Journal of Experimental Psychology:
Learning, Memory, and Cognition
1990, Vol. 16, No. 3, 446-456

0278-7393/90/$00.75

Suffix Interference in the Recall of Linguistically Coherent Speech

David A. Balota
Washington University

Nelson Cowan
University of Missouri at Columbia

Randall W. Engle
University of South Carolina

Four experiments are presented that address the stimulus suffix effect for linguistically coherent spoken materials. In Experiment 1, definitions of low-frequency words were presented for on-line written recall. Each definition was followed by a nonword speech suffix presented in the same voice as the definition, the same nonword presented in a different voice, or a tone. The results yielded a significant reduction in the recall of the terminal words of the definitions in the speech suffix conditions compared with the tone control. This general pattern was replicated in Experiment 2, in which subjects did not begin their recall until the suffix item or tone was presented, although the magnitude of the suffix effect was reduced in this experiment. In Experiment 3, sentences that were part of a cohesive story were presented for on-line recall. Here, the suffix effect was considerably reduced compared with the suffix effect found with the definitions presented in Experiments 1 and 2. This pattern was replicated in Experiment 4, in which subjects did not begin their recall of the story sentences until the speech suffix or tone was presented. Overall, the results suggest that auditory memory interference can take place for linguistically coherent speech, although the magnitude of the interference decreases as one increases the level of linguistic structure in the to-be-recalled materials. Implications of the present results for current models of natural language processing are discussed.

Within the past two decades there has been considerable research, with a variety of experimental procedures, examining different types of retroactive interference effects on the recognition and recall of spoken words and sounds (e.g., Crowder & Morton, 1969; Massaro, 1970, 1972; for reviews, see Cowan, 1984; Crowder, 1976, 1978; Penney, 1989). The primary goal of this research has been to understand better the theoretical mechanisms underlying auditory memory. A secondary, and quite important, goal in this research is to understand better the role that such an auditory memory system might play in the comprehension of spoken language. Here, the progress has been limited primarily to speculations that the same types of memory for auditory stimuli addressed in experimental studies are also important in the processing of ordinary language and that some of the same interference mechanisms also operate in everyday speech comprehension (e.g., Crowder, 1976; Darwin & Baddeley, 1974; Massaro, 1972, 1975; McNeill & Repp, 1973; Pisoni, 1973).

The primary goal of the present study is to begin filling the void in the current experimental literature by addressing auditory memory for materials, such as phrases and sentences, that are linguistically more natural than those materials typically used in studies of auditory memory. In pursuit of this goal, we attempted to use an experimental paradigm that (a) has been widely used to address the auditory memory system and, therefore, provides a large literature base, (b) appears to be sensitive to mechanisms that might play a role in speech perception and comprehension, and (c) can be easily extended to investigate auditory memory with phrases and sentences.

One experimental procedure that is particularly intriguing for studying auditory memory for natural language materials is the stimulus suffix procedure (e.g., Crowder & Morton, 1969). In a typical suffix experiment, a list of to-be-recalled words or syllables is followed by an additional redundant sound called the *suffix*. For instance, a list of eight digits might be presented with the word *go* appended to the end of the list. The subject's task is to serially recall the list of eight digits. Results from such studies indicate that if the suffix item is acoustically similar to the list items, there is considerable interference with memory for the last few list items (e.g., Morton, Crowder, & Prussin, 1971).

Most accounts of the stimulus suffix effect suggest that the suffix interferes with a *surface-level* representation of the last few list items (e.g., Balota & Duchek, 1986; Crowder, 1976; Greene & Crowder, 1984). Here, by *surface level* we mean a level of storage that occurs before there is meaning-level integration across the words in the to-be-remembered list. In the most celebrated model, the suffix effect presumably reflects interference with a prelexical representation, referred to as *echoic memory*, that primarily holds auditory sensory information (see Crowder, 1976; Crowder & Morton, 1969).

This research was supported by Grant 54126A to David A. Balota from the National Institute of Aging and by Grant HD21338 to Nelson Cowan from the National Institutes of Health. David A. Balota would also like to express his gratitude to the Netherlands Institute for Advanced Studies for their support while portions of the present article were written. Finally, we would like to thank Robert Crowder, Robert Greene, Catherine Penney, and Keith Rayner for their helpful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to David A. Balota, Department of Psychology, Washington University, St. Louis, Missouri 63130.

The more recent model developed by Greene and Crowder (1984) still emphasizes a prelexical locus for suffix interference. The fact that most studies of the stimulus suffix effect have been conducted with stimuli that provide very little opportunity for higher level integration (i.e., lists of digits or unrelated words) clearly promotes a surface-level account.

If the suffix primarily interferes with surface-level representations, then one might question whether suffix-type effects would be obtained with more natural language materials. Most models of language processing assume that surface-level representations are held only briefly and that the predominant form of representation changes as the words are integrated into a higher level linguistic representation that conveys primarily the meaning of the utterance. Consider, for example, the classic work of Sachs (1967). She found that memory for the precise wording of auditorily presented sentences (surface-level information) dropped considerably between a 0-syllable and 40-syllable delay but that memory for meaning remained relatively high at the 40-syllable delay (also see Fillenbaum, 1966, and Hanson & Bellugi, 1982, for similar findings). Moreover, the work of Marslen-Wilson (1975) provides compelling evidence that words are integrated into a meaning-based representation rather immediately. In these studies, Marslen-Wilson addressed the impact of semantic and syntactic information in the shadowed message for *close shadowers*. Close shadowers are individuals that can shadow a message at a very short lag (approximately 250 ms) between the onset of the to-be-shadowed stimulus word and the onset of the shadower's pronunciation. Even though these individuals could repeat the message virtually in synchrony with the signal, they still corrected any errors in the shadowed message to conform to meaning-level information. These results suggest that in speech comprehension words are very quickly integrated into a higher level meaning representation (also see Cole & Jakimik, 1980; Marslen-Wilson, 1989; Marslen-Wilson, Tyler, & Seidenberg, 1978; Potter & Faulconer, 1979).

In addition, to the results in the auditory modality, similar findings have been found in the visual modality. The work by Frazier and Rayner (1982) using eye-tracking methodology provides rather dramatic support for the notion that words are very quickly integrated with earlier words within sentences. They found that longer fixation durations were associated with the very first fixations in the area of a sentence that syntactically disambiguated the meaning of the sentence. Moreover, subjects often made regressive eye-movements to earlier portions of the sentences to attempt to recapture the surface-level representations for the appropriate reading. Thus, it appears that the human sentence parsing mechanism immediately computes at least some of the structural consequences of fixated material for the analysis of preceding material, and according to the regressive eye-movement data, does not always have a surface-level representation available in memory to recompute the new reading of the sentence. Just and Carpenter (1980) even more strongly emphasize the impact of on-line integration of words within a higher order text structure. According to their *immediacy assumption*, all processing associated with a word is completed before the eyes leave that word.

The important point to note here is that according to such immediate processing models, words are very quickly integrated within a higher level linguistic representation. In addition, as noted, there is some evidence that surface-level information about a given combination of words is lost relatively quickly. If indeed the suffix item interferes with surface-level information and if higher level linguistic integration quickly renders that surface-level information relatively superfluous, then one might expect little, if any, suffix effect with linguistically coherent materials.

Although there are clear reasons why one might expect relatively little suffix effect for linguistically coherent materials, there are also reasons why one might still predict a suffix effect with such materials. For example, the work of both Caplan (1972) and Jarvella (1971) suggests that although higher level integration may occur very rapidly, the surface-level representation is not fully lost until the subject reaches a clause boundary (see, however, Von Eckardt & Potter, 1985). Thus, it is possible that the suffix item might interfere with the surface-level representation in the current clause. Furthermore, despite the coherence of natural speech, it is possible that at some points it might overload the subject's working memory capacity (see Baddeley, 1986; Miller, 1956). The integration of individual words into a higher level text representation may lag behind presentation of the information, and hence, the listener may be forced to rely on a surface-level representation. Under these conditions, one would clearly predict a suffix effect with linguistic materials.

In the present study, four experiments are presented that address suffix effects with natural language materials. The basic paradigm was similar to that of taking notes from a lecturer: Subjects simply listened to each sentence and attempted to write it down verbatim. The materials varied between experiments in the amount of syntactic and semantic redundancy available for higher level encoding and retention (lower redundancy: dictionary definitions for low-frequency words, Experiments 1 and 2; higher redundancy: consecutive sentences from a child's story, Experiments 3 and 4). The experiments also varied in the method of the subject's response. Subjects either began writing as soon as the sentence began (Experiments 1 and 3) or waited until the sentence ended (Experiments 2 and 4). An immediate response may tend to shift the memory load toward the end of the sentence when the subject's writing cannot keep up with the stimulus presentation rate. In contrast, a delayed response may allow the subject to utilize the higher level information in the utterance before responding.

Following the procedure of a number of previous studies (e.g., Balota & Duchek, 1986; Greenberg & Engle, 1983; Morton et al., 1971), each stimulus sentence was followed by one of three suffixes: (a) a nonword (*snoke*) spoken in the same voice as the sentence, (b) the same nonword spoken in the voice of a person of the opposite sex, or (c) a tone. To the extent that some sort of speech-specific representation operates, recall should be superior with a tone suffix compared with either the same-voice suffix or the different-voice suffix. This should be the case if the relevant speech memory is echoic storage as it is generally conceived (e.g., Crowder, 1976), a phonetic store (e.g., Baddeley, 1986; Cheng, 1974),

or even an abstract speech memory representation that is derived from auditory and visual speech information together (Greene & Crowder, 1984). Furthermore, to the extent that a memory based on acoustic characteristics is involved (e.g., echoic storage as it is generally conceived; see Crowder, 1976), performance should also be higher in the different-voice compared with the same-voice suffix condition because the amount of acoustically based interference should be greater with the same-voice suffix. Of course, the ability to detect a difference between same- and different-voice suffixes is limited by the size of the overall suffix effect. As noted, there are clear reasons from current language processing models for expecting a relatively small suffix effect with more natural language materials, thereby limiting the potential to detect a voice-specific impact of the suffix item.

Experiment 1

In the first experiment the suffix effect was examined, with the to-be-remembered items being dictionary definitions of low-frequency words. Subjects were asked to transcribe tape-recorded definitions verbatim. They were instructed to begin writing down each definition as soon as the stimulus definition began. The definitions were recorded in natural rhythm and intonation and were based on actual definitions obtained from a college-level dictionary. At the end of each definition, one of the three suffix items was presented.

Method

Subjects. A total of 24 undergraduates from the University of Massachusetts at Amherst participated in this experiment as partial fulfillment of a psychology course requirement.

Apparatus. A Sony reel-to-reel recorder was used to record and present the stimuli. The tone-control suffix was produced by a Zenith Horizon microcomputer.

Materials. A total of 66 definitions of low-frequency words (i.e., words with less than 12 occurrences per million according to the Kučera & Francis, 1967, norms) were obtained from a college-level dictionary. Half of the 66 definitions were 15 words in length, and the other half were 20 words in length. (Some of the definitions had to be slightly modified to equal these word lengths.) Six of these definitions (three of each length) served as practice items, with the remaining 60 definitions serving as target items. The following two definitions are examples of 15-word and 20-word definitions for the words *chaperone* and *rudder*, respectively: (a) *a person, especially an older or married woman, who accompanies young, unmarried people in public;* (b) *a flat, movable piece of wood or metal hinged vertically at the stern of a boat and used for steering.*

The recorded definitions were produced by a male speaker in a normal intonation and speaking rate. After each definition was read, there was a delay of approximately 0.5 s, and then one of the three following events occurred: (a) The male speaker pronounced the nonword *snoke*; (b) a female speaker pronounced the nonword *snoke*; (c) a 500-ms, 500-hz tone was presented. Care was taken to make all of the stimuli as similar as possible in perceived loudness and timing of the suffixes. There was a 33-s silent period after each 15-word definition, and a 42-s silent period after each 20-word definition to allow subjects to complete their transcriptions.

The 60 test definitions were presented in random order, with the constraint that there was an equal representation of the two definition lengths and three suffix items in both the first and the second block of 30 test trials. The same random order of definitions was used for each of the three test tapes. Each test tape included 10 definitions for each combination of suffix type and definition length, but the suffixes were matched to particular definitions differently across the three tapes. For instance, if a particular definition was followed by a male suffix on Tape A, then that same definition was followed by a female suffix on Tape B and a tone suffix on Tape C. A different group of subjects received each tape. Thus, across subjects, each sentence was used an equal number of times for each of the three suffix conditions.

Procedure. Subjects were told that on each trial they would hear a definition of a relatively rare word. Their task was simply to begin writing down the definition as soon as the first word was presented. Subjects used a response booklet for their written recall, with each trial number indicated in the margin. The three suffixes that occurred at the ends of the definitions were also described, and subjects were instructed to ignore these stimuli.

Subjects were told that the sentences varied in length and that on some trials they might not be able to recall all of the words. When they could not recall some of the words in the definition, they were instructed to leave blank space on the line to indicate the portion of the sentence they could not recall.

The experiment was conducted in a sound-attenuated room. The stimuli were played through two speakers facing the subjects. A total of 2 to 4 subjects participated simultaneously in each session, with each subject working alone. Subjects were monitored via an intercom connected to a control station and were instructed to remain silent during the testing session. They were given a short break to ask questions after the last (6th) practice trial and were also given a short rest period after the 30th test trial.

Results

The primary method of scoring each subject's data in Experiment 1 and all subsequent experiments had several important characteristics. First, the scorers were blind to condition; that is, they did not know which suffix was used on a given trial. Second, the data were scored according to the serial position of each word in the stimulus definition. The subject received a *0* for each incorrect or omitted word and a *1* for a correctly recalled word. This scoring procedure resulted in an ordered series of 15 to 20 digits (*0* or *1*s) for each trial. Third, extra words inserted by the subjects did not affect their scores, provided that the surrounding words in the response could be clearly identified in the stimulus definition. For example, if the phrase *a large wooden container for bread* was transcribed as *a large wooden container of very high quality for the storage of bread*, no credit would be lost for the words *a large wooden container* or for the word *bread*. Fourth, subjects were given credit for a content word in a definition (i.e., a noun, verb, adjective, or adverb) whenever the root word was reported correctly. For example, if the stimulus word was *skirt*, the subject received credit for *skirts* but not for *dress*. Fifth, subjects received credit for function words (defined here as articles, prepositions, and conjunctions) only if the function word was written correctly and placed before or after an appropriate content word from the definition. For example, in a definition that began *The part of a jacket which is made . . .* a subject would receive credit for the word *of* if the subject wrote *Parts of jackets* or *A part of a new coat . . .*, but not if the subject wrote *Jackets of which part . . .* because in the last rendition, *of* neither follows

the correct content word (*part*) nor precedes the correct content word (*jacket*).

Two of the present authors (DB and NC) each scored half of the data. The two scorers used pilot data to ensure that they were using the same coding method. Analyses of the data with scorer as a factor indicated that this factor did not produce any significant main effects nor interactions with any of the variables of interest. Finally, it should be emphasized that any error in scoring should equally affect all conditions because, as noted, each definition occurred in each suffix condition, and the scorers were blind to condition.

The mean percent correct as a function of serial position and suffix condition for the 15-word definitions is displayed in Figure 1. There are two major points to note in Figure 1. First, recall accuracy produced the well-documented U-shaped serial positions curve. Second, there was a difference between suffix conditions in the expected direction primarily at the last four serial positions. For these serial positions, the tone-control condition produced the highest accuracy, followed by the different-voice condition, which, in turn, produced higher accuracy than the same-voice condition.

These observations were supported by a 3 (suffix type) × 15 (serial position) within-subjects analysis of variance (ANOVA). This analysis yielded main effects of suffix type, $F(2, 46) = 6.10$, $MS_e = 0.0331$, and serial position $F(14,322) = 49.90$, $MS_e = 0.0148$, along with a significant Suffix Type × Serial Position interaction, $F(28,644) = 2.14$, $MS_e = 0.0109$. (Unless otherwise specified, all significant effects are at least at the .05 level.) Post hoc comparisons at each serial position yielded significant main effects of suffix condition only at Serial Positions 12, 13, 14, and 15. In pairwise comparisons of the suffix conditions at these four final serial positions, the tone-control and same-voice conditions differed at all four positions. The tone-control and different-voice conditions significantly differed at Serial Positions 13 and 14, whereas the same- and different-voice conditions significantly differed at Serial Positions 12 and 15.

Figure 2 displays the mean percent correct for the 20-word definitions as a function of suffix type and serial position.

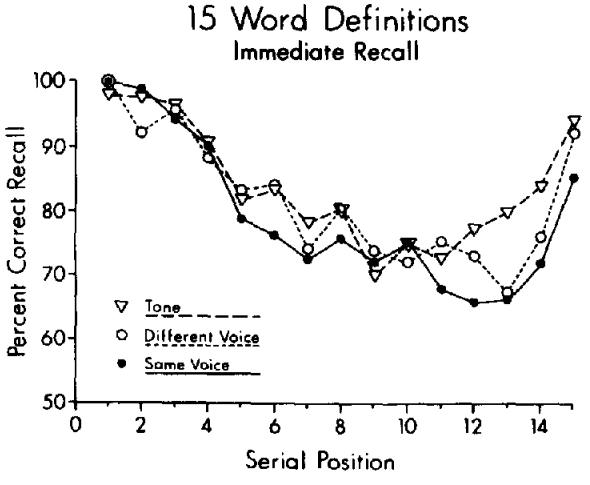

Figure 1. Mean percent correct immediate recall as a function of suffix type and serial position for the 15-word definitions.

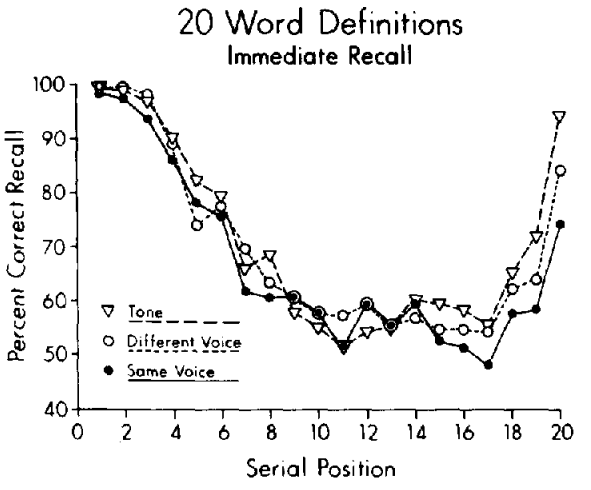

Figure 2. Mean percent correct immediate recall as a function of suffix type and serial position for the 20-word definitions.

The recall accuracy again conformed to the expected U-shaped serial position function. More important, suffix condition influenced performance primarily at Serial Positions 16–20.

An ANOVA on the 20-word definitions yielded main effects of suffix type, $F(2, 46) = 6.71$, $MS_e = 0.0321$, and serial position, $F(19, 437) = 78.70$, $MS_e = 0.0237$, along with a significant interaction between suffix type and serial position, $F(38, 874) = 1.56$, $MS_e = 0.0165$. Post hoc comparisons at each serial position indicated that there were significant main effects of suffix type at Serial Positions 16–20. As shown in Figure 2, the main effect of suffix type across Serial Positions 16–20 is primarily due to the differences between the same-voice and tone-control conditions. However, it is noteworthy that the different-voice condition is significantly different from *both* the same-voice and the tone-control conditions at Serial Positions 19 and 20, with the different-voice condition being intermediate between the same-voice and the tone-control conditions.

Discussion

There are a number of noteworthy aspects of the data from Experiment 1. First, it appears that one can produce a suffix effect with linguistically structured information that is produced with a natural rhythm and intonation pattern. Thus, it appears that at least in writing down definitions of low-frequency words, redundant speech information (i.e., the suffix) can selectively interfere with the recall of the last few list items. Moreover, there does appear to be some sensitivity to the physical characteristics of the suffix. There was more interference for the same-voice suffix than for the different-voice suffix, which in turn produced more interference than the tone-control condition. As noted, this sensitivity to the voice of a speech suffix has been viewed as support for the acoustic specificity of the storage mechanism with which the suffix interferes (e.g., Crowder, 1976).

One might ask whether subjects were actually using the linguistic constraints available in the definitions. It is possible

that the suffix effects were obtained only because, in this test situation, subjects treated the definitions as lists of unrelated words. If this were the case, then the present results would provide no new information concerning suffix interference for linguistically structured materials and would have no relevance to the models of language processing described earlier. However, in scoring the data, it was clear that subjects were attempting to recall the data in linguistically structured units, that is, as sentences or phrases. Subjects rarely produced syntactically unrelated sequences of words. In fact, only 8 protocols of the total 1,440 produced strings could be classified as unrelated sequences of words. In addition, the notion that subjects were using linguistic structure to supplement their recall of the definitions is further supported by the relatively high recall performance. Recall accuracy across serial positions was 82% correct for the 15-word definitions and 69% correct for the 20-word definitions.[1]

A second intriguing aspect of the data from Experiment 1 was that there appeared to be a larger terminal suffix effect for the 20-word definitions compared with the 15-word definitions. In fact, an ANOVA on only the last serial position for the 15- and 20-word definitions yielded a significant Suffix × Definition Length interaction, $F(2, 46) = 5.08$, $MS_e = 82.52$, indicating that the suffix effect was larger for the 20-word definitions than for the 15-word definitions. Furthermore, as noted in the *Results* section, for the 15-word definitions the difference between the tone-control and the different-voice conditions did not reach significance at the last serial position, whereas for the 20-word definitions this difference was highly significant. Thus, it does appear that there was a larger impact of the suffix for the 20-word definitions compared with the 15-word definitions.

Obviously, increasing definition length would place a greater load on the subject's memory system. Possibly, when there is a greater load, there may be an increased reliance on a surface-level trace to supplement other storage mechanisms. This could occur because memory for speech sounds (in the form of echoic, phonetic, or articulatory memory) is more helpful when the processing system becomes overloaded and lags behind the linguistic analyses of the incoming speech. If so, then when the 20-word definitions were presented, subjects would have been more likely to rely on a surface-level representation than when the 15-word definitions were presented. This would account for the greater potency of the speech suffix for the 20-word definitions.

Experiment 2

Experiment 2 was conducted as a replication and extension of Experiment 1. There were two major differences in the procedures. First, the tapes were constructed in a different fashion. In Experiment 1, the tapes were recorded naturally, without precisely controlling the delay interval between the definitions and suffix or the intonation patterns of the definitions across the three suffix conditions. In Experiment 2, this precise control was attained through computer digitization of the stimuli. Second, whereas subjects transcribed the definitions as they were presented in Experiment 1, in Experiment 2 they did not begin their written recall until the suffix was presented. It is possible that with the delayed-recall procedure, subjects might be more likely to integrate the linguistic information across the sentence, making the procedure more akin to natural speech processing. Theoretically, this could decrease the reliance on a surface-level representation and therefore decrease the influence of the stimulus suffix.

Method

Subjects. Twenty-four undergraduates from the University of South Carolina participated in this experiment for partial fulfillment of a course requirement.

Materials. To construct the tapes for this experiment, the definitions were first recorded in a male voice, always with the nonword suffix (*snoke*) appended to the end of each definition in the same voice. These sentences and suffixes were then digitized and stored separately in a Godbout microcomputer by using the CPM/86 operating system and an I/O Technology, analog-to-digital converter connected to a 24 db/Octave lowpass filter set at 5000 hz (to eliminate sampling noise). This procedure yielded same-voice suffixes that always were spectrally matched to the corresponding sentences (which naturally varied in pitch somewhat from trial to trial). An equal number of tokens of the suffix spoken in a female voice also were recorded and digitized, as was the 500-hz, 500-ms tone.

The tapes for use in the experiment were constructed via digital-to-analog conversion, with the analog signal channeled through the lowpass filter. The definitions and suffixes were recorded on these tapes in the same arrangements as in the first experiment. However, in this experiment, unlike Experiment 1, the recording method permitted the silent period between the definition and suffixes to be set at precisely 500 ms.

Procedure. The procedure used in Experiment 2 was precisely the same as in Experiment 1 except that subjects were instructed to begin their written recall only after they heard either the nonword *snoke* or the tone.

Results

Figure 3 displays the mean percent correct for the 15-word definitions as a function of serial position and suffix condition. There are two points to note in Figure 3. First, there is again the U-shaped serial position function found in Experiment 1. Second, unlike Experiment 1, there appears to be a relatively small impact of suffix condition. This effect appeares to be restricted to the same-voice condition versus tone-control condition across the last few serial positions.

An ANOVA on the 15-word definitions yielded a significant main effect of serial position, $F(14, 322) = 34.22$, $MS_e = 0.0209$. The main effect of suffix type, $F(2, 46) = 1.34$, $MS_e = 1.34$, $MS_e = 0.061$, and the interaction between suffix type and serial position, $F(28, 644) = 0.96$, $MS_e = 0.0182$, did not reach significance. However, it should be noted that a separate analysis on Serial Positions 13–15 yielded a significant main

[1] Of course, one reason that recall is relatively high in Experiment 1 is that subjects began recalling the list as soon as the first word was presented. However, this was not the only factor that produced the relatively high recall performance in Experiment 1. As shown in Experiment 2, when recall is delayed until the suffix or tone is presented, performance levels are still fairly high, that is, 70% for the 15-word definitions and 55% for the 20-word definitions.

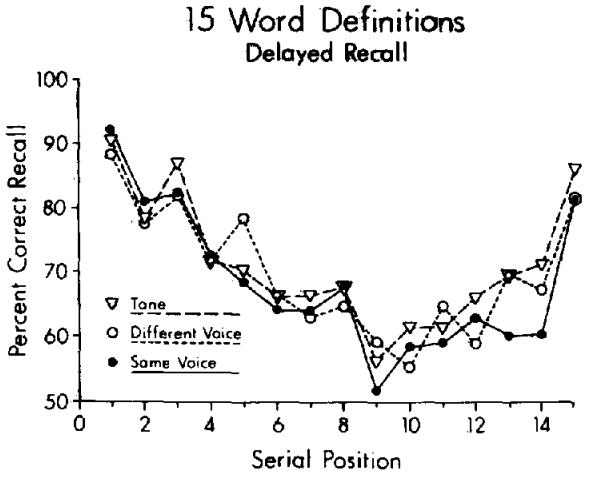

Figure 3. Mean percent correct delayed recall as a function of suffix type and serial position for the 15-word definitions.

effect of suffix type ($p < .05$). As shown in Figure 3, this is due to the difference in performance between the same-voice condition and the tone-control condition; the different-voice condition did not significantly differ from the remaining two conditions.

The mean percent correct recall for the 20-word definitions as a function of suffix type and serial position is displayed in Figure 4. There are four points to note here. First, one again finds the U-shaped serial position function. Second, there is relatively little difference between the same-voice and the different-voice conditions. Third, the advantage of the tone-control condition, compared with both of the voice-suffix conditions, occurs for the last half of the list. Fourth, there is a quite large (17%) advantage of the tone-control condition over the same-voice conditions at the last serial position.

An ANOVA on the 20-word definitions yielded a significant main effect of suffix type, $F(2, 46) = 9.59$, $MS_e = 0.0422$, and serial position, $F(19, 437) = 50.54$, $MS_e = 0.0251$. The interaction between these two variables did not approach significance, $F(38, 874) = 0.64$, $MS_e = 0.0230$. A post hoc comparison indicated that there was a highly significant suffix effect across the last 10 serial positions ($p < .001$), whereas, for the first 10 serial positions, this effect did not reach significance. As shown in Figure 4, the impact of suffix type is completely due to the difference between the tone-control and the two voiced-suffix conditions. In fact, the same-voice condition overall differed by less than 1% from the different-voice condition.

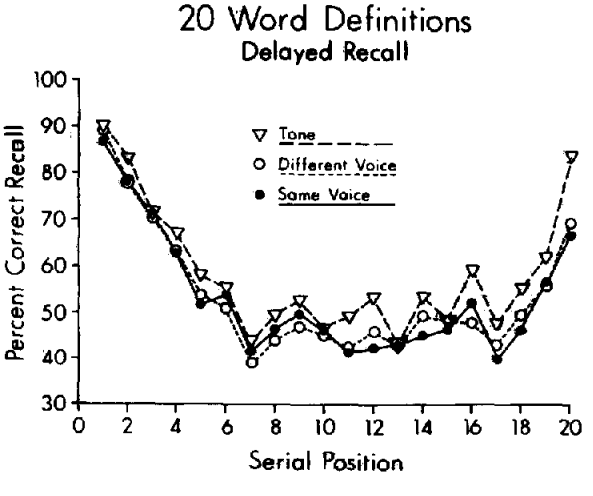

Figure 4. Mean percent correct delayed recall as a function of suffix type and serial position for the 20-word definitions.

As in Experiment 1, there appears to be a larger impact of the suffix at the last serial position for the 20-word definitions compared with the 15-word definitions. This was again confirmed by a separate ANOVA on the last serial position that yielded a significant Suffix Type × Definition Length interaction, $F(2, 46) = 4.23$, $MS_e = 109.4$. As in Experiment 1, the two voiced suffixes had more of a disruptive influence on the recall of the 20-word definitions than in the recall of the 15-word definitions.

Discussion

The results of Experiment 2 both extend and constrain the interpretation of the results from Experiment 1. First, a significant suffix effect was again obtained with linguistically coherent materials. Thus, one cannot totally attribute the pattern of data obtained in Experiment 1 to the recall procedure used or to the particular method of constructing the stimulus tapes. Also, as in Experiment 1, the effect was larger for the 20-word definitions than for the 15-word definitions. Thus, the hypothesis that the suffix produces greater disruption when there is greater load placed on the memory system was verified in this study.

However, the results of Experiement 2 did not yield any evidence of the voice-specific impact of the suffix that was obtained in Experiment 1. The suffix effect in Experiment 2 was primarily due to the difference between the tone-control condition and the two voiced-suffix conditions. One obvious difference across Experiments 1 and 2 was the nature of the stimulus tape construction. It appears that when the characteristics of the stimulus materials are better controlled, as in Experiment 2, there is little impact of same- versus different-voice suffixes. In fact, it is possible that by controlling these dimensions (i.e., duration and amplitude), one sufficiently reduces the differences across same- and different-voice suffixes to eliminate the effect. Hence, in our attempt to introduce strict control over the characteristics of the suffix items, we may have eliminated some of the differences that typically occur across same- and different-voices suffixes that produce the voice-specific impact of the suffix.

A second and more intriguing reason why in Experiment 2 there was an elimination of the voice-specific impact of the suffix found in Experiment 1 is the difference in recall procedure. In Experiment 2, a delayed-recall procedure was used, whereas in Experiment 1 an immediate-recall procedure was used. Possibly, a delayed-recall procedure is more akin to natural language processing. That is, in the immediate recall procedure used in Experiment 1, there may have been insufficient time for the subjects to fully comprehend the sentences

during transcription, whereas in the more normal (delayed) recall procedure of Experiment 2, the subjects should have had the complete syntactic/semantic information available for comprehension before recall began. Thus, there may have been less reliance on surface-level representations in the delayed-recall procedure, and as predicted, there was a smaller suffix effect.

The results of Experiments 1 and 2 indicate that one can produce suffix effects with linguistic materials such as naturally read definitions of low-frequency words. Of course, definitions of low-frequency words could still be viewed as linguistically impoverished, given that they involve a large proportion of adjectives and nouns that cannot be predicted from the linguistic context. The effects of speech suffixes on the recall of these definitions thus may have emerged only because of the low level of linguistic predictability in these stimuli, much as in more typical list-learning experiments. Hence, the extension to more natural language processing may still be quite limited in the first two experiments.

Experiment 3

In Experiments 3 and 4, an attempt was made to obtain a suffix effect with sentential materials that have relatively greater linguistic constraint. In this experiment, the materials were consecutive sentences adapted from a short story for primary-grade-level readers. Unlike the dictionary materials used in the first two experiments, each sentence placed general contextual constraints on the sentences that followed. Moreover, both vocabulary and sentence structures were relatively simple.

In the results of the first two experiments, the suffix effects were larger for the longer definitions. Therefore, in order to maximize the likelihood of obtaining a suffix effect, the sentences used in Experiments 3 and 4 (24 words in length) were slightly longer than those of Experiments 1 and 2 (15 and 20 words in length). The important question of interest is whether making the materials more linguistically coherent will override the disruptive impact of the voiced suffixes.

In order to make comparisons with Experiments 1 and 2, Experiment 3 involved immediate recall, similar to Experiment 1, and Experiment 4 involved delayed recall, similar to Experiment 2.

Method

Subjects. Twenty-four undergraduates at the University of Massachusetts at Amherst participated in this experiment as partial fulfillment of a course requirement.

Materials. The stimulus materials were based on a short story written for primary-grade readers. A total of 45 sentences were used. There were 36 target sentences and 9 filler sentences. The target sentences all contained 24 words, and the fillers contained from 15 to 22 words. The filler sentences were included to produce some variation in sentence length. In addition six practice sentences were constructed that were unrelated to the study. The practice sentences ranged in length between 20 and 24 words.

Each sentence was counterbalanced across the three suffix conditions (i.e., same voice, different voice, and tone), thereby again generating three separate tapes. The three tapes were constructed in the same fashion as in Experiment 2. There was a 45-s blank pause at the end of each sentence to allow time for recall.

Procedure. Subjects were first given the six practice sentences to acquaint them with the recall procedure. They were instructed to begin writing each sentence as soon as it began and to use the suffix item simply to mark the end of the sentence. After the practice sentences were presented, subjects were told that they would receive a series of sentences that formed a story. They were told to begin writing each sentence as soon as it began. Subjects were also told that they should try to comprehend the overall content of the story because at the end of the experiment they would be asked questions about the story. In all other ways the procedure of this experiment was the same as that used in Experiment 1.

Results

Figure 5 displays the mean percent correct as a function of serial position and suffix condition. There are two points to note in Figure 5. First, one again finds the U-shaped serial position curve, with recall performance relatively high for 24 word lists. The relatively high overall recall suggests that subjects were attending to the sentence structure, that is, not treating the sentences as lists of unrelated words. Second, and more important, there is only a small influence of the suffix condition, which appears to be largest at the last two serial positions.

The ANOVA yielded a main effect of serial position, $F(23, 529) = 86.95$, $MS_e = 0.0128$. Neither the main effect of suffix type, $F(2, 46) = 2.89$, $MS_e = 0.0462$, $p < .10$, nor the interaction between suffix type and serial position, $F(46, 1058) = 0.91$, $MS_e = 0.0120$, reached significance.

A post hoc analysis on the last two serial positions did not yield a significant influence of suffix condition, $F(2, 46) = 3.01$, $MS_e = 0.0184$. However, it should be noted that a post hoc comparison of only the same-voice versus the tone-control condition at the last two serial positions did reach significance ($p < .05$). The different-voice condition did not significantly differ from either the same-voice condition or the tone-control condition.

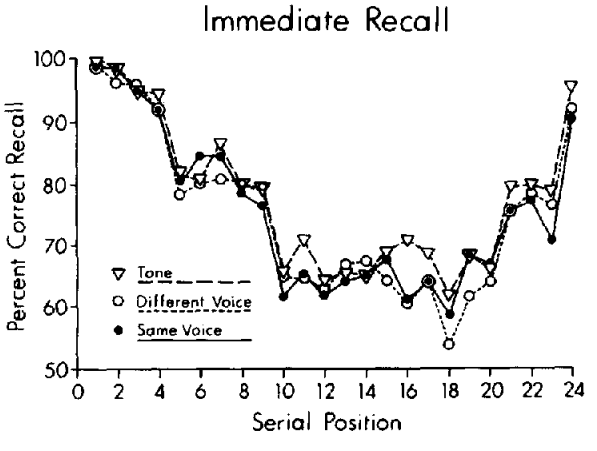

Figure 5. Mean percent correct immediate recall as a function of suffix type and serial position for the story sentences.

Discussion

The results of Experiment 3 indicate that when one provides easily comprehensible linguistic stimuli that are composed of a simple vocabulary and that include considerable contextual constraints both within a sentence and across sentences, there is only a small impact of suffix condition. Taken together, Experiments 1–3 suggest that the nature of the linguistic material modulates the appearance of a suffix effect and is more critical than sentence length. That is, Experiment 3 did not yield a significant main effect of suffix condition with sentences lengths of 24 words, whereas Experiments 1 and 2 did yield significant suffix effects with definition lengths of 15 and 20 words, with the effects being larger for the 20-word definitions.

Before further discussing the implications of Experiment 3, a fourth experiment will be presented which involves a delayed-recall procedure similar to Experiment 2. As noted earlier, the delayed-recall procedure should produce a greater reliance on higher level linguistic information.

Experiment 4

Method

Subjects. Twenty-four undergraduates at the University of Kentucky participated as partial fulfillment of a course requirement.

Materials and procedure. The same materials and procedures used in Experiment 3 were used in this experiment. The only exception was that subjects were instructed to begin their recall when they heard the nonword *snoke* or tone.

Results

Figure 6 displays the mean percent correct recall as a function of serial position and suffix condition. Again, one can see the U-shaped serial position effect. More important, the tone control condition produces a slight advantage over the same-voice and different-voice conditions that appears to extend across most serial positions. In addition, there is little, if any, consistent difference between the same- and different-voice conditions.

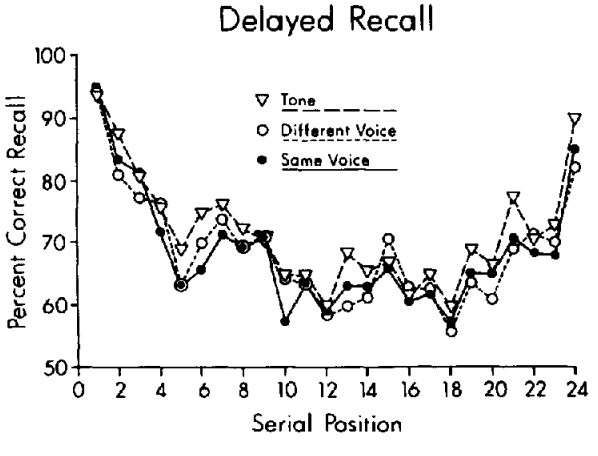

Figure 6. Mean percent delayed recall as a function of suffix type and serial position for the story sentences.

The ANOVA yielded a main effect of suffix type, $F(2, 46) = 4.50$, $MS_e = 0.0413$, and serial position, $F(23, 529) = 39.08$, $MS_e = 0.0147$. However, the interaction between suffix type and serial position did not approach significance, $F(46, 1058) = 0.75$, $MS_e = 0.0140$. An ANOVA on the last two serial positions did not yield a significant main effect of suffix condition, $F(2, 46) = 1.72$, $MS_e = 0.0237$; however, an ANOVA on the last six serial positions did yield a significant main effect of suffix type, $F(2, 46) = 3.96$, $MS_e = 0.0240$. The influence of the suffix was again entirely due to the difference between the two voiced-suffix conditions and the tone-control condition. In fact, the same-voiced condition differed from the different-voice condition by less than 1%.

Discussion

The results of Experiment 4 are quite consistent with those obtained in Experiment 3. With materials that include a high degree of cross-sentence linguistic constraint, there was only a small influence of suffix condition. In fact, a post hoc comparison at the last serial position did not yield a significant main effect of suffix condition in either Experiment 3 or Experiment 4. In addition, there was no evidence of a difference between the same-voice and different-voice suffix conditions in either experiment.

Thus, it appears that the high degree of constraint imposed by the story structure in Experiments 3 and 4 in large part eliminated the suffix effect that had been obtained with the definitions in Experiments 1 and 2. This can be seen most clearly in Table 1, where the mean percent correct and mean suffix effects are displayed as a function of list materials (definitions vs. sentences from a story) and suffix condition for the last serial position. As shown in Table 1, there is a considerable reduction in the suffix effects for the story sentences as compared with the definitions. In support of this observation, an ANOVA on the last serial position for Experiments 1 and 2 (collapsed across the length factor) versus the last serial position of Experiments 3 and 4 yielded a significant interaction between stimulus materials (i.e., definitions vs. story sentences) and suffix type, $F(2, 184) = 3.14$, $MS_e = 118.8$.

Table 1
Mean Percent Correct Recall and Suffix Effect at the Last Serial Position as a Function of List Materials and Suffix Type

List materials	Suffix type			Suffix effect	
	T	DV	SV	T–SV	DV–SV
Definitions (Exp. 1 and 2)	89.4	81.8	76.7	12.7	5.1
Story sentences (Exp. 3 and 4)	92.6	86.9	87.5	5.1	−0.6

Note. *T* refers to the tone-control condition; *DV* refers to the different-voice condition; and *SV* refers to the same-voice condition. Exp. = experiments.

General Discussion

The results of the present experiments are quite clear. First, the results from Experiments 1 and 2 yielded signficant suffix effects for definitions that were spoken with natural rhythm and intonation patterns under two different recall procedures. Second, although the results of Experiment 1 yielded quite large effects of suffix condition, the results of Experiment 2 produced a relatively smaller overall effect of suffix condition and eliminated the difference between same-voice suffixes and different-voice suffixes. The decreased impact of suffix condition in Experiment 2 is attributed to better control of the acoustic characteristics of the suffixes and/or a recall procedure that emphasized more integration across the words within the definitions. Third, the results from Experiments 3 and 4 indicated that cross-sentence linguistic constraint, more typical of a natural language processing situation, further reduced the impact of the suffix.

The present results are important for two reasons. First, the results indicate that one can produce suffix-type effects with linguistically coherent materials. Thus, the suffix effect is not totally restricted to lists of digits, letters, or unrelated words that are presented with a redundant presentation rate and in a monotonic fashion. We feel that this is important because, as noted earlier, researchers have speculated that a memory story similar to that influenced by the stimulus suffix also plays a role in ordinary language processing situations (e.g., Crowder, 1976). To our knowledge, the present study is the first empirical extension of the suffix paradigm to more linguistically constrained materials.[2]

Second, and more important, the results indicate that as one increases the similarity of the to-be-remembered materials to natural discourse, the suffix effect is greatly reduced. Support for this second observation can be found by comparing the present study with previous studies of the suffix effect and by comparing the experiments within the present study. For example, the suffix effects produced in the present study are quite small compared with studies with less linguistically constrained materials. Specifically, if one considers the difference in percent correct between the same-voice condition and the tone-control condition at the last serial position, the effect is only 13% in Experiments 1 and 2. This effect was reduced to only 5% with the more linguistically coherent materials in Experiments 3 and 4. Suffix effects with less constrained materials are very robust, on the order of 30% or greater (e.g., Ayers, Naveh-Benjamin, & Jonides, 1986; Balota & Engle, 1981; Morton et al., 1971). Moreover, when the physical characteristics of the same-voice and different-voice suffixes were carefully controlled in Experiments 2, 3, and 4, there was no impact of voice similarity. In contrast, voice similarity has also been shown to produce quite robust effects (approximately 15%) with more impoverished materials (e.g., Balota & Duchek, 1986; Morton et al., 1971; Watkins & Watkins, 1980). Thus, it appears that both across different studies and across the experiments in the present study, the strikingly robust suffix effect is considerably reduced by increasing the similarity to a more natural language processing situation.

Two recent studies addressing auditory memory in subjects that vary in comprehension performance are quite consistent with the notion that increasing linguistic structure can lead to a reduction in the suffix effect. First, Sipe and Engle (1986) found larger suffix effects in poor readers than in good readers. Second, Greenberg and Roscoe (1988) found that subjects who performed relatively low on a listening comprehension test produced larger suffix effects than individuals who performed relatively high on such tests. In fact, Greenberg and Roscoe argued that individuals with weaker listening comprehension may depend more upon vulnerable sensory codes in auditory memory while those with better comprehension may rely more on stable higher order codes, that is, linguistic structure. A similar argument may be made for the Sipe and Engle results. Hence, just as less constrained materials (digits and unrelated words produced with a redundant presentation rate and in a monotone fashion) produce larger suffix effects, individuals who are poorer at processing higher level linguistic information (poor readers and listeners) also produce larger suffix effects.

As described in the introduction, current notions in the language processing literature concerning rather rapid integration of words into higher order linguistic structure are quite consistent with the finding of a reduced suffix effect with increasing linguistic structure. That is, if the suffix effect is due to interference with a surface-level representation as generally conceived (e.g., Crowder, 1976), then such a representation might not be as crucial in a more natural language processing situation. In fact, one could argue that even with the story sentences used in Experiments 3 and 4, the task demands of verbatim recall of the sentences produced a situation that has less linguistic constraint than did a more natural listening situation. Thus, the small suffix effects that were obtained in Experiments 3 and 4 were, if anything, probably *overestimating* the suffix effect in more natural listening situations. At this level, the present results are quite consistent with the notion of rather immediate on-line integration of words within a higher order linguistic structure.

However, we clearly do not wish to dismiss the possibility of suffix-type effects in *all* natural language processing situations. In fact, our original motivation for conducting these experiments was to determine whether suffix-type effects might occur when an instructor is giving a lecture. For example, consider the situation where an instructor produces a definition for a new term to a class. The student/listener may be trying to transcribe part or all of a given utterance *A* at the same time that the instructor is producing utterance *B*, which, counter to the instructor's intention, may function as a stimulus suffix. Clearly, the results of Experiment 1 indicate that with materials such as definitions of new terms, one might expect substantial suffix effects. Here, the listener is not simply attending to the meaning of the definition, but rather may be attending to the surface details in order to accurately tran-

[2] One might argue that *linguistic* constraint per se is not the crucial factor in the present experiments, but rather some other form of informational constraint in the materials. Clearly, this is possible. However, our goal in the present study was to address a particular type of constraint (i.e., linguistic) that plays a role in natural language processing. Further studies are needed to address whether *all* types of constraint would produce the same pattern of data.

scribe the definition, the precise type of information that the suffix may be most likely to interfere with. Under this rather common listening situation, and on the basis of the present results, one might expect substantial suffix effects.

In addition to situations where the listener may be attending to surface-level information, there are further situations where the listener may be forced to recompute the meaning of a sentence, and hence reaccess surface-level representations. In fact, the work by Frazier and Rayner (1982) in reading is quite consistent with this possibility. As noted earlier, in the Frazier and Rayner study, subjects received *garden-path* sentences, that is, sentences that lead to a given syntactic analysis that was later disconfirmed. Frazier and Rayner found that the first fixation in the disambiguating region in such sentences produced increases in fixation duration for those sentences that were inconsistent with preferred parsing strategies. Moreover, when subjects were forced to recompute the syntax of the sentence, Frazier and Rayner found that subjects were more likely to make regressive eye-movements to earlier words in the sentence. Such regressive eye-movements might be viewed as an attempt to recapture the surface-level information concerning the earlier words in the sentence. Similar effects might be expected in a listening situation. That is, as long as each word is correctly integrated into a higher order linguistic structure, there should be little breakdown in performance due to a subsequent suffix item. However, if subjects are lead down the garden path and forced to reaccess surface-level information from earlier words in the sentence, then one might expect quite large suffix effects. We will have to await future research to determine whether one finds similar effects in listening as in reading.

Finally, we should note a recent study by Jakimik and Glenberg (1990), who have also attempted to extend the work on auditory memory to more natural sentential materials. Jakimik and Glenberg addressed the modality effect and the comprehension of anaphor, in which temporal order is important, for example, when someone uses an anaphor such as "the latter approach." On the basis of Glenberg and Swanson's (1986) temporal distinctiveness theory, these researchers argued that temporal order is more accurately represented in the auditory modality than in the visual modality. Hence, the resolution of temporal anaphor should be better when the materials are presented auditorily compared with visually. Consistent with this view, these researchers found that there was a modality effect in the resolution of temporal anaphors (e.g., *the latter approach*), whereas there was no evidence of a modality effect for semantically based anaphors (e.g., *the medical approach*). We clearly agree with Jakimik and Glenberg that an interface between research addressing auditory memory that has developed within the verbal learning tradition and work within psycholinguistics may provide some important insights into natural language processing.

References

Ayers, T. J., Naveh-Benjamin, M., Jonides, J. (1986). Recognition of the stimulus suffix. *Journal of Memory and Language, 25*, 619–626.

Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.

Balota, D. A., & Duchek, J. M. (1986). Voice-specific information and the 20-second delayed suffix effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 509–516.

Balota, D. A., & Engle, R. W. (1981). Structural and strategic factors in the stimulus suffix effect. *Journal of Verbal Learning and Verbal Behavior, 20*, 346–357.

Caplan, D. (1972). Clause boundaries and recognition latencies for words in sentences. *Perception & Psychophysics, 12*, 73–76.

Cheng, C. (1974). Different roles of acoustic and articulatory information in short-term memory. *Journal of Experimental Psychology, 103*, 614–618.

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–163). Hillsdale, NJ: Erlbaum.

Cowan, N. (1984). On short and long auditory traces. *Psychological Bulletin, 96*, 351–370.

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.

Crowder, R. G. (1978). Mechanisms of auditory backward masking in the stimulus suffix effect. *Psychological Review, 85*, 502–524.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics, 5*, 365–373.

Darwin, C. J., & Baddeley, A. D. (1974). Acoustic memory and the perception of speech. *Cognitive Psychology, 6*, 41–60.

Fillenbaum, S. (1986). Memory for gist: Some relevant variables. *Language and Speech, 9*, 217–227.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210.

Glenberg, A. W., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 3–15.

Greenberg, S. N., & Engle, R. W. (1983). Voice changes in the stimulus suffix effect: Are the effects structural or strategic? *Memory & Cognition, 11*, 551–556.

Greenberg, S. N., & Roscoe, S. (1988). Echoic memory interference and comprehension in a foreign language. *Language Learning, 38*, 209–219.

Greene, R. L., & Crowder, R. G. (1984). Modality and suffix effects in the absence of auditory stimulation. *Journal of Verbal Learning and Verbal Behavior, 23*, 371–382.

Jakimik, J., & Glenberg, A. (1990). *Verbal learning meets psycholinguistics: Modality effects in the comprehension of anaphora*. Manuscript submitted for publication.

Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior, 10*, 409–416.

Hanson, V. L., & Bellugi, U. (1982). On the role of sign order and morphological structure in memory for American Sign Language sentences. *Journal of Verbal Learning and Verbal Behavior, 21*, 621–633.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354.

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

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science, 189*, 226–228.

Marslen-Wilson, W. D. (1989). Access and integration: Projecting sound onto meaning. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 3–24). Cambridge, MA: MIT Press.

Marslen-Wilson, W. D., Tyler, L. K., & Seidenberg, M. (1978). Sentence processing and the clause boundary. In W. J. M. Levelt & G. B. Flores d'Arcais (Eds.), *Studies in the perception of language* (pp. 219–246). New York: Wiley.

Massaro, D. W. (1970). Perperceptual auditory images. *Journal of Experimental Psychology, 85*, 411–417.

Massaro, D. W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review, 70*, 124–145.

Massaro, D. W. (1975). *Experimental psychology and information processing*. Chicago: Rand McNally.

McNeill, D., & Repp, B. (1973). Internal processes in speech perception. *Journal of the Acoustical Society of America, 53*, 1320–1326.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Morton, J., Crowder, R. G., & Prussin, H. A. (1971). Experiments with the stimulus suffix effect. *Journal of Experimental Psychology, 91*, 169–190.

Penney, C. G. (1989). Modality effects and the structure of short-term memory. *Memory & Cognition, 17*, 398–422.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13*, 253–260.

Potter, M. C., & Faulconer, B. A. (1979). Understanding noun phrases. *Journal of Verbal Learning and Verbal Behavior, 18*, 509–521.

Sachs, J. D. S. (1967). Recongition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2*, 437–442.

Sipe, S., & Engle, R. W. (1986). Echoic memory processes in good and poor readers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 402–412.

Von Eckardt, B., & Potter, M. C. (1985). Clauses and the semantic representation of words. *Memory & Cognition, 13*, 371–376.

Watkins, O. C., & Watkins, M. J. (1980). Echoic memory and voice quality: Recency recall is not enhanced by varying presentation voice. *Memory & Cognition, 8*, 26–30.

Received July 5, 1989
Revision received October 16, 1989
Accepted October 30, 1989 ■