference between the two types of words is highly

Table 2
Mean, Standard Deviation, and Range of the Total Acceptance Points (in msec) of the
Monosyllabic Words and Polysyllabic Words, All and Low-Frequency
Subset, with Respect to Their Acoustic-Phonetic Offset

Monosyllabic Words (n=20)			Polysyllabic Words (n=20)			t Test		
Mean	SD	Range	Mean	SD	Range	t	df	Level (2-tail)
					All Frequencies (n=20)			
			−56.4	89.6	−330 to +124	6.95	38	.01
143.2	87.4	0 to +308						
					Low Frequencies (n=8)			
			−64.4	119.9	−330 to +124	4.89	26	.01

Note — t tests compare monosyllabic words with all polysyllabic words as well as with the low-frequency subset.

significant (t = 6.95, p < .01) and is not due to the lower frequency of the monosyllabic words: Two- and three-syllable low-frequency words were also accepted before their offset, as can be seen in Table 2.

The distribution of values obtained confirms this finding. Although 16 of the 20 polysyllabic words received a perfect confidence rating before (or at) their offset (the only exceptions were "camel" (+9 msec), "difference" (+64 msec), "partner" (+67 msec), and "trawler" (+124 msec), only 1 monosyllabic word ("snail") fell into this category. All the other words were accepted after their offset, some needing 200 msec or more of the next word or words to reach a confidence rating of 10. Among these, we find "boar" (+208 msec), "pout" (+244 msec), "bun" (+273 msec), and "doe" (+308 msec). Once again, these positive values are not the result of a small number of subjects who gave perfect ratings very late in the sentence and hence skewed the distribution of responses toward the positive values: for the 19 monosyllabic words that received a perfect confidence rating after their offset, 81.4% of the subjects, on the average, produced positive total acceptance values.

The pattern becomes even clearer if we examine the total acceptance point of a word with respect to its position in the sentence. Figure 2 presents the percentage of the words that received a confidence rating of 10 during the word itself or during the following preposition, article, or noun. As can be seen, polysyllabic words are either accepted totally before their offset or midway through the next word (a preposition). In contrast, the total acceptance point of monosyllabic words covers a much wider range, extending from the word itself, all the way to the last noun in the sentence, three words later. One word ("snail") is accepted totally at its offset, as we have already seen; 10 words receive a perfect confidence rating during the preposition [among these we find "flute" (7% of the way through the preposition), "stunt" (35%), "crust" (73%), and "brawl" (92%)]; 4 words have to wait until the following article to receive a perfect 10 ["pub" (13%), "plum" (24%), "dune" (41%), and "perch" (58%)]; and as many as 5 words (out of 20) are given a perfect confidence rating during the final noun ["pout" (3%), "boar" (10%), "bun" (12%), "doe" (19%), and "pram" (38%)].

These results strengthen our findings relative to the isolation point and stress once again that not all words are recognized before their acoustic offset. Whether or not the total acceptance point, which occurs some 140 to 150 msec on the average after the isolation point, is an overestimate of a word's actual recognition point is an open question. If we adopt the definition of word recognition proposed by Bradley and Forster (1984), which stresses the listener's fixation of belief for a particular word, then it probably is not. Whatever the reality, the two measures we used—a very conservative one (the isolation point) and a more liberal one (the total acceptance
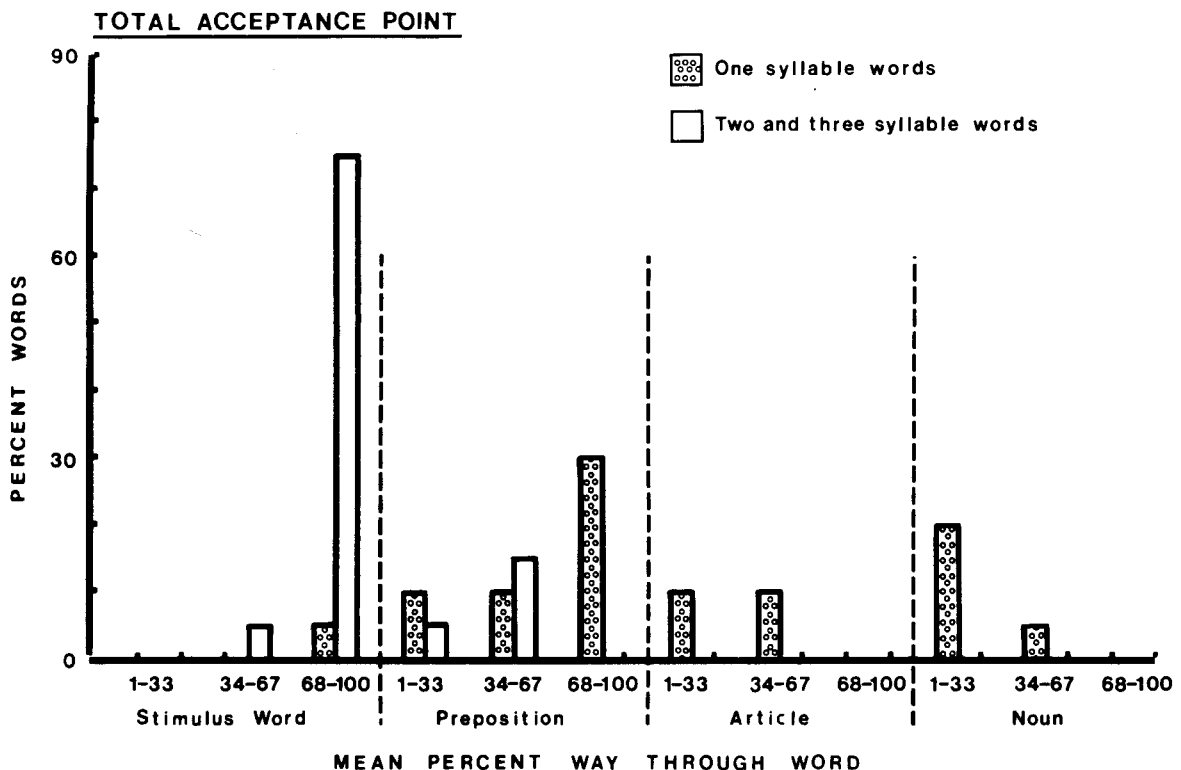

Figure 2. Percentage of the stimulus words that received a perfect confidence rating as a function of where this occurred in the sentence: during the word itself or during the words after. Percentages are based on 20 monosyllabic and 20 polysyllabic words.

point)—clearly indicate that the actual recognition point of a word can quite often occur after the word in question has been heard, that is, while the listener is hearing the next word or words.

Reasons for Delayed Recognition

Because there has been a bias in the literature toward immediate, word-by-word recognition, few studies have examined the causes for recognition delay. One obvious factor is the semantic or pragmatic context which at times can be so at odds with the bottom-up information that final acceptance of the word in question (and probably even its isolation) will be delayed for some time. Many misperceptions (slips of the ear) are the product of such conflicts between top-down and bottom-up information. Another reason pertains to the way the word is said; if it is articulated too rapidly, in too relaxed a manner or in some abnormal manner, then recognition may take more time. Linked to this is the paucity of word boundary cues. Although Nakatani and Dukes (1977) and Nakatani and Schaffer (1978) have shown how juncture cues such as glottal stops, allophonic variations, laryngealization, pauses, and stress patterns can mark word boundaries, it is well known that words are not always marked in these ways. We also know that words that have these cues can sometimes lose them in certain contexts, such as during rapid or poorly articulated speech. A fourth reason concerns the frequency of occurrence of a word. One

can imagine words that are heard so infrequently that complex processing must take place (involving erroneous narrowing-in, correcting, and back-tracking) before the word is finally isolated and recognized. One of the words used in our study—"pram"—happened to be such an item (all other words, as we saw above, were isolated by at least 11 out of the 12 subjects). "Pram" is a perfectly appropriate and quite frequent word in British English, but is rarely used in American English (Americans prefer the term "baby carriage"). This lack of usage resulted in the word's being isolated 81 msec after its offset and being totally accepted 211 msec after that (during the noun of the prepositional phrase). The candidates proposed before the isolation point are indicative of the processing difficulties encountered by the listeners. Among these we find "print," "prance," "parameter," and "perimeter."

A fifth reason for delayed recognition, and one that is problematic for models that emphasize bottom-up processing and the word's uniqueness point, is that many words are not phonotactically unique by the time they end. This means that the listener has to process the beginning of the next word as if it were part of the first word, and only then parse the words correctly. Several instances of this occurred in the isolation patterns we obtained. In Figure 3, we present the candidates proposed for the word "bun" (horizontal axis) as the length of the gate increased in duration (vertical axis). Note that we have marked the offsets of the words "bun," "in," and "the" with a horizon-
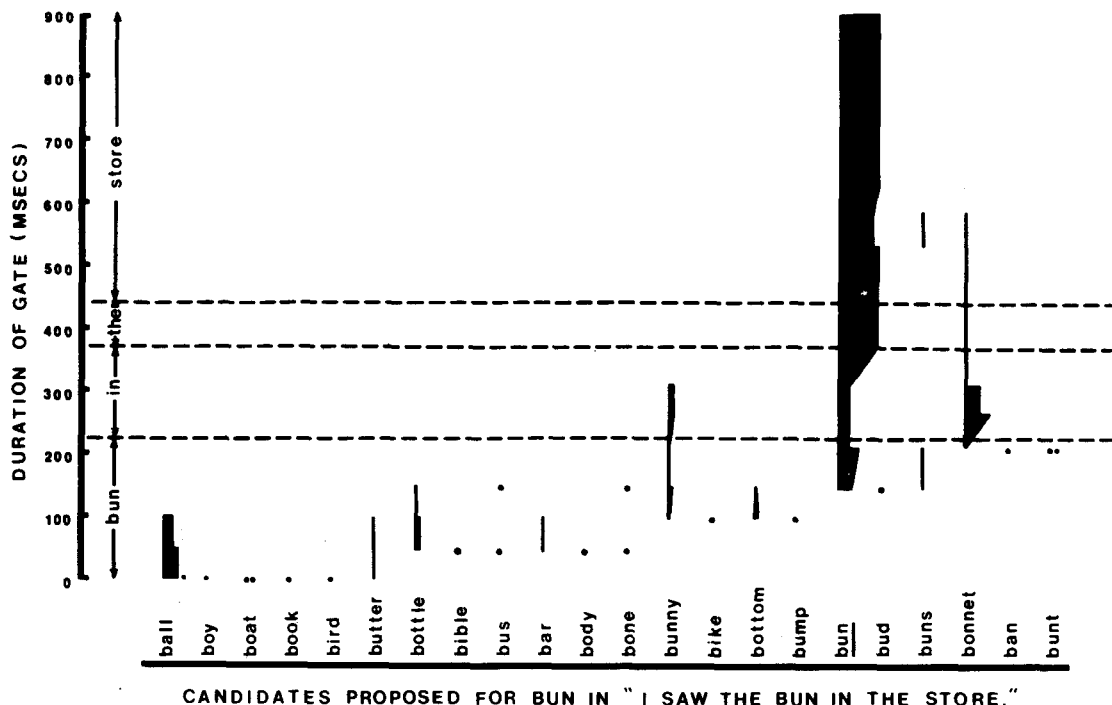

Figure 3. The candidates proposed for the word "bun" (horizontal axis) as the length of the gate increased in duration at each presentation (vertical axis). The dashed horizontal lines mark the offsets of the words "bun," "in," and "the," respectively. Candidates proposed at only one presentation (or gate) duration are depicted with a dot; those proposed over two or more presentations are depicted with a continuous line. The number of subjects proposing a particular candidate is represented by the thickness of the line—the more subjects, the thicker the line.

tal dashed line. Candidates proposed at only one gate duration are depicted with a dot; those that are proposed over two or more presentations are depicted by a continuous line. The number of subjects proposing a particular candidate is represented by the thickness of the line—the more subjects, the thicker the line. As can be seen in the figure, most subjects failed to isolate (and much less recognize) "bun" before its offset. This is because a word like "bun," in the type of nonconstraining context it was presented in ("I saw the bun"), can be parsed only after the listener has heard the segment that follows. The syllable /bʌn/ can be a word by itself, but can also be the beginning of numerous other words, such as "bunch," "bundle," "bunny," "bunt," and so forth. Although some subjects chose "bun" as a candidate before the word's offset, most integrated the syllable with the following syllable (/ɪn/) and proposed words compatible with this: "bunny" and "bonnet." It was only after the /n/ of /ɪn/ was heard fully that subjects realized they were dealing with two lexical items ("bun" and "in") and not just one. This is a case, therefore, of the simultaneous isolation of two words, and not of a word-by-word, left-to-right identification process.

Figure 4 presents a very similar example. The word "plum," in the context "I saw the plum on the tree," reached its uniqueness point only during the following word: /plʌm/ can be a word on its own but it can also be the beginning of "plummet," "plump," "plumber," and so forth. Some subjects opted for the "plum" solu-

tion, but others, as they heard more and more of the following preposition ("on"), proposed "plumber" as a candidate. Only the nasal consonant, and especially the beginning of the /ð/, made them backtrack and parse the sequence into two words: "plum" and "on." An examination of the candidates proposed for other items that must await the following word to be correctly identified showed similar isolation patterns. For example, the sequence "doe in" gave "donut," and "boar in" gave "board" and "born." It is interesting to note that words that are uniquely specified within themselves rarely posed any problems: "skit," "blimp," and "spout," for example, were all isolated before their offset (although they, too, had to wait for the next word or words to be fully accepted).

Thus, the uniqueness point of a word can explain, in part at least, how early (or late) a word will be recognized. We should stress two points that complicate the picture, however. First, the uniqueness point is only one among many factors accounting for the recognition of words in context. It interacts with syntactic, semantic, and pragmatic variables, the frequency and articulation characteristics of the word in question, as well as a host of recognition strategies (see the "word from a word garden path" in Grosjean, 1980, for example). As a result, the importance of the uniqueness point variable can be reduced quite substantially in the word recognition process. In the case of our 20 monosyllabic nouns, for example, the Pearson product-moment correlation between the uniqueness point
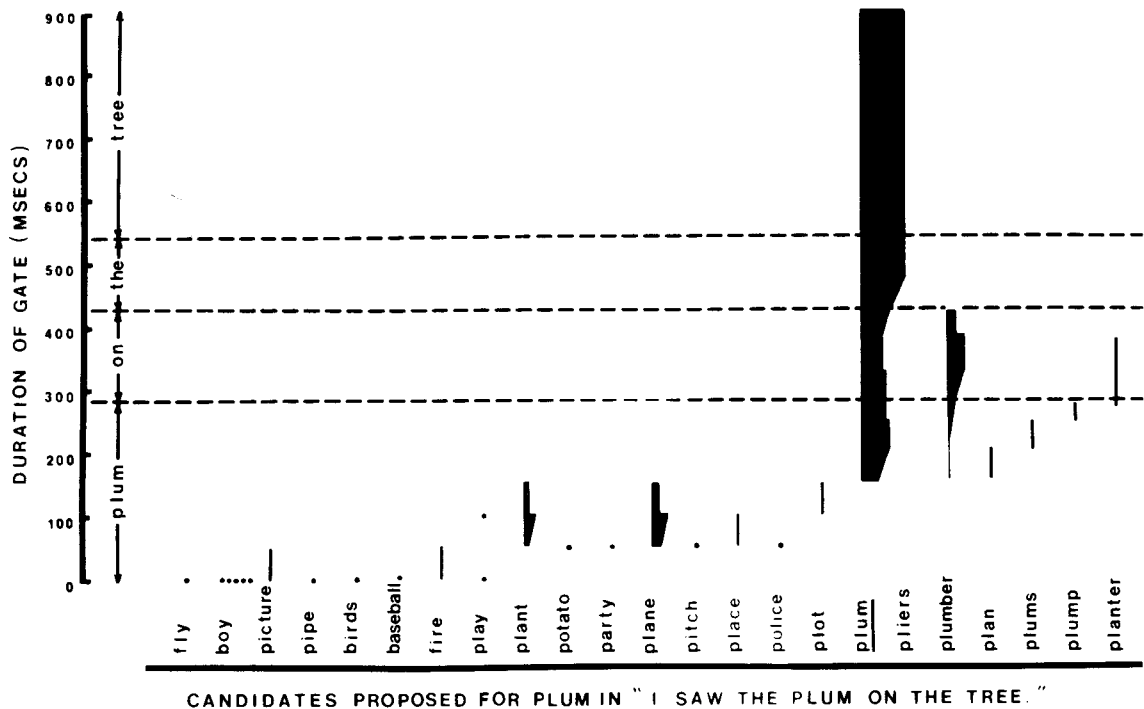

Figure 4. The candidates proposed for the word "plum" (horizontal axis) as the length of the gates increased in duration at each presentation (vertical axis). The dashed horizontal lines mark the offsets of the words "plum," "on," and "the," respectively. (See Figure 3.)

of the words and their isolation point was only 0.28, n.s. (The coefficient of correlation increased slightly when the uniqueness point was correlated with the total acceptance point: 0.37, n.s.) An examination of the scatter plot reveals that the low coefficients are due to the fact that certain words with uniqueness points *after* their last phoneme (e.g., "tramp" because of "trampoline," "flute" because of "flutist," "pub" because of "public," etc.) were nevertheless isolated at or before their offset. In each case, the subjects seemed to be opting for a monosyllabic noun in preference to a possible polysyllabic noun. This strategy is akin to the "word from a word garden path" mentioned above, and is simulated correctly, as it happens, by the McClelland and Elman (1985) trace model.

A second important point to keep in mind concerning the uniqueness point variable is that many words in context, as we saw above, are not unique before or at their last phoneme. When this is the case, as it was for 11 of our 20 nouns, then there is no predefined uniqueness point for the word. It can be determined only by examining what follows. Thus, in the case of "bun in the store," the uniqueness point of "bun" is at the /n/ of the "in," that is, two phonemes after the end of the word. It is only then that the listener knows that the sequence is "bun in" and not "bunny," for example. If the sentence had been "I saw the bun Mummy gave you," the uniqueness point would have been at the /m/, that is, only one phoneme later.

We should mention, at this stage, some recent findings obtained by Luce (1984). He used a 20,000-word computerized lexicon to compute the uniqueness point of all words with more than two phonemes, and found that when frequency of usage is taken into account, the probability of a word diverging from all other words prior to the end of the word itself is only 0.39. That is, the large majority of words become unique only *after* the fact, as the next word is being processed. What is more, short words, which tend to be the more frequent words, are especially prone to having uniqueness points after their last phoneme: 94.3% of words two phonemes long become unique during the next word; this is true for 74.17% of words three phonemes long and for 35.64% of words four phonemes long. Luce concludes that it is virtually impossible to recognize a word in fluent speech without first having heard the entire word as well as a portion of the following word. The present study brings experimental confirmation to this conclusion.

Implications for Models of Word Recognition

It has been shown in this paper that, during the continuous processing of speech, not all words are recognized before their acoustic offset. This has been shown by an examination of two points produced by the gating paradigm: the isolation point, that point at which the system has isolated a candidate but is still unsure about it, and the total acceptance point, that point at which the word can be said to have been recognized and can now be used in the construction of the interpretative representation.

Data obtained at these two points lead us to conclude that certain types of words, such as short words, cannot be isolated and much less totally accepted before their acoustic offsets. Although many words may indeed be recognized as they are being said (Grosjean, 1980; Marslen-Wilson & Tyler, 1980; Tyler & Wessels, 1983), there will be other words that will be recognized later—at a point in time, according to word-by-word models, when the system should be busy recognizing the next word, and should no longer be working on the preceding (and as yet unrecognized) word.

The results obtained raise some interesting problems, therefore, for those models of word recognition that state that words are recognized sequentially, left to right, one word at a time (Cole & Jakimik, 1979; Forster, 1976; Marslen-Wilson & Welsh, 1978; Morton, 1969). For example, in the cohort model (Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978), much stress is put on the fact that it is the initial segment of the word (the first one or two phonemes) that activates the initial cohort of word candidates. If the preceding word has been recognized before its offset, then the system is ready for the beginning of the next word and normal activation can take place. But what happens when the preceding word is still being recognized as the beginning of the next word occurs? The system does not know it is dealing with the initial segment of a word (it thinks it is part of the previous word) and does not therefore activate a cohort. When the previous word has finally been recognized, the system has some leftover material it has to deal with as well as the segments that have been "coming in" all along. Current versions of the cohort model will need to be modified to take this situation into account.

It would appear, then, that our results cause problems for models of word recognition that defend a totally sequential, left-to-right, word-by-word process. A possible fallback position for these models could be the following: The recognition process is indeed sequential but not always in synchrony with the acoustic-phonetic stream. Thus, when a particular word is recognized late, the recognition of the following word is itself delayed and occurs only once the first word has been recognized. The actual recognition process remains sequential, but it is, at times, "out of sync" with the acoustic-phonetic stream.

To assess this modified sequential position, we examined the individual candidates proposed for the words that were isolated after their offset. We took the 11 stimulus nouns that received positive isolation values, and for each, we analyzed the individual cases of late isolation. These we put into one of three categories: (1) those cases in which the preposition following the noun was isolated before the noun itself (e.g., at a gate before the isolation gate of the noun)—the case, for instance, where a subject proposes "bonnet in the window" when hearing the whole of "bun" and part of "in," in the sentence, "I saw the bun in the store"; (2) those cases in which the preposition following the noun was isolated at the same time as the noun (i.e., at the same isolation gate)—for example,

when a subject has been writing "bonnet on her head" and then shifts over to "bun in the oven," where both the noun and the preposition are correct; and (3) those cases in which the preposition following the noun was isolated after the noun (i.e., at a later gate)—for example, when a subject isolates "bun" first (he/she writes "bun at the bakery") and then, at a later gate, isolates the preposition by writing "bun in the bakery." We should note that the modified sequential model we proposed above hypothesizes that all the cases should fall into the third category: when the recognition of a word is delayed, the recognition of the next word will also be delayed and should occur only after the first word has been recognized. On the other hand, a nonsequential model would allow all three possibilities to occur, although in the case of low-frequency words with late uniqueness points (as in our study), it would predict that the majority of cases would fall into the first two categories, where the following preposition is isolated along with or before the noun.

The results we obtained clearly show little support for the modified sequential position. On average, only 7.3% of the 93 cases tabulated fell into the category "Preposition isolated after noun." The vast majority of the cases fell either into the category "Preposition isolated before the noun" (36.6%, on average) or the category "Preposition isolated at the same time as the noun" (56.1%, on average). We can conclude from this that a word embedded in a continuous speech stream can be recognized at various points in time: before, at the same time as, or after the next word is recognized. It is certainly not the case that a word is always recognized prior to the next word, as most sequential, word-by-word models would propose.

We should note that a number of recent models of word recognition do away with the strictly sequential, left-to-right, word-by-word aspect of word recognition. For example, in the trace model of speech perception, McClelland and Elman (1985) propose that information processing takes place through the excitatory and inhibitory interactions of a large number of simple processing units. Perception is treated as a graded phenomenon, and it is possible for the perceptual system to entertain several candidate interpretations for the same stretch of input simultaneously. The model is characterized by the use of continuous activation and competition processes in place of discrete decisive processes such as segmentation and labeling. In the extensive discussion McClelland and Elman give of their model, they show how it permits forward and backward interactions, how it is capable of processing successive words simultaneously, and how, like humans, it sometimes cannot identify a word until it has heard part of the next word. It is interesting to note that the trace model, unlike other models of word recognition, can account for the results obtained in this study.

Recently, Grosjean and Gee (in press) have proposed another view of word recognition that can also explain the nonsequential aspect of word recognition. In it, the output of the speech perception system (which may be in the form of a partial representation of weak and strong syllables) is processed in two different ways. On the one hand, stressed syllables are isolated and they (and only they) are used to initiate a lexical search. On the other hand, and concurrently, the weak syllables located on either side of the stressed syllable (functors, affixes, weak syllables of content words) are identified by means of a pattern-recognition-like analysis and with the help of the listener's knowledge of phonotactic and morphophonemic rules. The constant interaction of these two types of analysis—the analysis conducted on the weak syllables and the lexical search taking place with the stressed syllable—and the ever present information of other sources of information (the listener's knowledge of the world and of the immediate situation, the rules of the language, the discourse, etc.) will allow the appropriate segmentation of the speech stream into a string of units of representation (in this case, the lexical items in the lexicon).

According to this view, which is very similar to that proposed independently by Cutler (1976) and Bradley (1980), word recognition is not a word-by-word, strictly left-to-right process. Rather, it is a process that uses the prosodic structure of the speech stream and that puts particular emphasis on stressed syllables. In addition, the process is very much a feed-forward, feed-back operation, where there are constant adjustments being made to early and/or partial analyses, and constant predictions being made on what is to come.

Whether this view is correct remains to be seen. What is certain, however, is that existing models of word recognition will have to account for experimental findings that do not always correspond to what they predict. The one we have presented here is that words are not always isolated, and much less recognized, before their acoustic offset. Isolation and recognition may take place after the word has been heard, that is, while the listener is hearing the next word or words.

REFERENCES

BOND, Z., & GARNES, S. (1979). Misperceptions of fluent speech. In R. Cole (Ed.), Perception and production of fluent speech. Hillsdale, NJ: Erlbaum.

BRADLEY, D. (1980). Lexical representation of derivational relation. In M. Aronoff & M. L. Kean (Eds.), Juncture. Saratoga, CA: Anma Libri.

BRADLEY, D., & FORSTER, K. (1984). A reader's view of listening (Working paper). Clayton, Australia: Monash University, Department of Psychology.

COLE, R., & JAKIMIK, J. (1978). Understanding speech: How words are heard. In G. Underwood (Ed.), Strategies of information processing. New York: Academic Press.

COLE, R., & JAKIMIK, J. (1979). A model of speech perception. In R. Cole (Ed.), Perception and production of fluent speech. Hillsdale, NJ: Erlbaum.

CUTLER, A. (1976). Phoneme monitoring reaction time as a function of preceding intonation contour. Perception & Psychophysics, 20, 55-60.

FORSTER, K. (1976). Accessing the mental lexicon. In R. Wales & E. Walker (Eds.), New approaches to language mechanism. Amsterdam: North-Holland.

FORSTER, K. (1979). Levels of processing and the structure of the language processor. In W. Cooper & E. Walker (Eds.), *Sentence processing*. Hillsdale, NJ: Erlbaum.

FOSS, D. (1969). Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. *Journal of Verbal Learning & Verbal Behavior*, 8, 457-462.

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

GROSJEAN, F., & CORNU, E. (1985). *L'importance de la marque du genre dans la reconnaisance des substantifs français*. Manuscript in preparation.

GROSJEAN, F., & GEE, J. (in press). Prosodic structure and spoken word recognition. *Cognition*.

HOWES, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America*, 29, 296-305.

KLATT, D. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, 7, 279-312.

KUČERA, F., & FRANCIS, W. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

LUCE, P. (1984). *A computational analysis of optimal discrimination points in auditory word recognition*. (Research on Speech Perception Progress Report, No. 10, pp. 1-8). Bloomington: Indiana University, Department of Psychology.

MARCUS, S. (1984). Recognizing speech: On the mapping from sound to word. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X: Control of language processes*. Hillsdale, NJ: Erlbaum.

MARSLEN-WILSON, W. (1984). Function and process in spoken word recognition: A tutorial review. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X: Control of language processes*. Hillsdale, NJ: Erlbaum.

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

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

MCCLELLAND, J., & ELMAN, J. (1985). *The trace model of speech perception* (Working paper). Pittsburgh: Carnegie-Mellon University, Department of Psychology.

MEHLER, J., SÉGUI, J., & CAREY, P. (1978). Tails of words: Monitoring ambiguity. *Journal of Verbal Learning & Verbal Behavior*, 17, 29-35.

MORTON, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76, 165-178.

MORTON, J., & LONG, J. (1976). Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning & Verbal Behavior*, 15, 43-51.

NAKATANI, L., & DUKES, K. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, 62, 714-719.

NAKATANI, L., & SCHAFFER, J. (1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America*, 63, 234-245.

OHMAN, S. (1966). Perception of segments of VCCV utterances. *Journal of the Acoustical Society of America*, 40, 979-988.

PISONI, D. (1984). *Acoustic-phonetic representations in word recognition* (Research on Speech Perception Progress Report, Vol. 10, pp. 129-152). Bloomington: Indiana University, Department of Psychology.

POLLACK, I., & PICKETT, J. (1963). The intelligibility of excerpts from conversation. *Language & Speech*, 6, 165-171.

POLLACK, I., & PICKETT, J. (1964). Intelligibility of excerpts from fluent speech: Auditory vs. structural context. *Journal of Verbal Learning & Verbal Behavior*, 3, 79-84.

RUBENSTEIN, H., & POLLACK, I. (1963). Word predictability and intelligibility. *Journal of Verbal Learning & Verbal Behavior*, 2, 147-158.

SALASOO, A., & PISONI, D. (1985). Sources of knowledge in spoken word identification. *Journal of Memory & Language*, 24, 210-231.

SWINNEY, D. (1982). The structure and time-course of information interaction during speech comprehension: Lexical segmentation, access and interpretation. In J. Mehler, E. Walker, & M. Garrett (Eds.), *Perspectives on mental representation*. Hillsdale, NJ: Erlbaum.

TYLER, L. (1984). The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics*, 36, 417-427.

TYLER, L., & WESSELS, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception & Psychophysics*, 34, 409-420.

TYLER, L., & WESSELS, J. (1984). *Is gating an on-line task? Evidence from naming latency data* (Working paper), Nijmegen: Max Planck Institute for Psycholinguistics.

NOTE

1. We need hardly stress at this point the difficulty researchers are having in explaining exactly what is meant by the term "word recognition," in what way it differs from "lexical access," at what point in time it occurs during the on-line processing of the speech stream, and so forth. For recent efforts at distinguishing "word recognition" and "lexical access," but with contradictory outcomes, particularly in relation to the order in time of the two operations, see Bradley and Forster (1984) and Pisoni (1984). In this paper, we will not make a fine distinction between these two operations; instead, we will study in greater depth the narrowing-in operation that takes place during the recognition process, and examine the final acceptance phase the candidate goes through before the information that characterizes it in the individual lexicon is used in the construction of the interpretative representation of the message.

APPENDIX

The 40 stimulus words with their prepositional phrase endings. The beginning of every sentence was "I saw the...."

Monosyllabic Words	Polysyllabic Words
..doe in the woods	..president at the convention
..blimp in the sky	..donkey in the stable
..boar in the woods	..trawler out at sea
..pout on the face	..dissident in the prison
..bun in the store	..captain in the cockpit
..dune on beach	..parrot in the zoo
..spout out at sea	..doctor in the hospital
..gull on the beach	..traffic on the highway
..stunt in the show	..factory in the town
..brawl in the bar	..camel in the desert
..pub on the street	..family at the table
..gash on the hand	..partner in the office
..tramp on the street	..difference on the printout
..flute in the box	..candlestick on the shelf
..plum on the tree	..picture in the book
..perch in the lake	..candidate at the convention
..crust on the bread	..parsley in the store
..skit on the stage	..faculty in the department
..snail on the ground	..pharmacist in the drugstore
..pram on the street	..senator in the capitol