Spoken word recognition processes and the gating paradigm

FRANÇOIS GROSJEAN
Northeastern University, Boston, Massachusetts 02115

Words varying in length (one, two, and three syllables) and in frequency (high and low) were presented to subjects in isolation, in a short context, and in a long context. Each word was presented repeatedly, and its presentation time (duration from the onset of the word) increased at each successive pass. After each pass, subjects were asked to write down the word being presented and to indicate how confident they were about each guess. In addition to replicating a frequency, a context, and a word-length effect, this "gating" paradigm allowed us to study more closely the narrowing-in process employed by listeners in the isolation and recognition of words: Some delay appears to exist between the moment a word is isolated from other word candidates and the moment it is recognized; word candidates differ in number and in type from one context to the other; and, like syntactic processing, word recognition is strewn with garden paths. The active direct access model proposed by Marslen-Wilson and Welsh is discussed in light of these findings.

Two aims underlie this study. The first is to present an experimental paradigm that can be used to study the on-line processing of spoken language along with such existing paradigms as phoneme monitoring, detection of mispronunciations, and shadowing of continuous speech, and the second aim is to extend our knowledge of the word recognition process itself. Our paradigm—the gating paradigm—entails presenting a spoken language stimulus repeatedly and increasing its presentation time (duration from onset) at each successive pass. Depending on the questions at hand and the level of analysis, the stimulus can range from a simple CV syllable to a complex sentence and the presentation time (or gate) can also be made to vary. The subject's task is to guess the stimulus being presented after each pass and to give a confidence rating based on the guess. Although, in this

paper, we will only use this paradigm to study spoken word recognition, gating can also be used to study such issues as the effect of prosodic cues and perceptual strategies on sentence processing.

The tasks that have been employed to study spoken word recognition range from the early "perception in noise" or intelligibility paradigms, in which the dependent variable is response accuracy (see, e.g., Miller, Heise, & Lichten, 1951; Rubenstein & Pollack, 1963) through various monitoring tasks, including the monitoring of phonemes, syllables, and words, as well as mispronunciations, in which the dependent variables are response accuracy and speed of detection (see, e.g., Cole, 1973; Foss, 1969; Marslen-Wilson & Welsh, 1978; Morton & Long, 1976; Savin & Bever, 1970), to shadowing tasks in which one can study shadowing latencies, distribution of errors, and restorations or repetitions of mispronunciations (Miller & Isard, 1963; Marslen-Wilson, 1973, 1975).

Our first aim in this paper is to show that the gating paradigm is a useful and valid way of studying the processes involved in spoken word recognition. To do this, we will first validate the paradigm by replicat-ing three effects that have been obtained with other paradigms: the word frequency, the context, and the word length effects. The frequency effect—in which high-frequency words are processed more accurately and more rapidly than low-frequency words—has been found in the auditory modality with perception in noise tasks (Howes, 1957; Rosenzweig & Postman, 1957; Rubenstein & Pollack, 1963) and in phoneme monitoring tasks (Foss, 1969). Rubenstein and Pollack (1963), for example, report that for each 10-fold increase in word frequency, the S/N ratio may be reduced 3-4 dB to achieve a given level of intelligibility.

This effect has been replicated with a number of tasks and consequently has been integrated into a number of models of lexical access: Forster (1976), for example, in his autonomous search model, organizes his bins in the access files according to frequency, and Morton (1969), in his logogen model, gives lower thresholds to logogens corresponding to words of high frequency.

A second effect that has long been mentioned in the word recognition literature is that due to the presence or absence of syntactic and semantic context. This effect has been shown in intelligibility tasks (Miller, Heise, & Lichten, 1951; Rubenstein & Pollack, 1963), phoneme monitoring (Morton & Long, 1976), and detection of mispronunciations (Jakimik, 1979; Marslen-Wilson & Welsh, 1978). As Spoehr (1980) reports, earlier studies showed that violating either semantic rules (selection rules) or syntactic rules (strict subcategorization rules) affected the recognition performance of words (see Miller & Isard, 1963), whereas later studies have used only grammatical speech and have varied the transitional probability between the sentence context and the target word. Thus, for example, Morton and Long (1976), in a phoneme-monitoring task, presented sentences followed by either high- or low-transitional-probability target words and found that phonemes were detected, on the average, 73 msec faster when the target sound was in a high-transitional-probability word item than when it was in a low-transitional-probability word, and Marslen-Wilson and Welsh (1978), in a shadowing task, found that more mispronunciations were corrected (were fluently restored) in words that were constrained by the sentence context than in words not so highly constrained. These results, among others, indicate that context reduces the number of lexical possibilities and hence enhances word recognition. Again, like the frequency effect, the context effect (also called the congruity effect) has been integrated into a number of lexical access models. For example, in the logogen model (Morton, 1969), context increments the logogen so that less acoustic information is needed for the logogen to reach threshold, and in Marslen-Wilson and Welsh's (1978) active direct access model, context (or top-down information) has the effect of accelerating the process of partitioning the initial group (cohort) of word candidates down to a single choice.

Much less is known about the third effect we will study—the word-length effect. Mehler, Segui, and Carey (1978) do report, however, that in a phoneme-monitoring task, longer words are associated with faster detection of word-initial sounds on the following word. The explanation for this given by Cutler and Norris (1979) is that the decision that a short word has ended may not be arrived at until *after* it has ended, that is, after the target has occurred, thus

delaying the monitoring response, whereas the end of a longer, more redundant word, may be identified as it occurs, hence the shorter reaction times. Contrary to the frequency and context effects, little is said about the word-length effect in current word-recognition models, although some (the active direct access model, for example) do indicate that longer words will take more time to recognize than will shorter words—at least out of context.

Our first aim, therefore, will be to show that the point in the gating sequence at which a word is identified correctly by the listener will be influenced by its frequency, its length, and the amount of context that precedes the word. Less acoustic-phonetic information will be needed if the word is short, frequent, or in context, whereas more information will be needed if the word is long, infrequent, and presented in isolation.

In addition to replicating the frequency, the context, and the word-length effects with the gating paradigm, we will also examine the influence that each effect has on the other. In the visual modality, for example, Schuberth and Eimas (1977) and Schuberth and Spoehr (Note 1) found no interaction between context and frequency and concluded that the effects of the two variables were additive, whereas Becker (1979) did find an interaction between semantic context and frequency—the difference between lexical decision times of high- and low-frequency words was greater in the "low" context condition than in the "high" context condition. We will examine this question—unresolved in the visual modality and not studied in the auditory modality—with the gating paradigm.

As we noted above, a second aim of this study will be to extend our knowledge of the temporal aspects of the word-recognition process itself. We will do this in three ways: first, by examining the amount of acoustic-phonetic information needed from the onset of the word to the point at which it is isolated from other words, that is, guessed correctly by the listener (we will refer to this point as the "isolation point" rather than the "recognition point," as it is not a priori clear that isolation in the gating sequence necessarily corresponds to recognition in such tasks as word monitoring or shadowing); second, by analyzing the confidence ratings given at the isolation point as well as at the last presentation (i.e., when the whole word is presented); and third, by studying the erroneous guesses made up to the isolation point. We will compare the times needed to isolate a word with the recognition data published by Marslen-Wilson (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, Note 2). The latter showed that word recognition takes place within 175-200 msec of the onset of the word, that in context only about half or less of the acoustic signal corresponding to

a word is needed by the listener, that word candidates will remove themselves as more input is heard or more context given, and that, because of top-down effects, recognition decisions can be made *before* sufficient bottom-up information could have accumulated to allow a single word candidate to be selected on the basis of bottom-up information alone. We will also attempt to demonstrate that when a word is isolated, the listener's confidence may not always be very high, signifying thereby that the isolation point of a word may not necessarily be the same as its recognition point. This point will occur some time later, when the listener's confidence in the isolated word reaches some criterion level. And we will try to show that word candidates differ in number and in type from one context to the next and that, like syntactic processing, word recognition is strewn with garden paths. We will conclude the study with a general discussion of how these findings can best be integrated into an existing model of word recognition—namely, the active direct access model of Marslen-Wilson and Welsh (1978).

METHOD

Subjects

Twenty-four undergraduates, with no reported speech or hearing defects, served individually in sessions lasting 2 h.

Materials

Forty-eight nouns were chosen from the Kučera and Francis (1967) word-frequency list. Half the nouns were of low frequency (count of 1 on the list), and the other half were of high frequency (count of 30 or above on the list). Both groups of 24 words were made up of an equal number of one-, two-, and three-syllable words (all two- and three-syllable words were stressed on the first syllable). Thus, there were eight words in each cell of a frequency (2) by syllable (3) matrix, totaling 48 words. Care was taken to choose words beginning with a stop consonant or a fricative and to pair low- and high-frequency words on the first consonant and vowel whenever possible, for example, *bog* and *box, camel* and *captain, precipice* and *president*, etc.

Each word was embedded in the second clause of a two-clause sentence. For example, (1) Lost in the Scottish Highlands, he walked into a *bog* almost at once; and (2) At the zoo, the kids rode on the *camel* for a while. In the second clause, the stimulus word was always the noun in an object NP or in a prepositional phrase and was preceded by a subject NP and a verb which, together, never exceeded five syllables (e.g., *he walked into a* or *the kids rode on the*). Also, the stimulus word was always followed by an adverbial or prepositional phrase (e.g., *almost at once, for a while*) so as to make sure that it did not end the clause.

To determine the contextual importance of the information preceding the stimulus words, an independent group of judges was asked to read the sentences without the stimulus words and to fill in the missing words. A first group of 10 judges received only the beginning of the second clause (e.g., *He walked into a . . . ;* this will be referred to as the short context), whereas a second group of 10 judges was given the first and the second clause (e.g., *Lost in the Scottish Highlands, he walked into a . . . ;* this is the long context).

The responses were scored using the 1-4 rating scale proposed by Marslen-Wilson and Welsh (1978): 1 for a word identical to the stimulus word, 2 for a synonym, 3 for a related word, and 4 for an unrelated word. Responses were pooled across subjects

for each word and were then used to determine whether the short and long contexts were appropriate. A score that fell between 3.4 and 4.0 was deemed adequate for the short context, whereas a score that fell in the 2.4-3.0 range was necessary for the long context. The short and long contexts that did not meet these criteria were modified and retested with another 10 judges, until all contexts for all 48 words fell within the two required ranges.

The 48 complete sentences (e.g., Sentences 1 and 2) were then 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 manipulated by means of a computer-controlled cursor (Huggins, Note 3). First, the onset of the first and second clause and the beginning and end of the stimulus word were marked off. (As each word started with either a plosive consonant or a fricative, it was relatively easy to mark its onset: the release burst of the word initial plosive or the start of frication for the fricatives.) To measure the duration of the word, the time between onset and offset of the word was displayed on the oscilloscope screen to the nearest 100 μsec. The accuracy of each measurement was estimated to be within $\pm$5 msec (Sorensen, Cooper, & Paccia, 1978). Markers were then placed every 30 msec on the stimulus word by means of the cursor.

The recording of the test trials could now take place so as to meet the needs of the three experimental conditions: no context (NC), short context (SC), and long context (LC). For the no-context condition, only the stimulus word was output from the computer and recorded in presentations of increasing duration (+30 msec for each presentation) with an interstimulus interval of 7 sec. Thus, for a particular word, the first presentation corresponded to the first 30 msec of the word, the second presentation to the first 60 msec, the third presentation to the first 90 msec, and so on, until the whole word had been presented. Each word therefore was represented by a set of presentations, the size of the set depending on the length of the word [e.g., the word *smoke* lasted 330 msec and was therefore represented by a set of 11 presentations (or gates)]. When word durations were not an exact multiple of 30, the last presentation was incremented by the difference between the second-to-last gate time and the duration of the word. Each of the 48 stimulus words was gated in this manner, and the resultant 48 word sets were recorded randomly on tape.

For the short-context condition (SC), exactly the same procedure was used except that each presentation was preceded by the beginning of the second clause, that is, the preceding five syllables (e.g., *He walked into a* for the word *bog* or *The kids rode on the* for the word *camel*). Thus, in this condition, exactly the same sets of presentations were recorded as in the NC condition except that each presentation (or gate) was composed of a short context and a segment of the stimulus word. The interstimulus interval between presentations within sets was again 7 sec. Finally, for the long context condition, the first and second clauses preceded the stimulus word (e.g., *Lost in the Scottish Highlands, he walked into a* for the word *bog* or *At the zoo, the kids rode on the* for the word *camel*).

Below we illustrate the first three presentations (or gates) for the word *camel* as heard by the subjects in each of the three context conditions. We have lengthened the stimulus word in order to facilitate the illustration. The continuous black lines represent what the subjects heard in each of the three context conditions.

No context	c a m e l	
Gate 1	—	30 msec
Gate 2	—	60 msec
Gate 3	—	90 msec

Short context	The kids rode on a c a m e l	
Gate 1	_____	
Gate 2	_____	
Gate 3	_____	
etc.		

Long context   At the zoo, the kids rode on a c a m e l

Gate 1   _____
Gate 2   _____
Gate 3   _____
etc.
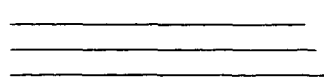
It should be noted that the prepositional or adverbial phrases following the stimulus words never appeared on the experimental tapes (they were inserted in the original recording so that the stimulus word was not in a sentence final position) and that the random order of the 48 sets was identical across the three experimental conditions.

Procedure

The 24 subjects were assigned randomly to one of the three experimental conditions (NC, SC, or LC) and were run individually. The subjects were instructed to listen to each presentation within each word set, to write down the word they thought was being presented, and to indicate with a vertical slash on a continuous scale (very sure—very unsure) how confident they were about the guess. They were asked to give a response (word and confidence rating) after each presentation, even though at first they might feel very unsure about the stimulus word.

Data Analysis

Response sheets were examined to determine the point at which each subject correctly guessed the stimulus word and did not change his/her guess subsequently. The duration of the gate at that point was adopted as the isolation time and was tabulated. When a subject did not isolate a stimulus word before the end of a set, he/she was assigned the duration of the word as the isolation time. An analysis of variance was then run with length, context and frequency as fixed effects and subjects and words as random effects (Clark, 1973).

The response sheets also yielded the subjects' confidence ratings at both the point of isolation of a word and at the last presentation of a word set. As no rating score could be assigned to subjects who did not isolate a word, mean ratings were obtained across subjects for each of the 48 words in each of the three context conditions. An analysis of variance was then run for confidence ratings, with length, context, and frequency as fixed effects and words as a random effect. Finally, the erroneous guesses given before the isolation point were analyzed with respect to frequency of occurrence, number of times proposed, total number of guess types, and phonotactic composition.

RESULTS AND DISCUSSION

Isolation Points

Figure 1 presents the mean isolation time of the test words as a function of the context preceding the words, the length of the words in syllables, and the frequency of occurrence of the words. Main effects were found for all three variables. First, as context becomes more constraining, the mean isolation time of words decreases substantially. We found a global mean of 333 msec for the no-context condition (NC), 245 msec for the short context (SC), and 153 msec for the long context (LC) [$F'(2,44) = 108.61$, $p < .01$]. Second, as the frequency of the words increases, isolation times decrease: low-frequency words are isolated in 274 msec on the average as compared with 213 msec for high-frequency words [$F'(1,44) = 8.21$, $p < .01$]. And third, as the length of the word in-
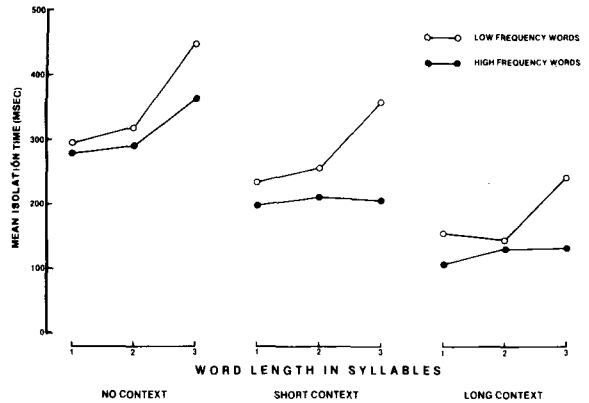

Figure 1. Mean isolation times for 48 words of varying length and frequency when presented in a no-context, a short-context, and a long-context condition.

creases, so does its isolation time: one-syllable words were isolated in 212 msec on the average, two- and three-syllable words in 226 and 292 msec, respectively [$F'(2,44) = 5.24$, $p < .01$]. A posteriori tests on context showed significant differences between the NC and SC conditions and the SC and LC conditions [Tukey HSD, Kirk (1967), $p < .01$], and similar tests on length showed a significant difference between two- and three-syllable words ($p < .01$), but not between one- and two-syllable words ($p > .20$).

Before turning to the interactions, we illustrate in Figure 2 the main effects with specific examples. As can be seen, the one-syllable words (*gull* and *gun*) are isolated more rapidly than the two-syllable words
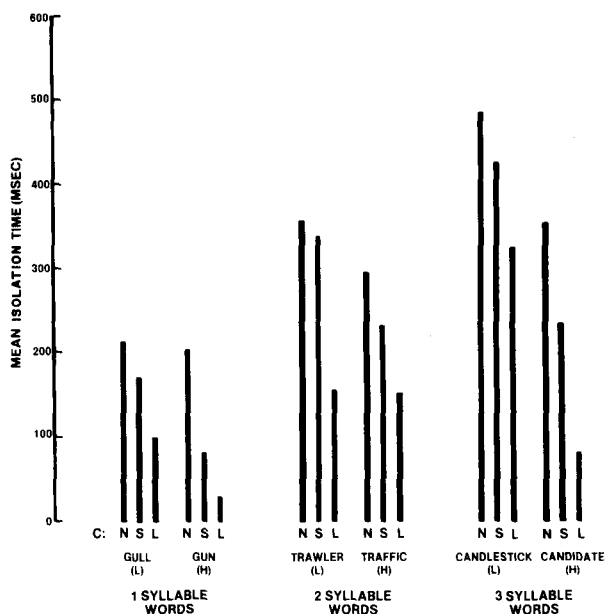

Figure 2. Mean isolation times for six exemplars of the 48 words used in the experiment. For each length group (one, two, or three syllables), we present the mean isolation times of a low (L) and a high (H) frequency word in each of three contexts: no context (N), short context (S), and long context (L).

(*trawler* and *traffic*), which in turn are isolated more rapidly than the three-syllable words (*candlestick* and *candidate*). Within each syllable group, low-frequency words take more time to be isolated than do high-frequency words and, for each word, context has a marked effect on isolation time.

The results obtained with the gating paradigm are similar to those of other spoken word recognition studies which have reported effects of frequency (Howes, 1957; Rubenstein & Pollack, 1963), context (Marslen-Wilson & Welsh, 1978; Morton & Long, 1976), and word length (e.g., Mehler, Segui, & Carey, 1978). It is interesting to note that, until now, no other study has examined all three independent variables together and that the three effects have never been separated out within the same paradigm. Thus, our results do seem to indicate that the gating paradigm is a valuable method of studying the processes underlying spoken word recognition.

The experimental design allowed us not only to determine the effect of length, frequency, and context, but also to examine the influence each effect has on the others. We found no significant interactions for Length by Frequency [$F'(2,42) = 1.93$, n.s.], Frequency by Context [$F'(2,60) = 1.75$, n.s.] [thus replicating the visual word recognition study by Schuberth and Eimas (1977) and contradicting that by Becker (1979), which reported a significant interaction], or Length by Frequency by Context [$F'(4,73) = .95$, n.s.]. The only significant interaction was for Length by Context [$F'(4,79) = 2.57$, $p < .05$]: As the context becomes more constraining, the length of the word plays less of a role in the time it takes to isolate the word. Thus, the difference between the mean isolation times for one- and three-syllable words was 118 msec in the NC condition but only 57 msec in the LC condition.

It is interesting to compare our isolation times with the word-recognition data obtained by Marslen-Wilson and his colleagues. Although we do not claim that isolation times obtained with the gating paradigm are necessarily equivalent to recognition times in word monitoring and shadowing tasks, our data are surprisingly similar to those obtained by Marslen-Wilson. Marslen-Wilson and Welsh (1978), for example, state that words in isolation can be identified approximately 300 msec after their onset. Our results confirm and refine this estimate: Averaging across frequency, one-syllable words in the NC condition were isolated, on the average, 289 msec after their onset, two-syllable words after 306 msec, and three-syllable words after 406 msec. Marslen-Wilson (Note 2) also reports that word-monitoring and speech-shadowing studies suggest that spoken words in normal prose contexts can usually be recognized within 175-200 msec after their onset. Our results show that, in the LC condition, isolation times range

from a mean of 105 msec (for one-syllable high-frequency words) to a mean of 245 msec (for three-syllable low-frequency words). Of course, as we have seen, the actual isolation times will depend, among other things, on the accompanying context (e.g., the word *doctor*, whose duration was 346 msec, was isolated in 30 msec by all eight subjects in the context *Because of her problem, the person saw a*), on the frequency of the word, and on the length of the word, in addition to a variable not studied in this study, the phonotactic configuration of the word. Jakimik (1979), for example, found shorter latencies to detect mispronunciations in words containing infrequent first syllables (e.g., spaghetti, vampire, etc.) than in words with frequent first syllables (e.g., inventor, convertible, etc.).

Marslen-Wilson (Note 2) also reports that words in normal contexts can be identified when perhaps only half or less of the acoustic signal corresponding to that word has been heard by the listener. Our results confirm this. Averaging across length and frequency, we found that in the LC condition only 37% of the word was needed to be isolated, that in the SC condition 60% of the word was required, and that in the NC condition as much as 83% of the word was needed. An analysis of variance of the percent of the word needed for isolation showed main effects for context [$F(2,84) = 186.8$, $p < .01$] and frequency [$F(1,42) = 9.61$, $p < .01$]. As might be expected, no main effect was found for length [transforming the isolation time data into a percentage of the total word neutralizes the length effect; $F(2,42) = 1.93$, n.s.], and no two- or three-way interactions were found.

Another point made by Marslen-Wilson (Note 2) concerns the interaction of top-down and bottom-up information in word recognition. His statement that recognition decisions can be made, in normal context, before sufficient bottom-up information could have accumulated to allow a single word candidate to be selected on the basis of bottom-up information alone is also true of our isolation times. As we have seen, normal context (LC) has the effect of reducing by about one-half the amount of bottom-up information needed for isolation when compared with the NC condition.

Finally, Marslen-Wilson and Welsh (1978) state that word recognition will be affected by the length of the word and that the strength of the effect will diminish as contextual constraints increase. Comparable trends are shown not only by our isolation time data (see Figure 1), but also by three Pearson product-moment coefficients of correlation calculated between the 48 mean isolation times within each context and the duration of these stimulus words. For the NC condition, $r = .88$, which indicates a strong relationship between the duration of the word and

the time it takes to recognize it. This correlation coefficient drops, however, in the SC condition ($r = .69$) and in the LC condition ($r = .60$), thereby confirming the prediction made by Marslen-Wilson and Welsh (1978).

As we will see in the next section, a distinction will need to be made between the isolation point and the recognition point of a word. The fact that the recognition data reported by Marslen-Wilson are so similar to the isolation data reported here will therefore have to be explained. It could be that, in both the shadowing and word-monitoring tasks, subjects are under time pressure to respond and are therefore isolating the words but not fully recognizing them before responding (the published data by Marslen-Wilson would therefore be isolation data) or that the similarity between the two sets of data—word monitoring and shadowing on the one hand, gating on the other— is only fortuitous. Only further research into the cognitive operations underlying all three tasks will throw light on this.

Confidence Ratings

Figure 3 presents the mean confidence ratings given by the subjects at two points: first, at the point at which they isolated the test words and, second, at the last presentation of the word (that is, when 100% of the word had been presented). Examining, first, the ratings at the isolation point, we notice that the patterns for the NC condition are very different from those for the SC and LC conditions. In the NC condition, the subjects became more confident (approached 1) as the word increased in length, whereas, in the SC

Figure 3. Mean confidence ratings for the 48 words varying in length and frequency when presented in three context conditions. Ratings were analyzed at the isolation point of the word (discontinuous lines) and at the last presentation of the word (that is, when the whole word is presented; continuous lines).

and LC conditions, no reliable pattern appears to emerge. However, the subjects do seem to be less confident about their guesses in these two conditions when compared with the NC condition. This is confirmed by an analysis of variance that revealed a main effect for context [$F(2,84) = 55.99$, $p < .01$]. This effect is due, however, to the difference between the NC and SC conditions and the NC and LC conditions ($p < .05$, Tukey HSD, Kirk, 1967) and not to that between the SC and LC conditions ($p > .20$). No main effect was found for frequency and word length [$F(1,42) = 2.83$, n.s., and $F(2,42)$, n.s., respectively], but a Context by Length and Context by Length by Frequency interactions were obtained [$F(4,84) = 5.37$, $p < .01$, and $F(4,84) = 2.59$, $p < .05$, respectively]. In an attempt to interpret the three-way interaction, it was broken down into its components two ways. Analyses of variance, holding context constant, revealed no frequency effect in any of the contexts [$F(1,42) = 1.37$, 3.76, and .91, n.s. for NC, SC, and LC, respectively] and only in the NC condition was a word-length effect obtained [$F(2,42) = 8.83$, $p < .01$, as opposed to $F(2,42) = .82$ and .00, n.s., for the SC and LC conditions, respectively]. The Length by Context interaction must be interpreted in terms of this three-way interaction.

From the above, we can conclude that subjects feel less confident about words in context at the point at which these are isolated than they do in a no-context condition (NC) and that, in the latter case, the longer the word, the more confident the subjects are when they isolate it. The first finding leads us to ask at which point in time subjects feel as confident about a word in the NC condition as they do about a word in the SC and LC conditions. Could it be that words are isolated more quickly in context than out of context (see Figure 1) but that as much acoustic (or bottom-up) information is needed in and out of context to reach an equal confidence rating for a given word? To answer this question, we took the mean confidence rating obtained for each word at the isolation point in the NC condition and determined for each subject how much acoustic information was needed in both the SC and LC conditions to reach an equal confidence rating. To do this, we found the point in time at which the subject's confidence rating changed from being above the mean NC rating to one at or below the rating. Results were pooled across subjects, and these new isolation times (now controlled for equal confidence ratings in all three conditions— NC, SC, and LC) were submitted to analyses of variance. Main effects were found for context [$F(2,84) = 142.22$, $p < .01$], length [$F(2,42) = 11.23$, $p < .01$], and frequency [$F(1,42) = 5.36$, $p < .05$], but no interactions reached significance. Words were isolated in 333 msec in the NC condition, and in 312 and 229 msec in the SC and LC conditions, respectively. One-
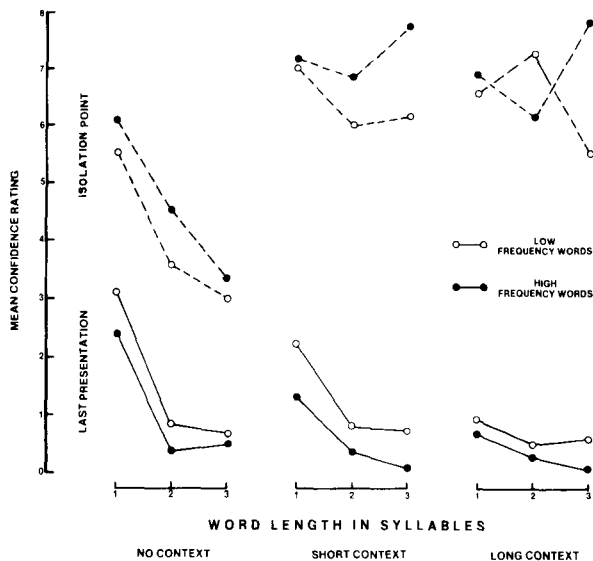
syllable words were isolated in 249 msec, two-syllable words in 271 msec, and three-syllable words in 355 msec, and, finally, low-frequency words took 314 msec to isolate, whereas high-frequency words took only 269 msec. In other words, when the mean ratings in the NC condition are used to determine the points in time at which identical ratings are obtained in SC and LC, isolation times are still dependent on the amount of context accompanying the word, the length of the word, and the frequency of the word. The only difference is that SC and LC isolation times are now closer to the NC times. Thus, to answer our question, the listener does not need as much acoustic information in context as he/she does out of context to reach an equal confidence rating for a given word. In context, words are still isolated more rapidly than are words out of context, even when confidence ratings are held constant.

As stated at the beginning of this section, we also examined the confidence ratings given by subjects to the responses they made to the last presentation in each word set, that is, when the whole word was presented (see Figure 3). A first point to notice is that these ratings are significantly different from those given at the point when the word was isolated (for the two sets of 18 means presented in Figure 3, t = 12.46, p < .01). As might be expected, subjects feel much more confident about a word when it is presented in full than when only part of it is presented. An additional, and perhaps more interesting, point is that confidence ratings at the last presentation are not identical across conditions. As with the isolation times presented in Figure 1, an analysis of variance showed main effects for context [F(2,84) = 8.51, p < .01], length [F(2,42) = 15.72, p < .01], and frequency [F(1,42) = 4.79, p < .05] and for one interaction [Context by Length, F(4,84) = 4.32, p < .01]. Thus, as context becomes more constraining, subjects are more confident about their guesses when the whole stimulus word is presented. Similarly, the longer the word and the greater its frequency of occurrence, the more confident they feel about their final responses.

Three main results emerge, therefore, from the study of confidence ratings. First, the isolation of a word can take place without the listener's feeling very confident about the candidate. Second, the subject's confidence in the proposed candidate increases between the point of isolation and the end of the word. And third, even after the whole word has been heard, the listener's confidence in the candidate proposed is rarely perfect and depends on such factors as the frequency of the word, its length, and the amount of information brought to the isolation and recognition processes by the surrounding context. These results led us to propose the notion of isolation point, that is, the point at which the listener has isolated a candidate but may still feel quite unsure about it. He or she

will therefore continue to monitor the acoustic-phonetic information until some criterion level of confidence is reached and the word is accepted or recognized. This distinction between the isolation point and the recognition point of a word may have to be integrated into a theory of word recognition.

The Word Isolation Process

The gating paradigm not only allows us to determine how much of the stimulus word is needed to be distinguished from other candidates and how confident subjects are about their answers; it also enables us to have an insight into the word-isolation process itself. This is done by analyzing the erroneous guesses subjects make before the isolation point of the stimulus word. In the following section, we will examine the word candidates proposed before isolation and will attempt to extract general patterns of word isolation. It is important to note that two underlying assumptions will be made throughout this section. The first is that we can learn something about the word-isolation process itself by examining responses across subjects and inferring from these the path followed by the individual listener. And the second assumption is that the word candidates proposed by subjects on the basis of gated information (the first 90 to 120 msec of the word, for example) are similar to those that would be available were we able to tap into the word-isolation process as it takes place online in normal listening. Only additional evidence from other paradigms will allow us to assess the validity of these assumptions.

Figure 4 presents the candidates proposed for the word *parrot* in each of the three context conditions. Candidates proposed at only one presentation (or gate) duration are depicted with a dot; those that are proposed over two or more presentations are depicted by continuous lines. The number of subjects proposing a particular candidate is represented by the thickness of the line—the more subjects, the thicker the line. As we can see in the left-hand side of the figure, in the NC condition the isolation process can rely only on acoustic-phonetic information and the candidates at each gate clearly reflect this. Some of the first guesses are random (for example, *fit*, *sit*, *rip*, etc.), but, by the second presentation (+60 msec of the word), every subject has heard the initial stop consonant (/p/ in *pen*, *pale*, etc.). Then, by the third gate (+90 msec), the vowel starts coming in—all subjects realize now that it will be a front vowel (*pen*, *pet*) and two of them have already identified the correct vowel (e.g., *pack*). It is at this gate (+90 msec) that one subject already proposes the correct word: *parrot*. As more of the stimulus word is given, the subjects now narrow in on the /r/ [five of the eight subjects have included it in their guesses by Gate 4 (+120 msec)] and then, with the next few
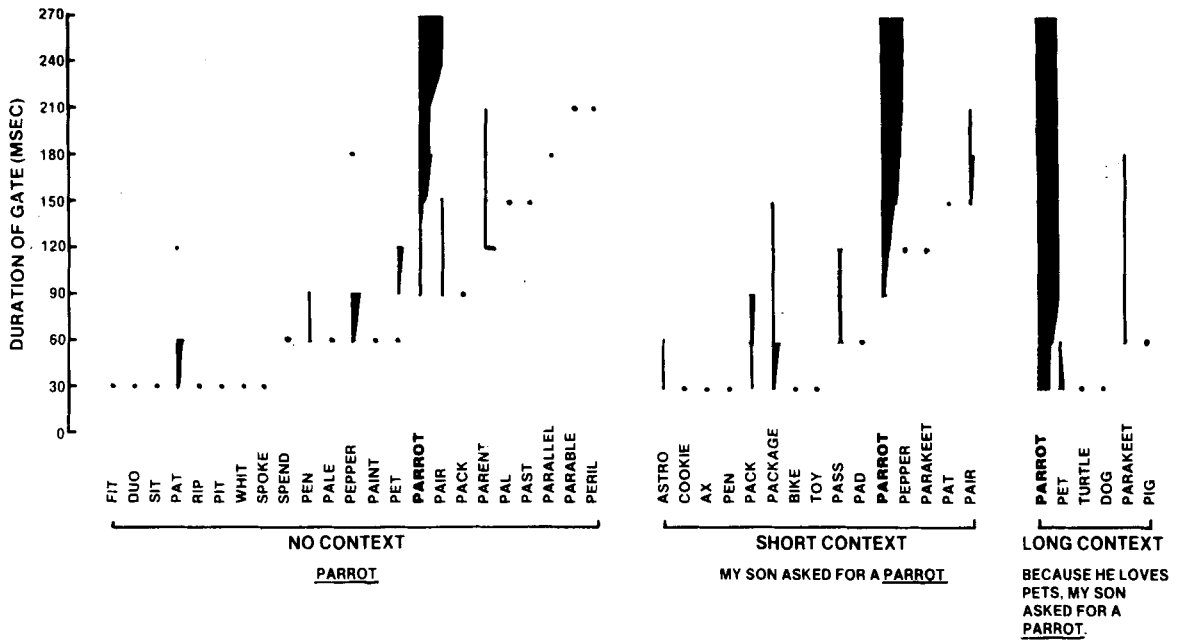
Figure 4. The candidates proposed for the word *parrot* in each of the three context conditions. Candidates proposed at only one presentation (or gate) duration are depicted with a dot; those proposed over two or more presentations are depicted by continuous lines. The number of subjects proposing a particular candidate is represented by the thickness of the line—the more subjects, the thicker the line.

presentations, the last vowel and consonant /ɔt/ are perceived. By the seventh gate (+210 msec), all but three subjects have isolated *parrot* (the others propose *parent*, *parable* and *peril*), and by the eighth gate (+240 msecs), all subjects propose *parrot* as the stimulus word.

In the SC condition, the word *parrot* is now preceded by the context *My son asked for a*, which gives both acoustic-phonetic and syntactic/semantic information. On the one hand, the word is preceded by rich phonetic information: prosody, rate, vowel transition of the preceding /ə/, etc., and on the other hand, the listener knows that the word is a noun which stands for an item that can be asked for by one's son. This top-down information speeds up the isolation process: The first consonant is identified more rapidly, there are fewer candidates overall (15 as opposed to 23 in the NC condition), and the stimulus word is isolated more quickly by the majority of subjects; by the fifth gate (+150 msec), six of the eight subjects have proposed *parrot*.

In the LC condition, the semantic information is much richer (*Because he loves pets, my son asked for a*) and helps to narrow down the possible candidates semantically (the word, in all likelihood, is going to reference a pet). The constraints are now so strong that very little bottom-up information is needed for subjects to propose *parrot*: Four subjects need only the first 30 msec of the word to isolate it, and, by the third gate (+90 msec), seven of the eight subjects

have guessed the word correctly.

It is interesting to note that candidates are very different from one context to the next: only five erroneous candidates are proposed in both the NC and SC conditions (out of a total of 36 words), only 1 out of 27 candidates is common to the NC and LC conditions, and only 1 out of 19 appears in both the SC and LC conditions. This clearly shows that the interaction of top-down and bottom-up information helps define a set of candidates that is unique to that interaction and that a change in either the semantic/syntactic information or the acoustic-phonetic information will change the set of possible candidates. It is, thus, very difficult to predetermine which words will be candidates for a stimulus word when it is presented in context, whereas this is relatively easy in the NC conditions, in which the inherent characteristics of the word such as frequency, length, phonotactic configuration, etc., will play an important role in determining the possible candidates.

In Figure 5, we present a second example of the route followed by a word (the one-syllable word *gull*) in three context conditions. In the NC condition, we notice once again, a phonetic narrowing-in on the word: after the first 30 msec, three out of the eight subjects propose words beginning with a plosive consonant, and, after the second gate (+60 msec), all word candidates begin with a plosive (four of them with a /g/) and seven subjects make it follow by a back vowel. It is interesting to note that the
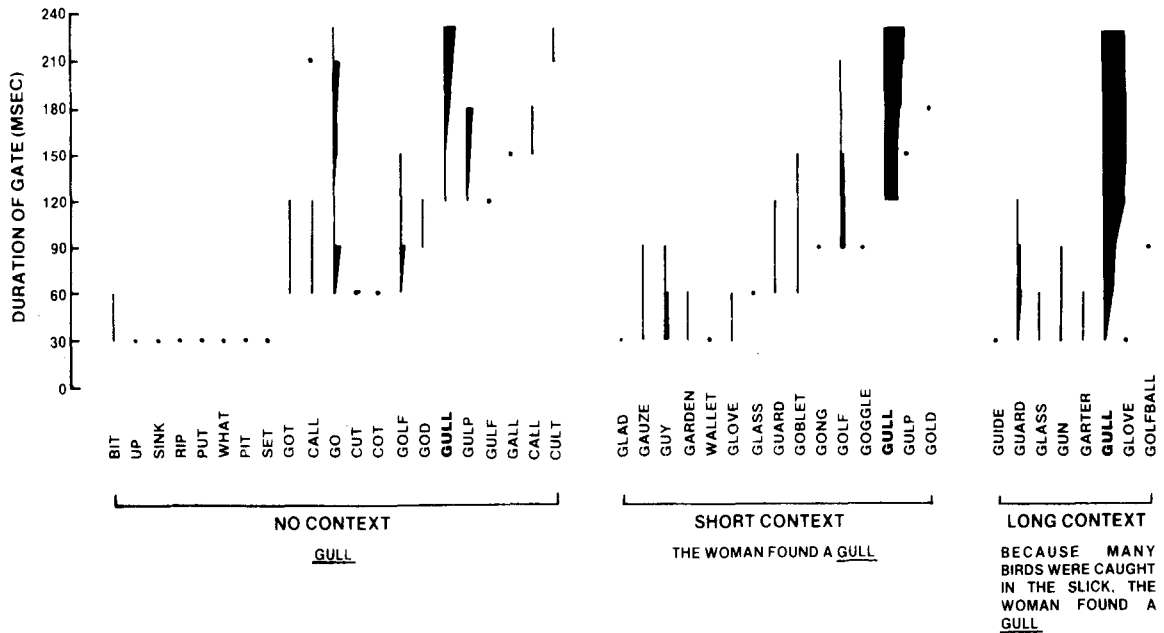
Figure 5. The candidates proposed for the word *gull* in each of three context conditions (see Figure 4).

stimulus word itself (*gull*) is late in coming in as a candidate (+120 msec) and that, at the end of the word (+240 msec), half the subjects still have not isolated it [although three of these do propose words (e.g., *gulf, gulp*) that share a number of phonetic traits with *gull*).

The SC condition brings additional acoustic-phonetic and prosodic as well as syntactic and semantic information to the isolation process (the word to be isolated is an object that can be found by a woman), and, not surprisingly therefore, we find that the first consonant is identified more rapidly (seven subjects hear a /g/ in the first 30 msec, as opposed to none in the NC condition). In addition, fewer candidates are proposed at the first gate (6 instead of 8 in the NC condition), fewer candidates are proposed overall (15 instead of 21), the first proposal of *gull* itself is made after 120 msec (as in the NC condition), but this time by half the subjects, and all eight subjects have isolated *gull* by the last presentation (instead of only four in the NC condition). And, again, as for *parrot*, very few candidates in this condition (2 out of 30) are also candidates in the NC condition. Finally, in the LC condition [in which the context constrains the semantic set to an animal (bird) or an object that can be found in the slick], the stimulus word *gull* is a candidate from the very beginning. Fewer candidates are proposed overall, and, by the fifth gate (+150 msec), all eight subjects have guessed the stimulus word correctly.

From these error patterns and those found for the other stimulus words, we propose the following main points concerning the isolation process in this task. First, in the NC condition, the narrowing-in on the

stimulus word is based in large part on the phonetic information available. Thus, in addition to the length of the word and its frequency of occurrence, the consonants and vowels involved and their organization into syllables (their phonotactic configuration) will play an important role in how quickly the word will be isolated. For example, an analysis of the consonants proposed in the guesses after the first gate (+30 msec) showed that some consonants are much more difficult to perceive in that time span than others. While 52% of guesses for a word beginning with /d/ started with that consonant (32% for /p/, 31% for /s/, 30% for /b/), only 17% of the guesses for a word with /k/ as its first consonant started with this consonant (13% for /t/, 6% for /g/) and *no* guess for a word starting with /f/ started with that consonant. However, as soon as a short context is added (in this case this is both prosodic information and syntactic/semantic information), the first consonants are perceived much more accurately in the time span of the first gate (+30 msec). For example, all of the guesses for stimulus words beginning with /g/ now start with this consonant (as opposed to only 6% in the NC condition), 94% for both /b/ and /d/ (as opposed to 30% and 52%, respectively, in NC), and even /f/ is now included in 46% of all guesses starting with that consonant. (Other percentages are 79% for /p/, 73% for /t/, 90% for /k/, 49% for /s/.)

A second point concerning the isolation process is that not all subjects isolate the stimulus word by the last presentation, that is, when it is presented in full (e.g., see *gull* in the NC condition in Figure 6). This is especially noticeable for words in the NC condition (48% of these words were not isolated by all

eight subjects, as opposed to 15% in the SC condition and 2% in the LC condition). This is also the case for words that are short (11% of the one-syllable words were not isolated by all eight subjects across all contexts, as opposed to 6% for two-syllable words and 4% for three-syllable words) and for words that are not frequent (14% for the low-frequency words as compared with 8% for the high-frequency words). This means that listeners probably use the information following the word not only to confirm their final candidate (as we saw with the rating data), but also at times to help them in their actual isolation when the information carried by the word itself is not sufficient.

A third point concerning the isolation process is that stimulus words are rarely candidates at the onset of the word. By taking the gate time needed for each word to be guessed correctly by at least one subject, we were able to determine which factors account for the point in time at which a word becomes a candidate for the subjects taken as a group. By means of an analysis of variance, we found main effects for context [$F(2,84) = 145.5$, $p < .01$], length [$F(2,42) = 4.18$, $p < .05$], and frequency [$F(1,42) = 8.41$, $p < .01$], and a significant Context by Length interaction [$F(4,84) = 3.76$, $p < .01$]. Thus, for example, a stimulus word becomes a candidate (that is, it is proposed by at least one subject) after 245 msec, on the average, in the NC condition but after only 65 msec, on the average, in the LC condition. In addition, one-syllable words are candidates after 135 msec, on the average, but three-syllable words need, on the average, 187 msec of their total time before they are candidates; low-frequency words become candidates after 177 msec, on the average, whereas high-frequency words need only 127 msec. The fact that across subjects a word is not a candidate from the onset of its presentation and that a number of factors affect the moment at which it will become a candidate is an interesting finding which may indicate that some words are, as it were, stronger candidates than others at various times during the isolation process. It would probably be erroneous to say that the stimulus word is not a member of a set of candidates from the onset of the word, but it may not have sufficient weighting to appear as a candidate early in the gating process.

A fourth point that can be made concerning the word-isolation process is that the number of candidates proposed at the onset of the word (in this study, after the first presentation) will depend on the context that accompanies the word (see Figures 4 and 5). An analysis of variance based on the different candidate types after the first gate (+30 msec) shows a main effect for context only [$F(2,84) = 38.6$, $p < .01$]. In the NC condition, there were almost as many candidates as there were subjects (7.5 candidate types,

on the average, for eight subjects), whereas in the SC condition there were 6.08 candidates, on the average, and in the LC condition only 4.73 candidates, on the average. Thus, context, as we have seen repeatedly, helps listeners narrow-in on the stimulus word, and, as a consequence, fewer candidates are involved in the selection process.

A fifth point, very much related to the preceding, concerns the total number of candidate types for each word across all presentations of that word. As illustrated by the isolation patterns of *parrot* and *gull* (Figures 4 and 5), we found that there were fewer candidate types overall in the LC condition (7.6 on the average) than in either the SC condition or the NC condition [13.0 and 26.0, respectively; $F(2,84) = 112.9$, $p < .01$] and fewer candidate types for short than for long words (e.g., there were an average of 14.1 candidate types for one-syllable words, as opposed to 18.4 for three-syllable words [$F(2,42) = 4.10$, $p < .05$]. No main effect was obtained, however, for frequency, although we did find a trend for low-frequency words to be linked to more candidate types (16.4, on the average, as opposed to 14.7 for high-frequency words). This means that the isolation time for low-frequency words is delayed (see Figure 1), not by the proposal of new (but erroneous) candidate types, but by the perseverance (or reiteration) of these types (in other words, by more tokens).

A sixth point concerning the isolation process, as exemplified by Figures 4 and 5, is that candidates are very different from one context to the next. The interaction of top-down and bottom-up information will determine a set of candidates, and a modification of either or both inputs will modify this set accordingly.

A seventh point is that some candidates have a short life span, whereas others are candidates for a longer time [e.g., *go* in the NC condition of *gull* is proposed continuously by up to three subjects from the second gate (+60 msec) until the very last presentation (+240 msec)]. These last two points (when and for how long a word will be proposed as a candidate) can best be explained by such factors as context and the phonotactic configuration of the word. However, we have studied three additional factors that may account for the sometimes unnaturally long life span of a word candidate. These we have labeled the "word from a word," the frequency, and the semantic garden paths.

The "word from a word" garden path. Figure 6 shows two examples of this type of garden path. On the left side of the figure, we have the error pattern for the word *captain* in the NC condition. The first errors reflect a normal phonetic narrowing-in on the first part of the word, but very rapidly a number of subjects follow the same route and propose the word *cap* as the stimulus word, so that by the 7th gate as many as five out of eight subjects have gone down
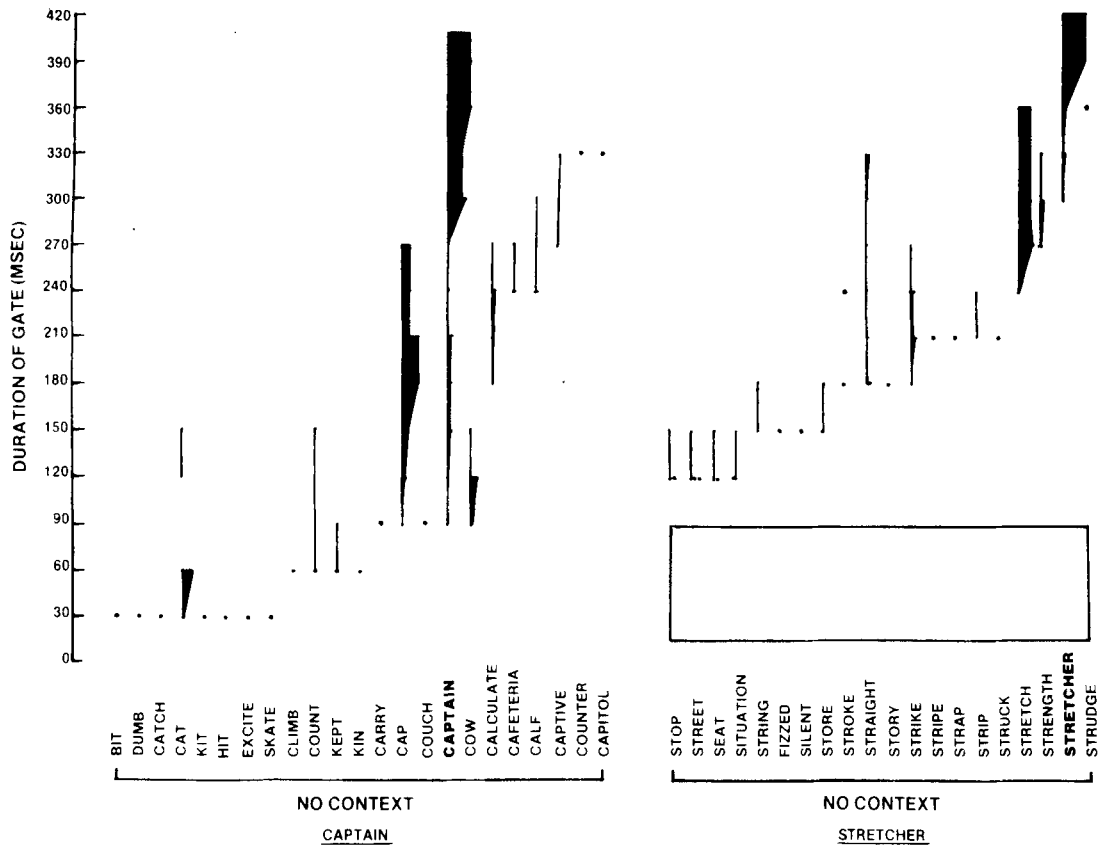
Figure 6. The "word from a word" garden path: Candidates proposed for the stimulus words *captain* and *stretcher* in the no-context condition (see Figure 4). For lack of space, the candidates proposed for *stretcher*, up to the +90 msec gate, are not included.

this garden path. When, at the next gate (+240 msec), they realize that another syllable follows, there is a general dispersion, and, by the 9th gate (+270 msec), as many as six candidate types are proposed by the eight subjects. Then, with longer presentations, they all narrow in on *captain*, and, by the 12th gate (+360 msec), every subject has proposed this word as the stimulus word. Thus, although some transition and coarticulation information could have led the subjects to think that the second syllable (*tain*) would follow the first (*cap*), the majority of the subjects nevertheless followed this "word from a word" garden path. What is interesting is that, in the SC condition, no subject ever proposed the word *cap*, as it did not fit the context *Margaret told the*.

On the right-hand side of Figure 6, we see another example of this garden path. Here, again, a number of subjects are led down the garden path and propose *stretch* for the stimulus word *stretcher* (five out of eight proposals at the 12th gate). But, at the 14th gate (+390 msec), as the following /ə/ comes in, they suddenly all switch to *stretcher*. And, again, in the SC condition (*John went to fetch a*), *stretch* is, of course, never proposed as a candidate.

From this we can propose that if one or any num-

ber of syllables from the onset of the word can stand on their own and make a word, and either the word is presented in isolation or the syntactic and semantic rules are not violated when the word is in context, then that word will be a candidate in the isolation process. For example, we analyzed the 10 two-syllable words (out of a total of 16) that fit this category (e.g., *strain* from *stranger*, *parse* from *parsley*, *pick* from *pickle* and *picture*, *cult* from *culture*, *dock* from *doctor*, etc.) and found that subjects followed a garden path for every one of these words in isolation. However, as soon as the words appeared in context, the first syllable usually disappeared as a candidate. For example, no subject proposed *parse* for *parsley* in the context *Michael chopped up the*. Across all 10 words, 60% of all guesses were the first syllable of the word in the NC condition (*strain* from *stranger*, *dock* from *doctor*, etc.), but this percentage dropped to 24% in the SC condition and 16% in the LC condition. The relatively high percentages in these two context conditions are due to garden paths for *precipice* in the SC condition (a number of subjects proposed *press* for the context *The man looked down the*) and for *picture* in the LC condition (subjects proposed *pick* in the context *Before climbing the North Face,*

Carl examined the). It is interesting to note that *pick* is also the first syllable of *pickle* and, as such, was proposed in the NC condition, but that it was quite naturally never proposed as a candidate in the LC condition (*To her tuna sandwich and french fries, Susan added a*).

From this we can conclude that, more often than not, context will repress the tendency to go down the "word from a word" garden path, but that this bias is quite natural in the NC condition when the syllable configuration of the word permits it.

The word-frequency garden path. The word-frequency effect that has been replicated in this study (see Figure 1) is probably due to a bias listeners have for high-frequency words. As can be seen in Figure 7, before recognizing a low-frequency word (in or out of context) subjects will propose higher frequency counterparts if these fit (to some extent at least) the phonetic, syntactic, and semantic specifications of the utterance. Thus, for example, in the NC condition, as many as five subjects (out of eight) proposed *ball* [frequency (f) of 110 on the Kučera and Francis (1967) list] when presented with 120 msec of *bog*; then, as more information was given, they switched over to *bought* (f = 56); and finally, when the evidence was really overwhelming, they proposed *bog* (f = 1).

Another example concerns the word *trawler* in the context *Stephen worked on a*. Based on the semantic information given by the context, subjects give words only for modes of locomotion (*train, trolley, tram*, etc.). The order in which these candidates are proposed follows the phonetic information given by each successive gate, and, again, the candidates are all of higher frequency than the stimulus word. The first two main candidates are based on the early consonantal information, and both are quite frequent: *train* (f = 82) and *truck* (f = 57). As more phonetic information enters, subjects (with only one exception) chose the candidate *trolley*, which, considering its use in the Boston area, is probably more frequent than is reflected by the Kučera and Francis list (f = 5). It is only when the final /ð/ is heard that the subjects finally backtrack and isolate the low-frequency stimulus word *trawler*.

A further example is the isolation pattern of *smog* (f = 1) in the SC condition *John clearly saw the*. After a variety of relatively strong candidates that fit the acoustic information available in the early gates and the semantic/syntactic context (*sun, sign*), subjects narrow-in erroneously on the word *smoke* and, by the seventh gate (+210 msec), five of the eight proposed this word, whose frequency is much higher than
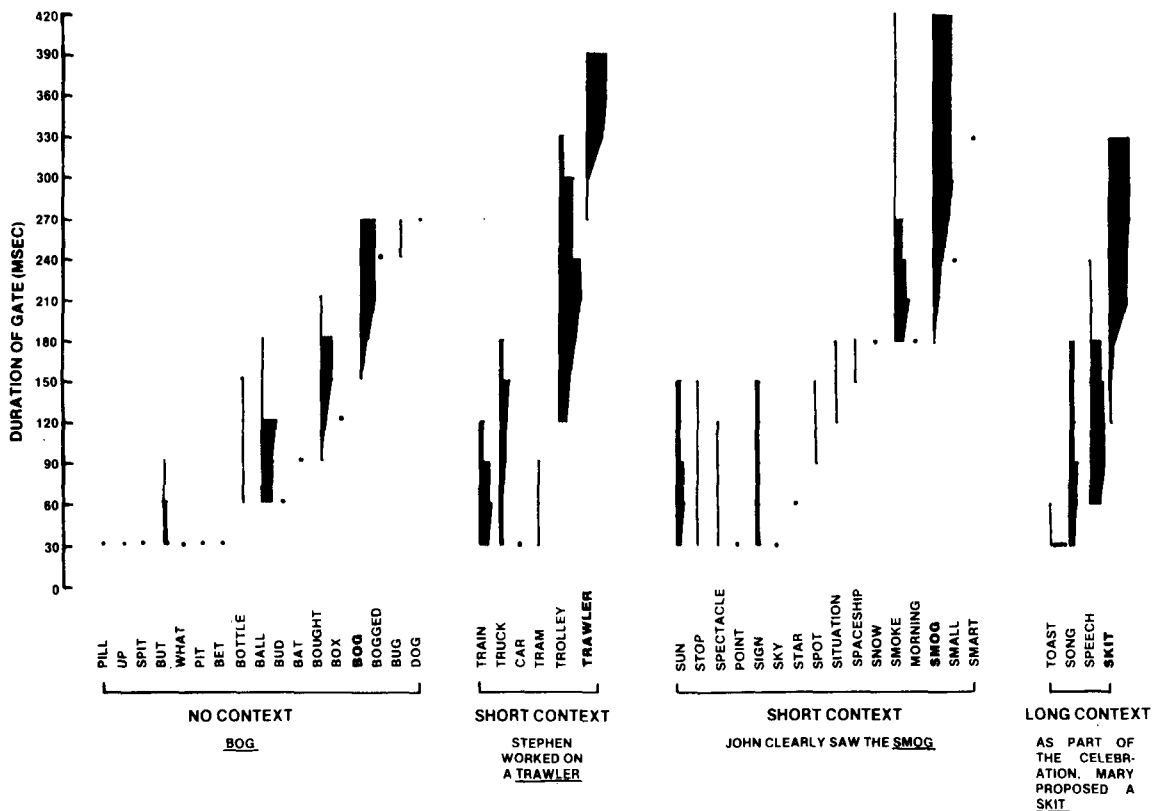

Figure 7. The frequency garden path: Candidates proposed for the stimulus words *bog* (no context), *trawler* (short context), *picture* (long context), and *culture* (long context) (see Figure 4).

that of *smog* (f = 49, as opposed to 1). Again, only additional phonetic information will force subjects to retrace their steps and propose the low-frequency *smog*. We should note that in the SC condition for *smoke* (*Ann caught sight of the*) only one subject ever proposed *smog*, and this for only two presentations.

A final example of the word-frequency garden path concerns the stimulus word *skit* in the LC condition: *As part of the celebration, Mary proposed a.* Again, quite naturally, all guesses are quite compatible with the context (*toast, song, speech*). What is interesting, however, is the series of garden paths that are based on the frequency bias and the acoustic information available. With the very first gate (+30 msec), five subjects propose *toast* (f = 19), but this is short-lived when they realize that the first consonant is a fricative. So *song* (f = 70) becomes a relatively strong candidate until more acoustic information is available. When subjects realize they are dealing with a CC word (the first consonant is a fricative and the second a stop), they all transfer to *speech* (f = 61), which is then proposed for a number of presentations by seven of the eight subjects. Finally, information concerning the final /t/ forces the subjects to backtrack; they now realize that the second consonant is not a /p/, as they all thought at first, but, in fact, a /k/, and they finally propose the correct word, *skit*.

To confirm the high-frequency bias illustrated in Figure 7, we analyzed the candidate types proposed for 18 stimulus words (nine high-frequency and nine low-frequency words taken from each of the three syllable groups in each of the three context conditions). Using the Kučera and Francis (1967) word-frequency list, we attributed a frequency to each candidate type and averaged frequencies (by taking the geometric mean) within words. No significant difference was found between the frequency of the candidate types proposed for high- and low-frequency stimulus words in either of the three context conditions (t = −.08, .59, −.31 for the NC, SC, and LC conditions, respectively). This means that subjects stay with high-frequency word candidates until they are forced down to the lower frequency words. Our next question was whether this shift to the lower frequency candidates was gradual, that is, whether as more acoustic information appears, subjects start to realize that they are dealing with a lower frequency word and reflect this in their proposals or whether the shift is abrupt. We therefore examined the frequency of the proposals just before a word was isolated (i.e., the last erroneous guess for each subject) and again found no significant difference between high- and low-frequency words (for all context conditions, t = −1.22, n.s.). This means that the shift to a low-frequency candidate is an abrupt one (it takes

place at the moment the low-frequency word is isolated) and that, whether the stimulus is of high or low frequency, subjects propose only high-frequency candidates. It is only when context and/or acoustic information forces them down to a low-frequency candidate that they backtrack on the frequency garden path and finally isolate the correct low-frequency word.

The semantic garden path. Overall, our two context conditions increased isolation times and improved confidence ratings. In this sense, our results confirmed that context usually has a facilitatory effect on lexical access. However, the technique we used to assess the strength of the contexts (SC and LC) preceding the stimulus words (Marslen-Wilson & Welsh, 1978) was not always foolproof, and this led to a few garden-path situations. Subjects, basing their decisions on the context and the existing phonetic information, went astray when the top-down and bottom-up information did not converge as well as they could. These garden paths are extremely revealing, inasmuch as, in actual speech situations, a conflict between top-down and bottom-up information may often exist. In Figure 8, we give the few examples found in which such a conflict seems to have existed. It should be noticed immediately that the frequency bias is involved to some extent in each of these examples, but that the conflict between semantic/syntactic information and the phonetic information is also evident.

The first example concerns the word *bog* in the short context *He walked into a*. It is striking to note that all eight subjects propose *bar* repeatedly before they realize that they are in error and disperse at Gate 8 (+240 msec). For 180 msec, they think *bar* is the logical place to walk into despite the conflicting bottom-up information, which should lead them to *bog*. What is interesting is that, in the NC condition, *bar* is never proposed and, in the LC condition (*Lost in the Scottish Highlands, she walked into a*), it is proposed by only one subject during the first five presentations.

The next example concerns a very early bias for *police* as a candidate for the stimulus word *partner* in the SC condition *John first talked with the* (see Figure 8). Based on only the first 30 msec of the stimulus word (which lasted 390 msec), seven of the eight subjects chose the candidate *police*. This can be explained by a semantic garden path linked to the context from which one may infer that something unusual has happened (e.g., an accident) and John has to tell a number of people about it. In this case, it seems logical that the police should be contacted first, and the first 30 msec seems to confirm this—a /p/ is clearly audible. Again, like many other garden paths, a general dispersion takes place when the subjects realize they have been led down a garden path [it is only at the 9th gate (+270 msec) that as
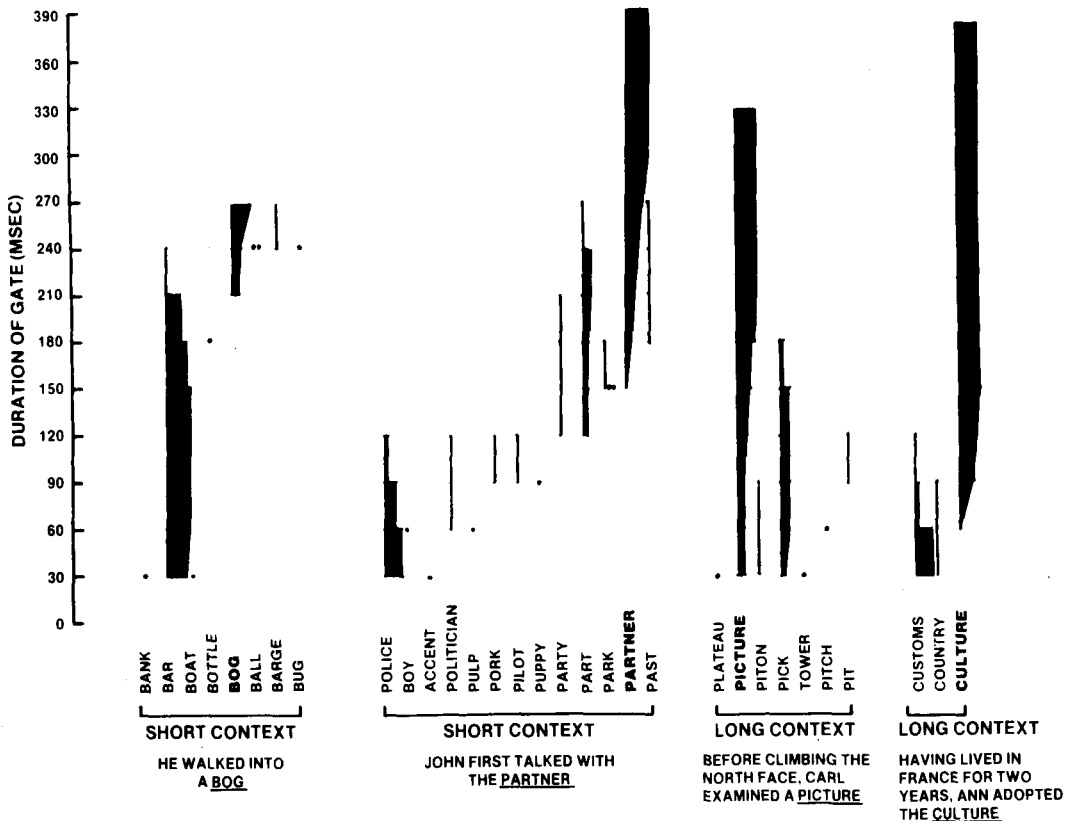
Figure 8. The semantic garden path: Candidates proposed for the stimulus words *bog* (short context), *partner* (short context), *picture* (long context), and *culture* (long context) (see Figure 4).

many subjects center in one proposal again (for *partner*)].

The third example (*picture*) is interesting in that, when it is presented in the context *Before climbing the North Face, Carl examined the*, a number of subjects choose the word *pick*, thus delaying the final isolation of *picture* by all subjects until the seventh gate (all eight subjects isolated *picture* by the third gate in the SC condition: *Carl examined the*). Subjects in the LC condition were clearly attempting to pick a candidate that would fit the context (what it is you examine before climbing a mountain) and that would also be in accord with the acoustic-phonetic information. As the first syllable (*pick*) fits both constraints, they propose it as a candidate, even though, after a while, more acoustic information (the /č/) contradicts their proposal. Once again, only additional bottom-up information forces the garden-path subjects to backtrack and change their candidate. (It is interesting to note that, for mountain climbers, *picture* would almost certainly have been an early candidate; one never undertakes a long and dangerous north-face climb without having carefully studied the itinerary on a diagram or a picture of the face.)

The final example in Figure 8 concerns the proposal for *customs* instead of *culture* in the context *Having lived in France for two years, Ann adopted the*. The

word *customs* is proposed by seven of the eight subjects for the first two gates, not because it is more frequent than *culture* (in fact, it is less frequent: f = 18 as opposed to f = 58), but because, on the one hand, it does not contradict the acoustic information given in the first two gates (+30 and +60 msec) and, on the other hand, it is probably a better continuation of the verb *adopt*. However, very quickly information concerning the third phoneme, /l/, comes in and subjects abruptly switch over to *culture*, which also fits the bottom-up and top-down information quite well.

These examples clearly call for a study that would investigate systematically the conflict that may arise when top-down and bottom-up information do not reinforce one another. Such a study would pit semantic constraint against acoustic-phonetic information and would allow one to determine how much bottom-up information is needed to override the top-down constraint. Instances could even be devised in which no amount of phonetic information would suffice to force subjects out of the semantic garden path.

GENERAL DISCUSSION

In this section, we will examine our results in the light of an existing word-recognition model and then

discuss the possible uses of the gating paradigm in the study of on-line processing of speech.

The Active Direct Access Model

Instead of examining a series of word-recognition models (Cole & Jakimik, 1979; Forster, 1976; Morton, 1969; Rubenstein, Lewis, & Rubenstein, 1971) and determining how our results match their predictions, we will concentrate on one model that was at the origin of this study—Marslen-Wilson's active direct access model (also called the cohort model; Marslen-Wilson & Welsh, 1978; Marslen-Wilson, Note 2)—and study our data in light of its predictions. The central concept of this model is that a word initial cohort, which is a directly accessed set of potential word candidates, is activated during the earliest phases of the word-recognition process. This set is accessed solely on the basis of bottom-up information and consists of the entire set of words in the language that begin with a particular initial sequence. It is this word initial cohort, which is based on bottom-up information alone, that is assumed to be the basis for word recognition in all contexts. A word is recognized at the point, going from left to right through the word, that the word in question becomes uniquely distinguishable from any other words in the initial cohort. The role of top-down information is to speed up the process of partitioning the initial cohort of word candidates down to a single choice. Just as memory elements will withdraw from the pool of word candidates when they no longer fit the acoustic-phonetic input, so they will also withdraw when they no longer fit specifications of context. This allows for a highly flexible balance between top-down and bottom-up information sources. In fact, the model can make precise predictions about the time course of recognition for individual words when they occur in isolation or a specified context. Once a single word choice has emerged, a less detailed assessment of the remaining input for that word will be required.

As we have already seen in the section on isolation times, the cohort model accounts well for many of our results. For example, words in normal context were identified with very little of the acoustic signal corresponding to that word, and, in fact, isolation took place before sufficient bottom-up information could have accumulated to allow a single word candidate to appear on the basis of bottom-up information alone. We also found that long words took longer to isolate than short words, as predicted by the model, and that candidates remove themselves (in our task, fewer candidates are proposed) as more input is heard or more context given.

However, a few additional characteristics may need to be built into the model in order to predict all of our results. The first, which has been mentioned by a number of researchers (e.g., Spoehr, 1980), is

that the cohort model does not account for the frequency effect that has been found in this and other studies. One way of building this effect into the model would be to propose that candidates are organized or weighted according to frequency of occurrence and that subjects have a bias toward high-frequency words. This would be somewhat akin to the assumption proposed by Foster (1976) in his autonomous search model, which states that the listener has to check through the high-frequency words in an access file bin before checking through the low-frequency candidates in his recognition procedure.

A second aspect on which the cohort model and our data do not match perfectly concerns the nature of the initial cohort of candidates. Marslen-Wilson and Welsh (1978) write, "Word recognition in continuous speech is fundamentally data-driven, in the specific sense that the original selection of word-candidates is based on the acoustic-phonetic properties of the initial segment of the incoming word. *We do not need to assume that top-down processes themselves preselect a class of word candidates*" (p. 61; italics added). Although we agree with the hypothesis that some acoustic-phonetic (bottom-up) information is practically always necessary for word isolation and recognition (how often do we hear constraining sentences of the type *She picked up the gun, aimed and fired*, in which the probability of hearing the last word "fired" is almost 1.0?), and although we are in agreement that only the inherent phonetic-acoustic characteristics of the word are the important factors for recognition of words out of context, we do believe that top-down information can help restrict the initial cohort of candidates in number and in kind. As we have seen, the SC and LC conditions have the effect of reducing the number of candidate types at the first presentation (+30 msec) and of greatly constraining the semantic set of these candidates (see Figures 4 and 5). Although it is true that candidates obtained across subjects may not reflect exactly those within subjects at the onset of a word, we nevertheless propose that top-down and bottom-up information will interact to select a class of word candidates and that bottom-up information will then help the listener narrow-in on the appropriate candidate.

A third point concerns the life span of candidates: When do they become candidates, how long do they remain as candidates, and when do they drop out? Although the cohort model is not explicit on this point, it does seem to propose a very large initial cohort which consists of the entire set of words in the language that begin with a particular initial sequence and a steady narrowing-in process (using top-down and bottom-up information) to finally arrive at one final candidate. We propose, based on our

data, that the narrowing-in process is probably much more complex and may not be a simple elimination of candidates from an entire set at the beginning of the word. For example, we find subjects narrowing in on only one candidate that is not the stimulus word (see the "word from a word," the frequency, and the semantic garden paths we have described) and therefore having to switch to other candidates before finally isolating the word. Only further research on the narrowing-in process with other paradigms will be able to describe the complex pattern that is followed in eliminating candidates and finally isolating and recognizing the word.

A final point about the cohort model (and any other word-recognition model) concerns the point at which a word is said to be recognized. The cohort model proposes that a word is recognized at the point at which the word in question becomes uniquely distinguishable from any other word in the word initial cohort (this point can be reached with bottom-up information only or with bottom-up and top-up information). Other models talk of a word-recognition threshold (the logogen model) or a criterion value for a match (the autonomous search model). The results we obtained from the confidence ratings—low confidence ratings at the recognition point and, at times, far from perfect confidence ratings at the end of the word—suggest a much stronger monitoring component to word recognition than that proposed in most models. For example, Marslen-Wilson proposes, for the cohort model, that as soon as a single word choice has emerged, the recognition system will have achieved its primary goal and a less detailed input for that word will be required. However, the emergence of a single candidate does not automatically mean that the stimulus word has been recognized: Various garden paths can lead to single, but erroneous, candidates, and some single candidates, although correct, are extremely tentative in that the top-down information is not constraining enough or that the acoustic-phonetic information is not redundant enough (e.g., one-syllable words). Thus, although some candidates emerge as single choices quite early (at the isolation point), we propose that the listener keeps monitoring the input so that a final decision on the word is put off until quite late, even on occasion some distance into the next word or two (this decision would be taken at the recognition point). Such monitoring would allow subjects to increase their confidence in a candidate and to switch to another candidate when, for example, they have been led down the garden path.

The Gating Paradigm

In this study, we have attempted to show that the gating paradigm is a useful and valid way of studying one aspect of on-line processing, namely lexical ac-

cess. We first validated the paradigm by replicating three effects that have been obtained with other paradigms (the word-frequency, the context, and the word-length effects) and by extending the data offered by Marslen-Wilson concerning the time it takes to identify words of varying lengths in and out of context. Then, by examining the confidence ratings of subjects at the isolation point of the word and at the end of the word, we were able to show that spoken word recognition may involve two distinct stages: isolating an appropriate candidate and, after monitoring the following acoustic-phonetic information, finally accepting or "recognizing" that candidate. Next, by analyzing the erroneous guesses made up to the isolation point, we were able to study more closely the narrowing-in process employed by listeners when isolating words. Although our results need to be confirmed with other paradigms, we have shown that the isolation process is not a simple narrowing-in; it is highly complex, and, as with syntactic processing, word isolation and recognition is strewn with garden paths. It is in this respect— uncovering the word isolation process itself—that the gating paradigm appears to be very useful.

We are now employing the paradigm in another word-recognition experiment in which the stimulus word is in conflict with the semantic context (this will help us to determine how much acoustic-phonetic information is needed to counteract the conflicting contextual information) and in a sign-language experiment in which we ask which factors in a sign (frequency, length, inflection, etc.) are important in the lexical access of this visually presented language. In addition, we are using this paradigm at the phrase and sentence levels to help us uncover the processes involved in on-line processing of sentences.

We believe that the gating paradigm will prove especially useful in determining the minimum amount of information (acoustic-phonetic, syntactic, semantic, and pragmatic) needed for the analysis of language during on-line processing. And, along with other "on-line" paradigms, we hope that this task will enable us to map out the cognitive operations that take place during real-time processing of language.

REFERENCE NOTES

1. Schuberth, R., & Spoehr, K. *Interaction of stimulus and contextual information in the lexical decision process.* Unpublished manuscript, Department of Psychology, Brown University, 1979.

2. Marslen-Wilson, W. *Sequential processes during spoken word recognition.* Paper presented at the annual meeting of the Psychonomic Society, San Antonio, Texas, 1978.

3. Huggins, A. *A facility for studying perception of timing in natural speech* (Quarterly Progress Report, Research Laboratory of Electronics, **95**, 81-83). Cambridge, Mass: M.I.T., 1969.

REFERENCES

BECKER, C. Semantic context and word frequency effects in visual

word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, **5**, 252-259.

CLARK, H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335-359.

COLE, R. Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics*, 1973, **11**, 153-156.

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

CUTLER, A., & NORRIS, D. Monitoring sentence comprehension. In W. Cooper & E. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, N.J: Erlbaum, 1979.

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

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

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

JAKIMIK, J. Word recognition and the lexicon. In J. Wolf & D. Klatt (Eds.), *Speech communication papers presented at the 97th meeting of the Acoustical Society of America*. New York: Acoustical Society of America, 1979.

KIRK, R. *Experimental design: Procedures for the behavioral sciences*. Belmont, Calif: Brooks/Cole, 1967.

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

MARSLEN-WILSON, W. Linguistic structure and speech shadowing at very short latencies. *Nature*, 1973, **244**, 522-523.

MARSLEN-WILSON, W. Sentence perception as an interactive parallel process. *Science*, 1975, **189**, 226-228.

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

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

MILLER, G., HEISE, G., & LICHTEN, W. The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, 1951, **41**, 329-335.

MILLER, G., & ISARD, S. Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 217-228.

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

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

ROSENZWEIG, M., & POSTMAN, L. Intelligibility as a function of frequency of usage. *Journal of Experimental Psychology*, 1957, **54**, 412-422.

RUBENSTEIN, H., LEWIS, S., & RUBENSTEIN, M. Homographic entries in the internal lexicon: Effects of systematicity and relative frequency of meanings. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 57-62.

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

SAVIN, H., & BEVER, T. The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 295-302.

SCHUBERTH, R., & EIMAS, P. Effects of context on the classification of words and non words. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, **3**, 27-36.

SORENSEN, J., COOPER, W., & PACCIA, J. Speech timing of grammatical categories. *Cognition*, 1978, **6**, 135-153.

SPOEHR, K. Word recognition in speech and reading: Toward a single theory of language processing. In P. Eimas & J. Miller (Eds.), *Perspectives on the study of speech*. Hillsdale, N.J: Erlbaum, 1980.