

The Role of Verbal Output Time in the Effects of Word Length on Immediate Memory

NELSON COWAN, LINDA DAY, J. SCOTT SAULTS, TIM A. KELLER, TOM JOHNSON, AND
LUIS FLORES

University of Missouri

In three experiments, we examined the role of delays within overt verbal responding in causing effects of word length on immediate recall. Although a phonological memory decay mechanism has been implicated by past research on word length effects, the exact basis of the effect remains unclear. The added difficulty of recalling longer words could arise both while subjects attempt to rehearse words silently and while they attempt to repeat words aloud. To examine the latter mechanism, the lengths of words in the first and second halves of lists to be recalled were varied independently, and both forward and backward recall orders were used. Recall of each word was found to be influenced by the total pronunciation time for all items to be recalled prior to that word, although there was an additional advantage for the last item output. The results clarify and generally support the theory of the articulatory loop, and the method permits an improved analysis of immediate memory into decay-based and other factors. © 1992 Academic Press, Inc.

In a classic paper, Miller (1956) described how short-term memory is limited to a small number of items. That paper did not settle all of the mysteries of memory's capacity limitations, but rather helped to open up a vigorous line of research on the topic that is still ongoing. One of the most exciting subsequent developments has been the finding that there is a time limit to immediate verbal memory (Baddeley, Thomson, & Buchanan, 1975), either instead of (Schweickert & Boruff, 1986) or in combination with (Zhang & Simon, 1985) the item limit that Miller had proposed. Specifically, the amount of verbal material that a subject can recall is closely related to the amount that he or she can pronounce in 1.5 to 2 s. Differences in immediate memory across

individuals, across different materials, across ages, and across different languages all can be accounted for largely on the basis of differences in the rate at which items can be pronounced (Baddeley, Thomson, & Buchanan, 1975; Case, Kurland, & Goldberg, 1982; Hulme, Thomson, Muir, & Lawrence, 1984; Nicolson, 1981; Schweickert & Boruff, 1986; Standing, Bond, Smith, & Isley, 1980; Zhang & Simon, 1985).

The finding that is most relevant to the present research is that the immediate serial recall of word lists depends on the length of words in the list, defined as the time it takes to pronounce each word. In a seminal experiment with 5-word lists, Baddeley et al. (1975) found better performance for lists of shorter words.

Although the relation between speech rate and immediate memory is well established, the full theoretical explanation for this relation is still uncertain. Baddeley (1986) put forth the "articulatory loop hypothesis" that is the predominant current view on the matter. According to this simple account, verbal short-term memory

This research was supported by NIH Grant HD21338 awarded to Nelson Cowan. We thank Lora Kaye for collecting data and David Balota, Jerry Busemeyer, Peter Frensch, Richard Kunze, Mark McDaniel, Judith Wiley, and Noelle Wood for making helpful comments. Address correspondence and reprint requests to Nelson Cowan, Department of Psychology, 210 McAlester Hall, University of Missouri, Columbia, MO 65211. E-mail: psycowan@umcvmv.

depends upon an articulatory loop that includes two components; (1) a memory store in which incoming phonological information can be passively held, subject to decay at a fixed rate, and (2) a covert rehearsal process that can reactivate items in the passive store and thereby postpone the loss of these items through decay. The amount of phonological material that one can remember is assumed to equal the amount that can be reactivated by rehearsal, in a repetitive loop, before it decays from the passive store. At least in normal individuals, the rate of memory decay is assumed to be roughly 1.5 to 2 s, because Baddeley et al. (1975) found that, in this amount of time, a subject can pronounce about the number of items that fit within his or her memory capacity for a particular set of materials (also see Schweickert & Boruff, 1986; Standing et al., 1980). Stimulus differences and individual differences in memory both are assumed to result from differences in the rate of covert or overt articulation of items.

Although the concept of the articulatory loop has worked rather well (see Baddeley, 1986), there are some important issues that have not been resolved. The issue that is to be the primary concern of the present research is the source of the effects of the item articulation rate on immediate memory. One possible source is the covert rehearsal of the words. The exact protocol of rehearsal is uncertain, but it has to be cumulative in that the words throughout the list must be kept active within the phonological buffer until the recall period arrives (e.g., for words from the beginning of the list, 5 to 10 s within the procedures of Baddeley et al.). At least with longer, clearly supraspan lists, there is positive evidence for a cumulative rehearsal strategy (e.g., Palmer & Ornstein, 1971). Such covert rehearsal is the articulatory activity that has been the primary focus of most past research on the articulatory loop.

The focus of the present research is another, potentially important type of articulatory activity that has not been adequately

examined in the past: the subject's articulation during the recall phase of the trial. During this phase, while the subject is pronouncing the initial items on a list, the representations of the other items may be lost from the phonological buffer through the proposed decay process. The faster that the list items could be pronounced, the larger would be the number of items that the subject could repeat before these items decayed.

In their research, Baddeley and his colleagues usually have dealt with the multiplicity of articulatory activities primarily by attempting to control output time, so as to allow an examination of rehearsal. For example, in their study of word length effects in immediate memory, Baddeley et al. (1975) used a paced recall procedure in some experiments, in which subjects were to recall the words in synchrony with a metronome tick so that short and long words would be recalled at the same rate. Similarly, Baddeley, Lewis, and Vallar (1984) used paced, written recall of the first three letters in each word. The presence of a word length effect in these conditions and the abolishment of the effect with articulatory suppression (Baddeley et al., 1984) indicates that at least part of the effect of word length can be attributed to rehearsal. However, this research has not attempted to determine if output articulation time also plays a role in recall.

The results of Baddeley and Hull (1979) provide preliminary evidence suggesting that there could be an effect of output time. They presented lists of digits for recall, each followed by a stimulus suffix (presented by the experimenter) or a response prefix (pronounced by the subject) that was interposed between the list and the subject's spoken recall of the list. Performance on prefinal items was poorer when the suffix or prefix was longer. However, there is some question about the nature of the effects obtained in that study. Word length was manipulated by varying the number of syllables in the suffix or prefix (from one to

five syllables), so the effects could have been based on the amount of phonological interference from these items rather than on the amount of time that elapsed as they were pronounced.

It is possible to construct lists of words that are relatively short versus long in pronunciation time but do not differ in the number of phonemes or syllables, as did Baddeley et al. (1975, Experiments 4 & 5) in their investigation of immediate memory and articulation speed. One theoretically could use the controlled stimulus sets of Baddeley et al. (1975) as prefixes or suffixes in a study like that of Baddeley and Hull (1979) to examine the effects of imposed response delays more cleanly. Unfortunately, though, the differences in duration between those controlled short and long word sets were, of necessity, not very large. We were not confident that the differences between short and long suffixes matched for the number of phonemes and syllables would be large enough to produce robust effects, and so did not take that experimental route.

Instead, our experiments followed the procedure of Baddeley et al. (1975, Experiment 5). Those investigators found that memory was poorer for lists of words that took longer to pronounce, even when the shorter and longer lists were matched for the number of phonemes and syllables and for word frequency. Further, the difference in performance was well accounted for by the difference in the time it took to pronounce words within the two sets. The success of the very subtle manipulation of word length in that study probably depended upon the fact that differences in word length between the two sets of words accumulated throughout the five-word lists to be recalled.

We used word lists similar to those of Baddeley et al. (1975, Experiments 4 & 5), but we manipulated the lengths of words in the first versus the second half of the list independently (in our Experiments 2 & 3). We also examined recall of items in both

the forward and backward directions (Experiment 3). In combination, these manipulations permitted an assessment of the effects of item articulation rate at the time of recall. The basic notion is that the length of any word could affect memory performance in two ways. First, short words might be remembered better than long words, even within a mixed list. Additionally, though, a word's length might affect the recall of some of the other words on the list. Specifically, a word taking longer to pronounce would impose a longer response delay on other words, permitting memory for those other words to decay further. These are not mutually exclusive possibilities.

Before launching into an investigation of output delay, we found that we first had to address a potential methodological problem within the well-known and otherwise elegant study of Baddeley et al. (1975, Experiments 4 & 5), which we wished to adapt to our own purposes.

EXPERIMENT 1

The experiments of Baddeley et al. (1975, Experiments 4 & 5) have provided the most decisive evidence in favor of a decay-based account of immediate verbal memory. An effect of word length was obtained under conditions that ruled out accounts based on the larger phonetic content of longer words. In these experiments, short and long words were equated in the number of syllables and phonemes, but still differed in the amount of time that it took to pronounce the words in the two sets. Lists of shorter words, which could be pronounced and covertly rehearsed more quickly, were commensurately better recalled.

Upon close inspection of the stimuli, though, we began to fear that there may have been an undue amount of phonetic similarity among the five words that were included in the long set (*coerce*, *harpoon*, *friday*, *cyclone*, *zygote*). Specifically, the last three of these words all have /aI/ as the

first, stressed vowel, and the last two share both of their vowels. Although there is no accepted metric for the degree of phonological similarity among lexical items, it can easily be argued that there was less phonetic similarity among the words included in the short-word set (*pectin*, *pewter*, *phallic*, *bishop*, *wicket*), in which only the last two of these words share a stressed vowel. It has long been known that the phonological similarity between items impedes serial recall (Baddeley, 1986; Conrad, 1964), and it has recently been found that, in contrast to this effect on memory, pronunciation times are no slower for similar than for dissimilar lists (Hulme & Tordoff, 1989; Schweickert, Guentert, & Hersberger, 1990). Therefore, if phonological similarity between items inadvertently was a factor in the Baddeley et al. (1975, Experiments 4 & 5) findings, that could have distorted the estimate of the relation between word length and serial recall.

We composed a different set of words in the same way that Baddeley et al. did, by selecting certain words from a larger set used in their earlier experiments. However, we paid special attention to minimizing the degree of phonological similarity between words in a set. This stimulus set allowed us to determine if phonological similarity between items could have accounted for all of the effect of word length obtained by Baddeley et al. (1975, Experiments 4 & 5).

Method

Subjects. The subjects were 16 college students (4 male, 12 female) who had no known hearing losses and received course credit for their participation.

Apparatus and stimulus materials. A difficulty for this and the previous research is that neither linguists nor psycholinguists totally agree on the phonemic analysis of the English language. However, we obtained perfect agreement among four local, trained transcribers, two of whom were among the present authors (N.C. and L.D.). Thus, the

words that were included in the "short" set, along with broad transcriptions according to our dialect (shown with periods between phonemes), are *decor* /d.eI.k.ə.r/, *hackle* /h.æ.k.ə.l/, *wiggle* /w.I.g.ə.l/, *pewter* /p.j.u.t.ə.r/, and *ember* /ε.m.b.ə.r/. The "long" set included the words *zygote* /z.aI.g.o.t/, *voodoo* /v.u.d.u/, *coerce* /k.o.ə.r.s/, *morphine* /m.ə.r.f.i.n/, and *humane* /h.j.u.m.eI.n/. The word sets are approximately matched for word frequency. Among the 26 phonemes in each word set, 19 phonemes in the short set are unique (i.e., occur within only one word in the set) and 21 phonemes in the long set are unique. Therefore, any advantage for short words should occur in spite of, rather than because of, the degree of phonemic similarity.

Each word was printed in black lettering 1 cm high on a laminated, 8 × 13 cm (3 × 5 in.) index card. For the reading rate tests, words were presented in lettering 0.5 cm high, with 0.5 cm blank space between words vertically, in the same arrangement that Baddeley et al. (1975) used (one page for short words and one for long words, each with 50 words arranged in two columns). During the memory phase of the experiment, the experimenter listened through headphones to a tape recording of tones spaced 2 s apart, in order to time the word presentations. Speaking and reading rates were timed with a stopwatch.

Procedure. The procedure was modeled closely after that of Baddeley et al. (1975, Experiment 5). Subjects were tested in a quiet room one at a time. The first phase of the experiment was the memory test. Each list to be recalled contained all five short or all five long items in a unique serial order. That is, the order of the five short and the five long words were rearranged to construct multiple trials of each length. The experimenter spoke the word "ready" and then showed the cards with the words one at a time, at a 2-s per item rate. The experimenter signalled that the list had ended by placing the last card face down, and then the subject was to attempt to immediately

repeat the words in their correct serial order. Subjects were told to remember the words by rehearsing them silently. Subjects received four series of five trials each (two series of short words and two of long words), with the four series ordered according to a latin square.

Following the memory test, speaking and reading rates were assessed. Half of the subjects received the speaking rate assessment first, and half received the reading rate assessment first. Of these groups, half received short words first within each of these assessments, and half received long words first.

In each trial within the speaking rate assessment, a sequence of three words from a complete set of five was randomly selected and presented visually for the subject to memorize and then repeat 10 times in a row, as rapidly as possible. The time from the beginning of the first repetition to the end of the tenth repetition was recorded. There were four trials with the short words and four with the long words, and those sets of four trials were averaged for each subject. In each trial within the reading rate assessment, the subject read the set of 50 words on a page as quickly as possible, and the reading time was recorded. As in the speaking rate assessment, the subject's score was the average of four trials for each word length.

Results

Proportion correct. The proportion correct was analyzed in a 2×5 ANOVA with word length (short vs. long) and serial position (1–5) as fixed-effect, within-subject factors. The results, which are shown in Fig. 1, replicated Baddeley et al. (1975) There was an effect of word length, $F(1,15) = 16.03, p < .002, MS_e = 0.02$, and also of serial position, $F(4,60) = 33.13, p < .001, MS_e = 0.01$. Unlike Baddeley et al., we also obtained a Word Length \times Serial Position interaction, $F(4,60) = 3.49, p < .02, MS_e = 0.01$. Performance was at ceiling

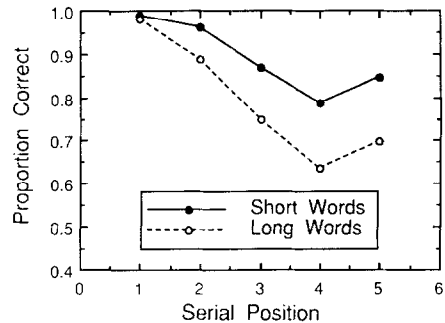


FIG. 1. Mean proportion correct recall of short words (solid line) versus long words (dashed line) by serial position in Experiment 1.

level for both word lengths at the first serial position, and the word length effect grew larger across serial positions (see Fig. 1).

Pronunciation times and memory capacity time estimates. In order to further check the loop hypothesis, speeded reading and speaking pronunciation times were used to estimate (as did Baddeley et al., 1975) what we will call the "memory capacity time." This will be defined as the mean duration that it would take a subject to pronounce as many words as he or she could remember, on the average, in a particular type of word list. It was calculated by multiplying the mean number of words correctly recalled on a particular type of trial by the pronunciation rate (as measured in the speaking or reading posttest) expressed in seconds per word. According to the theory of the articulatory loop, this time estimates the persistence of information in the phonological buffer, inasmuch as memory presumably is limited to the amount that the subject can articulate before it decays. The data of Landauer (1962), showing that overt and covert speech rates are similar, strengthens the legitimacy of this measure.

Separate estimates of memory capacity time were obtained using speaking versus reading estimates of pronunciation rate. According to Baddeley et al. (1975) and current theory (Baddeley, 1986), estimates for all conditions should be similar, falling in the range of 1.5 to 2.0 s. This was indeed the case. The estimates based on speaking

were, for short words, 1.55 s ($SD = 0.28$); for long words, 1.59 s ($SD = 0.36$). The estimates based on reading were, for short words, 1.65 s ($SD = 0.28$); for long words, 1.60 s ($SD = 0.36$). An ANOVA of these data with word length and speed estimate (speaking versus reading) as fixed-effect, within-subject factors revealed no significant differences.

Discussion

In this experiment, the results of Baddeley et al. (1975, Experiment 5) were replicated with special attention to the potentially confounding factor of phonological similarity. Thus, word length effects really do appear to occur even with the stimuli matched for the number of phonemes and syllables.

On the other hand, it should be noted that the mean difference between the overall proportion correct for short words (.89) versus long words (.79) was only about half of what Baddeley et al. observed in their Experiment 5. Therefore, it is possible that part of their effect was, in fact, a spurious consequence of the phonological similarity between items.

EXPERIMENT 2

The second experiment was conducted in order to determine if subjects' performance in the span task is consistent with what one would expect if output effects are present. Because words at the beginning of the list are the first to be repeated aloud, the faster these words could be pronounced the less time would elapse from the end of the list presentation until words at the later serial positions could be recalled. However, the closer one gets to the end of the list, the fewer subsequent words there are to be affected. Thus, because the length of each word can affect the recall of subsequent words on the list, but not the recall of prior words, the prediction is that the length of words in the earlier serial positions should

be more important for recall throughout the list.

There is a precedent to the method and interpretation that we used in this experiment. Watkins (1977) manipulated independently the word frequency of words in the first and second halves of lists to be recalled in a span procedure and found that the frequency of words in the first half of the list had a larger effect than the frequency of words in the second half of the list. Watkins interpreted this in terms of a dual memory system, but others (Schweickert & Boruff, 1986; Wright, 1979) have interpreted the findings in terms of pronunciation times. Specifically, it takes more time to pronounce low-frequency words than to pronounce high-frequency words. When low-frequency words occur in the first half of the list, this imposes a longer delay before the subject can recite the words in the second half of the list. According to this interpretation, similar results would be expected with a manipulation of first- and second-half word lengths rather than word frequencies.

For our purposes, our design affords one advantage over that of Watkins (1977). Because he used a span procedure in which lists varied in length and recall of each list was counted right or wrong, it was not possible to assess performance for each serial position separately. Because we used a fixed list length instead, we could examine performance at each serial position separately and so could determine if an item's length affects performance on that word itself, on other words in the list, or both.

Method

Subjects. The subjects were 16 college students (6 males, 10 females) who had not participated in Experiment 1.

Stimulus materials. New lists were formed from the same short and long word sets as were used in Experiment 1. In principle, we wanted to use lists in which the lengths of words in the first and second

halves of a list were manipulated independently. However, we wished to continue to use five-word lists as in Experiment 1, because we knew that substantial word length effects could be obtained in a sensitive range of measurement with that list length. Because five serial positions could not be divided evenly into first and second halves, the word in the medial serial position was selected so that its length matched the first half of the list in half of the trials and the second half of the list in the other half of the trials. This resulted in six trial types. With S = "short" and L = "long," the six types can be represented as 5S, 2S/3L, 3S/2L, 2L/3S, 3L/2S, and 5L (e.g., 3S/2L refers to a list with three short words followed by two long words). For the sake of an ANOVA, these six types were collapsed into four reflecting the first and second list halves: SS, SL, LS, and LL.

Procedure. Each subject received five blocks of 12 trials. Each block included two SS trials, two LL trials, and one of each of the four mixed-length trial types. Each block also contained four filler trials, in which two or three short words and a complement of long words were randomly arranged across serial positions. The purpose of the filler trials was to prevent subjects from learning to group the words within each list on the basis of the short and long word sets.

There was a designated set of word lists for each of five trial blocks. However, these five trial blocks were presented in a different random order for each subject, as were the 12 lists within each block.

Following the memory test, each subject took the same reading and speaking rate tests as in Experiment 1. In every other way, the procedure was the same as that of Experiment 1.

Results

Proportion correct. The results were entered into an analysis of variance with the length of words in the first half of the list

(short vs. long), the length of words in the second half of the list (short vs. long), and serial position (1–5) as fixed-effect, within-subject factors in a $2 \times 2 \times 5$ ANOVA. There was a main effect of the length of words in the first half of the list, $F(1,15) = 5.67, p < .03, MS_e = 0.03$. In contrast, the effect of the length of words in the second half of the list did not approach significance, $F(1,15) = 0.36, MS_e = .03$. Finally, there was an effect of serial position, $F(4,60) = 37.20, p < .001, MS_e = 0.03$. No other effect approached significance. The mean proportions correct at each serial position for lists with short versus long words in their first half are shown in Fig. 2.

The effect of the length of words at the beginning of the list was significant in a separate analysis of the first two serial positions, $F(1,15) = 4.81, p < .05$, and also in a separate analysis of the last two serial positions, $F(1,15) = 5.09, p < .04$. The latter effect demonstrates the interesting finding that performance even in the last two serial positions was influenced by word length in the *first* two serial positions. In contrast to these findings, neither analysis revealed an effect of the second-half word length.

The mean proportions correct for the four length combinations in the experiment, collapsed across serial positions, were: short–short, .84; short–long, .85; long–short, .82; and long–long, .78. There are two points to be made on the basis of these

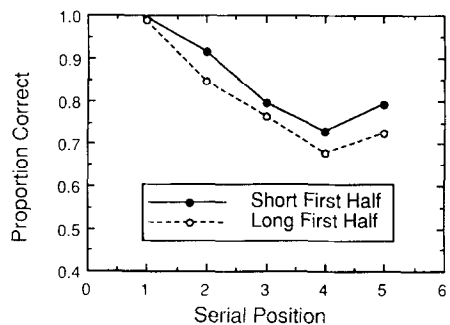


FIG. 2. Mean proportion correct recall of lists in which the first half contained short words (solid line) versus long words (dashed line) by serial position in Experiment 2.

TABLE 1
MEMORY CAPACITY TIME ESTIMATES FOR EACH
CONDITION IN EXPERIMENT 2

List/recall condition	Speaking estimate	Reading estimate
Short-short	1.71 (0.41)	1.71 (0.33)
Short-long	1.88 (0.35)	1.83 (0.30)
Long-short	1.81 (0.39)	1.75 (0.26)
Long-long	1.87 (0.36)	1.76 (0.28)

Note. Standard deviations in parentheses.

means. First, notice that performance in the short-short and long-long conditions nicely replicates the results of Experiment 1 (short, .89; long, .79). Second, although the effects of the length of words in the second half of the list in Experiment 2 did not reach significance, the means suggest that performance levels actually may have been lower in the long-long condition than in the other three conditions. The experiment may not have been powerful enough to detect the interaction of First-Half Word Length \times Second-Half Word Length, which was marginally significant in the overall analysis, $F(1,15) = 3.22$, $p < .09$, $MS_e = 0.02$. Therefore, we do not conclude that there is absolutely no effect of second-half word length. In fact, that would not be expected on the basis of an output account, because there should be some effect of the length of the item output fourth on recall of the item output last. The results do indicate, however, that first-half word lengths were more important than second-half word lengths.

Memory capacity times. Based on the observed speaking and reading pronunciation rates, memory capacity time estimates (the time it would take to pronounce the mean number of items that the subject could remember on a trial) were calculated as in Experiment 1. In the mixed-length conditions, the average of pronunciation times for short and long words together was used in the calculations. The results are shown in Table 1. As in Experiment 1, the estimates (which ranged from 1.71 s to 1.88

s) are in keeping with the generalization that subjects can remember about as much as they can pronounce in 1.5 to 2.0 s.¹

Memory decay function. Because estimates of pronunciation time were obtained (see above), it is possible to display an approximation of memory decay throughout the response period. Specifically, in Fig. 3, the mean proportion correct for each word is shown as a function of the estimated time that it took to pronounce all prior list items. That time was estimated by adding up the pronunciation time estimates for the particular combination of short and long words recalled before the word in question. The combination of prior short (S) and long (L) words contributing to the delay in respond-

¹ Although the memory capacity time estimates generally are similar to one another, there were some differences between them. In an ANOVA of these scores with the first-half word length, second-half word length, and speed estimate (speaking vs. reading) as fixed-effect, within-subject factors, there was a significantly lower estimate for lists with a short second half (1.75 s) than with a long second half (1.84 s), $F(1,15) = 5.32$, $p < .04$, $MS_e = .06$. This difference is to be expected if one carefully considers the consequences of the weak magnitude of second-half word length effects within the proportion of correct scores. Differences between the mean number correct used within the memory capacity time estimates for words with short versus long second halves would tend to be minimal, without any compensatory reduction in differences between the pronunciation rates. There also were several interaction effects in these data (First-half Word Length \times Estimate, $F(1,15) = 8.96$, $p < .009$, $MS_e = .003$; Second-half Word Length \times Estimate, $F(1,15) = 5.08$, $p < .04$, $MS_e = .004$; and First-half Word Length \times Second-half Word Length \times Estimate, $F(1,15) = 7.50$, $p < .02$, $MS_e = .000$). Based on the means in Table 1, memory capacity time estimates for long words appear to have been higher when based upon a speaking measure of speech rate than upon a reading measure, with less difference among estimates for short words. Subjects may have had difficulty rapidly repeating triads of the longer words, as is required within the speaking estimate. Our informal observations suggest that the long vowels within these words quickly cause neuromotor fatigue, as in tongue-twisters. Although these effects indicate that estimates of memory capacity time are imperfect and somewhat situation-specific, this should not overshadow the fact that all of the estimates still were quite similar.

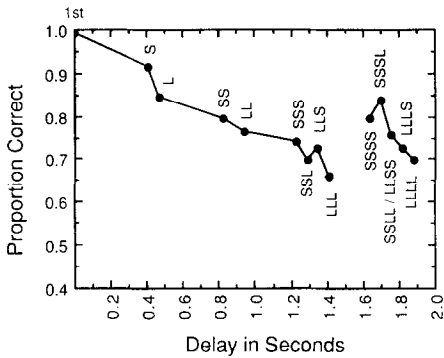


FIG. 3. For Experiment 2, the mean proportion correct recall as a function of the estimated total time of pronunciation of items recalled before the item in question. Pronunciation time estimates were derived from the average of reading and speaking times obtained for each subject following the memory test. Data labels indicate the sequence of short (S) and long (L) words occurring before the item in question.

ing is indicated by the label above or below each data point. It can be seen that the proportion correct was inversely related to the estimate of response delay, except that there was an additional advantage for items in the last serial position on the list (indicated by the last five data points in Fig. 3). To confirm this pattern, a linear fit to the data shown in Fig. 3 was obtained, but with a constant adjustment for the last serial position. This was accomplished by using a least-squares, nonlinear regression program (BMDPAR). Three parameters were entered into the regression equation: the slope and intercept of the linear equation, and the value of a constant that was added only for items at the last serial position of the list.

The fit to the data was excellent, $R^2 = .94$. However, not all of this variance can be unequivocally attributed to temporal delay and the final-position correction. The apparent delay effects across serial positions actually could reflect interference based on the number of items spoken. In order to establish that an unambiguous, within-position source of temporal delay also contributed to the model, a second model was constructed in which all items were assumed to cause equivalent effects,

regardless of their different pronunciation times. This model fit the data less well, $R^2 = .81$. A hierarchical regression analysis confirmed that an improvement in the fit arose from the inclusion of differential temporal estimates for short vs. long words, Partial $R^2 = .12$, $F(1,11) = 21.76$, $p < .001$.

Discussion

The results of this experiment clearly showed that immediate memory was not affected equally by the length of words in all parts of the list. Specifically, the word length effect was obtained with the manipulation of words in the first half of the list, but not with the manipulation of words in the second half of the list. Moreover, first-half word length had an effect throughout the list, not just in the first half of the list.

The findings are consistent with a process in which, within the response phase of each trial, the pronunciation of the first words on the list impose an output delay on later words. Although this result is entirely consistent with the model of the articulatory loop offered by Baddeley (1986), the manner in which word length has been manipulated in prior research (i.e., across the entire list) suggests that researchers have not usually anticipated results such as those described here.

The plot of proportion correct for each word as a function of the estimated total duration of pronunciation for preceding words on the list (Fig. 3) provides an interestingly direct view of the memory decay process. With one qualification, the results clearly suggest that recall of an item is inversely related to the duration of the preceding output delay, at least within the observed time frame of about 1.8 s. The qualification is that there was an added advantage for items in the last serial position. The slope of the decay function for the last serial position was similar to that observed at the earlier serial positions, but the Y intercept was higher. The basis of the added advantage for the final list item can-

not be ascertained from the present results alone, but the last experiment will help to clarify that finding.

EXPERIMENT 3

A third experiment was conducted because the results of Experiment 2 can be interpreted in two ways. Although they may indicate that immediate memory was affected by the delay in output caused by the lengths of the earlier list items, they alternatively could reflect a rehearsal process in which the first items on the list were rehearsed more often than later items. By enhancing the long-term memory representation of those early items or by causing the later items to be neglected, differential rehearsal conceivably could have produced the effect of the manipulation of word length in the first half of the list that was observed in Experiment 2.

The third experiment replicated Experiment 2, but with the inclusion of conditions in which the output duration and rehearsal accounts made opposing predictions. Specifically, in half of the trials, subjects were to repeat the list items in backward order. The cue to recall in a forward versus backward order was not presented until the termination of the list to be recalled. Therefore, subjects could not use a different rehearsal strategy during the input phases of these two types of trials. Thus, the distinction that is examined in the present experiment is one of articulatory processes based on the input order versus the required output order.

The account based on the differential access of words to rehearsal during the input phase would lead to the same predictions for either forward or backward recall. In contrast, according to the output delay account, it should be the length of the items repeated first that make a difference, as they help to determine the retention interval for words repeated later. In the backward condition, unlike the forward condi-

tion, it is the second half of the list that is to be repeated first.

Method

Subjects. The subjects were 32 college students (14 males, 18 female) who had not participated in either of the other two experiments.

Stimuli. The same stimulus set as in Experiment 2 was used. Additionally, there was a card on which a rightward-pointing, blue arrow was drawn (to be used as a forward recall cue) and another on which a leftward-pointing, orange arrow was drawn (to be used as a backward recall cue).

Procedure. Each subject received four trial blocks. Within each block there were 12 trials, which included the same set of word list types (and the same number of filler trials) as in Experiment 2. However, in this experiment, each word list was succeeded, in the same presentation rhythm as the list, by a card cue to recall the words in a forward or backward order. In each block, four of the eight test lists and two of the four filler lists were to be recalled in the forward order, and the remaining lists were to be recalled in the backward order. The order of trials was random within a block, so that the subject did not know the direction of recall until the postlist cue was presented.

The recall order for variants of each mixed list were alternated across blocks within each subject. For example, if within Block 1 a subject received the 2S/3L trial with backward recall and the 3S/2L trial with forward recall, then this same assignment of trial types to recall orders would be used on Block 3, whereas the opposite assignment would be used on Blocks 2 and 4. Across subjects, the assignments were completely balanced for each block.

The procedure for the estimation of pronunciation time was identical to that of the previous experiments, but capacity estimates were calculated separately for for-

ward versus backward recall. All other aspects of the present procedure were as in Experiment 2.

Results

Proportion correct. In addition to an overall ANOVA of proportion correct, proportion correct scores were entered into separate ANOVAs for the forward and the backward recall orders. For the sake of exposition, it will be simplest to begin with these separate ANOVAs. They both included the length of words in the first half of the list (short vs. long), the length of words in the second half of the list (short vs. long), and serial position (1–5) as fixed-effect, within-subject factors.

In the forward recall order, the effect of the length of words in the first half of the list was significant, $F(1,31) = 10.18, p < .004, MS_e = 0.06$, as was the effect of serial position, $F(4,124) = 60.00, p < .001, MS_e = 0.06$. As in Experiment 2, the effect of the length of words in the second half of the list did not reach significance.

In strong contrast to the above findings, in the backward recall order it was the length of words in the *second* half of the list that reached significance, $F(1,31) = 4.62, p < .04, MS_e = .07$, in agreement with the output time hypothesis; the effect of the word length in the first half of the list did not approach significance, $F(1,31) = .01, MS_e = .54$. The serial position effect was again significant, $F(4,124) = 50.17, p < .001, MS_e = 0.07$ but, unlike forward recall, the higher performance was found for items presented at the end of the list (and recalled first). The significant effects of the length of words in the first half of the list (in the forward recall trials only) versus the second half of the list (in the backward recall trials only) are shown for each serial position, in Fig. 4.

Within the overall ANOVA, which included the order of recall as a fixed-effect, within-subject factor along with all of the

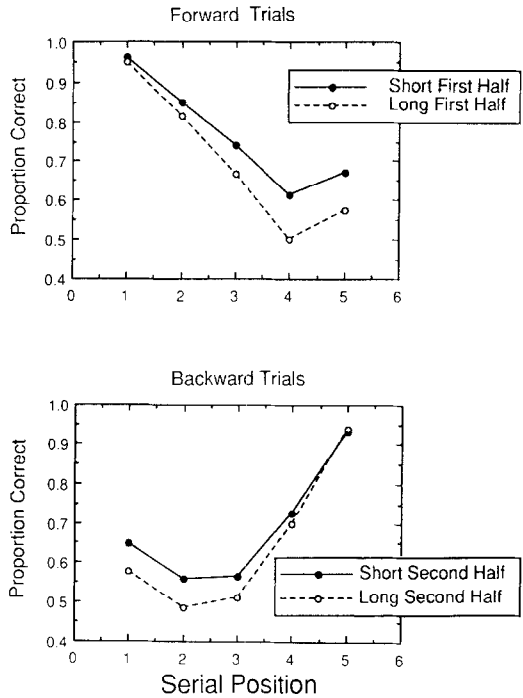


FIG. 4. A depiction of the significant main effects within separate analyses for each recall order in Experiment 3. Top panel: forward recall, proportion correct on lists with a short (solid line) versus long (dashed line) first half; bottom panel: backward recall, proportion correct on lists with a short (solid line) versus long (dashed line) second half.

factors of the separate ANOVAs for each recall order, only effects that involved the order of recall are of interest. First, there was an overall advantage of the forward order (forward, .73; backward, .66), $F(1,31) = 7.82, p < .009, MS_e = 0.20$. In keeping with Fig. 3, there was also a highly significant interaction of Recall Order \times Serial Position, $F(4,124) = 85.84, p < .001, MS_e = 0.06$.

In the overall ANOVA there was also a three-way interaction of Recall Order \times Serial Position \times First-Half Word Length, $F(4,124) = 3.54, p < .01, MS_e = 0.03$. This effect, which is shown in Table 2, illustrates that a consistent effect of the first-half word length across all serial positions emerged only in the forward recall order. Finally, instead of a comparable interaction for sec-

TABLE 2
MEAN PROPORTION CORRECT FOR EACH RECALL
ORDER, SERIAL POSITION, AND FIRST-HALF-LIST
WORD LENGTH IN EXPERIMENT 3

List/recall conditions	Serial position				
	1	2	3	4	5
Forward recall order					
First half-list short	.961	.848	.742	.613	.672
First half-list long	.949	.816	.668	.500	.574
Backward recall order					
First half-list short	.621	.555	.539	.684	.930
First half-list long	.650	.488	.539	.742	.937

ond-half word length, the main effect of second-half word length was significant across recall orders (short, .720; long, .678), $F(1,31) = 7.39, p < .02, MS_e = 0.08$. One possible explanation of the more pervasive effect of second-half word length than of first-half word length across recall orders is that some subjects somehow tried to prepare for the more difficult, backward recall order even before the recall order cue was received, knowing that they had to recall backward on half of the trials.

In Fig. 4, one can see that the advantage for the forward recall order can be attributed to some of the medial serial positions. Because the most important influence on the recall of any particular item was clearly its output position, we assessed the additional effects of recall order by conducting separate F tests on the forward versus backward orders of recall for each output position (i.e., Serial Position 1 in the forward order was tested against Serial Position 5 in the backward order, because those data reflect Output Position 1 for the two recall orders; Position 2—Forward was tested against Position 4—Backward; and so on). The advantage of forward recall occurred for the second output position, $F(1,31) = 12.57, p < .002, MS_e = 0.29$, and the third output position, $F(1,31) = 24.54, p < .001, MS_e = 0.29$, but did not approach significance for the other output positions. An intuitive description of these results is that, in the forward order, subjects could rather often report the first two or three

items; but, in the backward order, only the first one or two items were often reported correctly. This is understandable given the need to mentally reverse the list items in the backward order and the added response difficulty that this would appear to engender.

In sum, the largest, most robust effect on recall was of the length of words in which-ever half of the list was output first, namely the first half of the list in forward recall and the second half of the list in backward recall. Moreover, across recall orders it was the second half of the list in which the stronger and more general effect of word length was found in this experiment. All of this is the opposite of what would have been expected according to a rehearsal mechanism favoring the beginning of the list.

Memory capacity times. The mean capacity time estimates are shown in Table 3. The estimates in forward recall are consistent with the decay hypothesis, ranging from 1.47 to 1.62 s. They do appear to be somewhat lower than in Experiments 1 and 2, which suggests that the requirement of bidirectional recall impaired the functional memory capacity within the present experiment. Memory capacity times in the backward recall order, ranging from 1.31 to 1.51 s, were significantly lower than in forward

TABLE 3
MEASURES OF MEMORY CAPACITY TIME FOR EACH
CONDITION IN EXPERIMENT 3

List/recall condition	Speaking estimate	Reading estimate
Forward recall		
Short-short	1.49 (0.39)	1.57 (0.47)
Short-long	1.62 (0.31)	1.60 (0.32)
Long-short	1.56 (0.40)	1.55 (0.44)
Long-long	1.55 (0.43)	1.47 (0.43)
Backward recall		
Short-short	1.35 (0.34)	1.40 (0.34)
Short-long	1.33 (0.44)	1.31 (0.42)
Long-short	1.43 (0.43)	1.41 (0.41)
Long-long	1.51 (0.40)	1.43 (0.41)

Note. Standard deviations in parentheses.

TABLE 4
CORRELATIONS BETWEEN PRONUNCIATION TIME MEASURES AND MEAN PERFORMANCE IN FORWARD AND BACKWARD RECALL

Recall order	Pronunciation time measure			
	Spkg-short	Spkg-long	Rdng-short	Rdng-long
Forward	.42*	.62**	.30	.32
Backward	.47**	.49**	.57**	.50**

Note. Spkg = speaking measure of pronunciation time; rdng = reading measure. "Short" and "long" refer to the words spoken or read. Recall order means entering into these correlations were averages across all word lengths and serial positions.

* $p < .05$, two-tailed.
** $p < .01$, two-tailed.

recall, $F(1,31) = 7.33, p < .02, MS_e = 0.41$.² This further suggests that factors other than fixed decay and pronunciation rates must affect immediate memory. For example, forward speech measures may well overestimate the rate of speech that is actually used in backward recall responses.

One might further question if speaking rate plays the same role in backward recall that it does in forward recall, given the potential for strategies unique to backward recall. We examined this question by calculating correlations between forward and backward recall and the pronunciation time measures, which are shown in Table 4. The table shows that these correlations are at least as high in backward recall as in forward recall, suggesting that speaking rate is

² This effect occurred within an ANOVA of the memory capacity estimates with recall order, first-half word length, second-half word length, and speed estimate (speaking vs. reading) as fixed-effect, within-subject factors. Additionally, several interaction effects conformed to the same pattern as in Experiment 2 (First-half Word Length \times Estimate, $F(1,31) = 16.83, p < .001, MS_e = 0.01$; Second-half Word Length \times Estimate, $F(1,31) = 19.18, p < .001, MS_e = 0.01$; and First-half Word Length \times Second-half Word Length \times Estimate, $F(1,31) = 5.64, p < .03, MS_e = 0.0005$), again suggesting that subjects may have had difficulty rapidly repeating triads of the longer words. Finally, there was a three-way interaction of Recall Order \times Second-half Word Length \times Estimate, $F(1,31) = 4.77, p < .04, MS_e = 0.0006$, but the correct interpretation of that effect is uncertain.

relevant to both types of recall despite the fact that the pronunciation measures that entered into these correlations always involved a form of forward recital rather than backward.

Memory decay function. In Fig. 5, the memory decay function was plotted in the same manner as in Experiment 2, for forward recall (top panel) and backward recall (bottom panel). For backward recall it was of course the items presented later in the list than the plotted item, but recalled earlier in the response, that determined the estimate of response delay. Although performance appeared to decline as a function of response delay more quickly here than in Experiment 2 (cf. Fig. 3 & 5), the patterns of performance were quite similar. For the forward condition, a temporal delay model

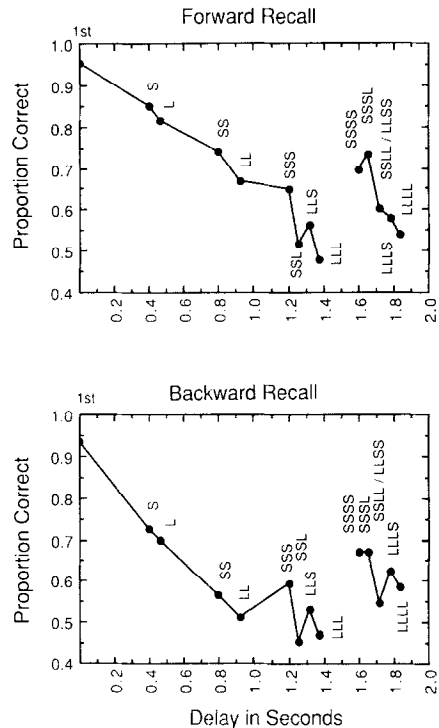


FIG. 5. For Experiment 3, the mean proportion correct recall as a function of the estimated total time of pronunciation of items recalled before the item in question. Data labels indicate the sequence of short (S) and long (L) words occurring before the item in question. Top panel: forward recall; bottom panel: backward recall.

that was linear except for a constant adjustment for the last output position was again quite good, $R^2 = .96$, and was again better than when pronunciation time information was omitted, $R^2 = .81$. The difference between models was confirmed in a hierarchical regression analysis, Partial $R^2 = .14$, $F(1,11) = 34.89$, $p < .001$.

For the backward condition (bottom panel of Fig. 5) a roughly similar pattern of results was obtained. Notice that in this condition it was the first input position (i.e., still the item output last) for which there was a recall advantage above what would be expected on the basis of the response delay alone. The data for the backward recall condition do appear somewhat noisier than the forward condition, and therefore the delay model yielded a good but less impressive fit, $R^2 = .83$. The delay model appeared to account for, at best, slightly more of the variance than the contrasting model in which any differential temporal information was omitted ($R^2 = .77$), Partial $R^2 = .05$, $F(1,11) = 3.37$, $p < .1$. It is worth noting that the reduced effect of delay within output positions in this regression analysis in no way contradicts the strong effect of second-half word length that was obtained in the ANOVA of the proportion correct for the backward recall condition (see above). That ANOVA clearly indicated that a delay-based factor operates in backward recall. The relevant difference between analyses probably is that, in the regression, unlike the prior ANOVA, forward pronunciation times had to be used (to estimate the speech rate in backward recall). Understandably, the imprecision of that method limited the precision of the output delay model in the regression.

Discussion

The pattern of performance observed in the present experiment clearly indicates that there was an effect of the output delay, which was modulated by the length of whatever words were to be output first.

This was demonstrated clearly in the separate analyses of the forward and backward recall orders. The length of words in the first half of the list yielded a significant effect in the forward order only, whereas the length of words in the second half of the list yielded a significant effect in the backward order as well as the overall analysis across orders. This contradicts what one would have expected on the basis of a rehearsal process at input, which could only favor the words presented earlier in the list.

The analyses of pronunciation speeds reassure us that the articulatory processes that occur on backward trials are related to recall in a manner similar to articulatory processes in forward recall. In fact, even though the estimation of pronunciation speed was a forward pronunciation measure, this measure was more highly correlated to recall in the backward condition than in the forward condition.

Finally, the estimated functions of phonological memory decay during the output delay (Fig. 5) reaffirm the importance of output delays in verbal recall. In general, recall was a decreasing function of output delay. However, in both recall orders there was an added advantage for the item that was output last (i.e., the last input serial position in forward recall and the first input position in backward recall). This advantage may reflect some contribution to recall outside of the articulatory loop. It plausibly may reflect greater temporal distinctiveness for items at both ends of a list (e.g., see Lee & Estes, 1981). Ceiling effects would have obscured the contribution of distinctiveness at the position output first.

GENERAL DISCUSSION

The outcome of the present research indicates quite clearly that word length effects in immediate verbal memory for visually presented word lists must be accounted for largely by the effect of output delays, rather than primarily by covert articulatory processes carried out during the input of

the list. This conclusion is fully consistent with the operation of the articulatory loop as described by Baddeley (1986), but that description did not include a strong commitment to which of several different, theoretically possible articulatory mechanisms actually produce the word length effect in experiments using spoken serial recall.

Experiment 1 provided a straightforward but important replication of the finding of Baddeley et al. (1975, Experiments 4 & 5) that the word length effect could be obtained with short and long word sets matched for phonemic and syllabic content. The potentially confounding factor of phonological similarity between items in a set, evident upon close inspection of the stimuli Baddeley et al. used, turned out not to be critical for the effect. Thus, pronunciation time per se, rather than a form of phonological interference, is sufficient to cause a word length effect.

Previous research has not localized the effects of word length. Any word's length theoretically might affect the recall of that word alone, or it might affect the recall of other words on the list. Experiment 2 showed that the length of words in the first half of the list was more critical than the length of words in the second half of the list, with the first-half length effect extending across all serial positions. This result is consistent with what one would expect on the basis of a mechanism in which later words can decay in memory as earlier words are pronounced during the recall period. The findings of this experiment alone would be consistent also with alternative explanations based on rehearsal in the input phase, but Experiment 3 rules out those accounts (see below). In Experiment 3, the separate manipulation of first-half and second-half word length was again used, but this time subjects were to recall the list in reverse order on half of the trials. They did not know the recall order until the list was complete, so any difference between recall orders could only be accounted for by processes occurring in the output phase of the

trial. A profound difference between output orders was obtained. Specifically, an analysis of the forward recall order showed only a first-half length effect, whereas the analysis of the backward recall order showed only a second-half length effect. In fact, the performance functions obtained for the two recall orders appear as near mirror images (see Fig. 3), except that output positions 2 and 3 show an advantage for the forward order. In both recall orders, the word length effect was obtained for the half of the list that was output first. Thus, the results support an account in which the length of words output first influences the recall of words output subsequently.

One could argue that the manipulation of output order in Experiment 3 leads to a departure from strategies that are found in the ordinary task of immediate, serial recall. However, the results of a pronunciation speed test indicated that the same relation between pronunciation speed and memory occurred in the backward conditions of Experiment 3 as in the forward condition in all three experiments. All conditions replicated the finding of Baddeley et al. (1975) that subjects can remember as many words as they can pronounce in about 1.5 to 2.0 s.

In explaining the word length effect, a devil's advocate position might be that another aspect of the words, such as difficulty on some unspecified scale, could have covaried with the length of the words. However, this account seems odd in light of the finding that performance on words within the half of the list output last was significantly affected by the length of words output earlier, but not by their own length. It would appear unlikely that performance on any word would be influenced more by another word's difficulty than by its own difficulty. A difficulty-based account also would provide no explanation for the finding that articulatory suppression eliminates the word length effect with visual materials (Baddeley et al., 1975).

One of the strongest sources of evidence in favor of the output delay account is the

orderly effect of the estimated output delay on recall performance across output positions (see Fig. 3 and 5). This relation is important, because it estimates the function of phonological memory decay more directly than in past research. The previous finding was simply that memory performance was a linear function of speaking rate. The explanation of that linear relation (e.g., Baddeley, 1986; Baddeley et al., 1975) was that a variable articulatory process operates on a relatively fixed memory decay function, but there was no proof that this parsimonious account was correct. The present observation of decay across about 2 s during the response period nicely confirms the theoretical assumptions that Baddeley made.

On the other hand, the decay functions also illustrate the limitations of the articulatory loop hypothesis. Performance in the serial position output last (i.e., the last input position in forward recall and the first input position in backward recall) was better than expected on the basis of phonological memory decay. One likely explanation is that there is an added distinctiveness of items at either end of the list. The advantage would not be observable for the item output first, because performance on that item was already at ceiling level. The results suggest that memory decay may never account for *all* of the variance in a serial recall procedure.

An additional point of interest is that performance was considerably better in Experiment 2 (forward recall only) than in the forward recall condition of Experiment 3. The only procedural difference is that, in Experiment 3, subjects did not know the direction of recall until after the words were presented. This suggests that the mnemonic processing that occurs during list input depends in part on the expected type of output. Thus, although the present results indicate that there was an effect of delay during verbal output, they are consistent also with the possibility that there could be an additional effect of the duration of covert

rehearsals during the list presentation, as some previous research (Baddeley, Lewis, & Vallar, 1984) appears to suggest.

In sum, the results provide support and clarification for a model (Baddeley, 1986) that includes a transient phonological memory store that is subject to decay. The length of words to be repeated first helps to determine how much time will elapse before the remaining words can be repeated and, therefore, how much phonological memory decay will occur.

The present method is useful because it permits an analysis of the immediate recall performance into multiple memory components. A decaying phonological store that was merely inferred in past research can be more directly observed in the present study (see Fig. 3 & 5). Additionally, the figures illustrate that there is a separate distinctiveness component for the last-output item that was independent of the phonological decay function. Thus, the data allow a clearer, more constrained view of Baddeley's (1986) articulatory loop.

REFERENCES

- BADDELEY, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- BADDELEY, A. D., & HULL, A. (1979). Prefix and suffix effects: Do they have a common basis? *Journal of Verbal Learning & Verbal Behavior*, 18, 129-140.
- BADDELEY, A. D., LEWIS, V. J., & VALLAR, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, 36, 233-252.
- BADDELEY, A. D., THOMSON, N., & BUCHANAN, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14, 575-589.
- CASE, R., KURLAND, D. M., & GOLDBERG, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, 33, 386-404.
- CONRAD, R. (1964). Acoustic confusion in immediate memory. *British Journal of Psychology*, 55, 75-84.
- HULME, C., THOMSON, N., MUIR, C., & LAWRENCE, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology*, 38, 241-253.
- HULME, C., & TORDOFF, V. (1989). Working memory development: The effects of speech rate, word

- length, and acoustic similarity on serial recall. *Journal of Experimental Child Psychology*, 47, 72-87.
- LANDAUER, T. K. (1962). Rate of implicit speech. *Perceptual and Motor Skills*, 15, 646.
- LEE, C. L., & ESTES, W. K. (1981). Item and order information in short-term memory: Evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 149-169.
- 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.
- NICOLSON, R. (1981). The relationship between memory span and processing speed. In M. Friedman, J. P. Das, & N. O'Connor (Eds.), *Intelligence and learning* (pp. 179-184). New York: Plenum Press.
- PALMER, S. E., & ORNSTEIN, P. A. (1971). The role of rehearsal strategy in serial-probed-recall. *Journal of Experimental Psychology*, 88, 60-66.
- SCHWEICKERT, R., & BORUFF, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 419-425.
- SCHWEICKERT, R., GUENTERT, L., & HERSBERGER, L. (1990). Phonological similarity, pronunciation rate, and memory span. *Psychological Science*, 1, 74-77.
- STANDING, L., BOND, B., SMITH, P., & ISELY, C. (1980). Is the immediate memory span determined by subvocalization rate? *British Journal of Psychology*, 71, 525-539.
- WATKINS, M. J. (1977). The intricacy of memory span. *Memory and Cognition*, 5, 529-534.
- WRIGHT, C. E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory and Cognition*, 7, 411-419.
- ZHANG, G., & SIMON, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and the acoustical loop hypothesis. *Memory and Cognition*, 13, 193-201.

(Received November 14, 1990)

(Revision received March 10, 1991)