

Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition

Alan Baddeley, Dino Chincotta, Lorenzo Stafford, and David Turk

University of Bristol, Bristol, UK

Although it is generally accepted that the word length effect in short-term memory operates through output delay or interference, there is less agreement on whether it also influences performance through its impact on rehearsal. We investigated this issue by studying the effect of word length on recall and on a recognition task in which output delay was controlled. Word sequences were repeated exactly, or with one pair of words reversed. Two experiments using auditory presentation showed clear word length effects for both recall and serial recognition, although the magnitude of the effect tended to be less for recognition. A third experiment using visual presentation studied the effect of articulatory suppression during the recognition test; again we found a clear word length effect. It is concluded that the word length effect can influence retention through both rehearsal and output factors, as proposed by the phonological loop hypothesis.

Baddeley, Thomson, and Buchanan (1975) observed that immediate serial recall for word sequences decreased systematically with the number of syllables within the words comprising the sequences to be recalled. As recall required the writing of only the first three letters of each word, they discounted output delay as an explanation, attributing the word length effect to trace decay during the real-time rehearsal of the words. Although the word length effect has been replicated many times (Hulme, Thomson, Muir, & Lawrence, 1984; Mackworth, 1963; Neath & Nairne, 1995; Schweickert & Boruff, 1986), at least two aspects of the interpretation have been challenged.

The first of these concerns the assumption of trace decay, rather than some form of interference. The preference for a decay interpretation in the study by Baddeley et al. (1975) came from the observation, in an experiment involving two sets of disyllabic words, that the set of items that took longer to articulate showed poorer recall than did more rapidly articulated disyllables. It has, however, been claimed that the disyllabic word set used by Baddeley et al. is atypical, with other word sets supporting an interpretation in terms of complexity rather than duration (Caplan & Waters, 1994; Lovatt, Avons, & Masterson, 2000). Unfortunately the

Requests for reprints should be sent to Alan Baddeley, Centre for the Study of Memory and Learning, Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Bristol, BS8 1TN, UK. Email: Alan.Baddeley@bristol.ac.uk

available pool of suitable words tends to be limited, the words potentially idiosyncratic, and the effects rather small. Although complexity is almost certainly an important determinant of memory span for nonwords (Hulme, Maughan, & Brown, 1991; Service, 1998), this may simply reflect the fact that subjects are being required to store more novel information with the more complex pseudowords. The fact that subjects have been familiarized with the novel pseudowords does not mean that the pseudowords will be retrieved as readily as words that are already in the subject's lexicon. Performance may well be strongly influenced in nonword recall by the need for storage of item information, a factor that is typically minimal in the characteristic word recall procedure when a limited set of familiar words is used. Hence, although these nonword studies are of considerable interest, their implications for interpreting the findings of Baddeley et al. (1975) remain equivocal.

Cowan and colleagues have tried to avoid the problem of item comparability by using the same set of words for long and short conditions but instructing subjects either to prolong or to shorten their pronunciation of the words (Cowan, Nugent, Elliott, & Geer, 2000; Cowan, Wood, Nugent, & Treisman, 1997, Experiment 2). They obtained clear evidence that prolonged pronunciation impaired recall. However, Service (1998, 2000) has suggested that the magnitude of the effect is less than that predicted by decay theory and may reflect the demands of the somewhat artificial task of articulating slowly. (See Cowan, Nugent, & Elliott, 2000; Service, 2000, for further discussion.) It is clear, however, that the nature of the word length effect, with its implications for the classic issue of whether forgetting in short-term serial recall reflects trace decay or interference, is still to be resolved.

The second point of issue concerns the question of whether the principal cause of impairment resulting from word length is based on the slower rehearsal of long words, or from the increased time it takes to recall and articulate the longer words during the process of recall. Such output delay is certainly likely to be the case with spoken recall. Implicit in the Baddeley et al. (1975) interpretation was the assumption that rehearsal and recall were basically similar processes, involving the ordered retrieval of the constituent items. In the case of rehearsal, this was typically assumed to be covert and to potentially be interleaved with item presentation, being continued during an unfilled delay, but disrupted by articulatory suppression, the requirement to utter some irrelevant speech sound (Murray, 1968). The importance of forgetting during the process of recall was demonstrated by Cowan et al. (1992) in a series of experiments in which delay was manipulated by varying the position of long and short words in the sequence. They show that locating long words early in the sequence leads to a longer average delay in recall and to poorer overall performance. Such a result is consistent with the conclusions of an earlier study by Baddeley and Hull (1979) in which they varied the length of redundant response prefix or stimulus suffix words following a digit sequence. In both cases, short interpolated words led to better recall than did long words.

The question therefore arises as to whether the word length effect is attributable entirely to output interference, or whether rehearsal also plays a role. Although it could be argued that Baddeley et al.'s (1975) procedure of only requiring the subjects to write the initial three letters of each response equates output, it is, as Longoni, Richardson, and Aiello (1993) suggest, entirely possible that subjects subvocalize the whole word before writing the first three letters, hence producing a subvocal output effect.

Henry (1991) used a probe technique whereby the list presentation was followed by the presentation of a single item, with the subject required to respond by producing the item that

had followed the probe. She found that the word length effect was substantially reduced by this procedure, but was still present for all but her youngest subjects. Avons, Wright, and Pammer (1994), using adult subjects, also noted that the probe technique reduced the size of the word length effect substantially, but they still noted a significant effect. These studies are consistent in emphasizing the importance of output factors, while still suggesting some contribution from other variables such as rehearsal. One might argue, however, that the probe item itself acts as a suffix; given the demonstration by Baddeley and Hull (1979) that long suffixes disrupt recall more than do short suffixes, it is possible to argue that even the small effect of word length in probe studies may be attributable to output factors rather than to rehearsal. Further evidence for the importance of output is provided by Coltheart and Langdon (1998), who used very rapid visual presentation of the stimulus words, hence allowing insufficient time to rehearse during input. They obtained both acoustic similarity and word length effects.

All the studies that we have reviewed so far have emphasized the importance of the process of recall *per se* on the impact of word length, but have not denied the possibility that there is also a role of word length in rehearsal. However, the possibility that the word length is entirely attributable to forgetting during output is suggested by Doshier and Ma (1998). They report a study in which "equal output times accounted precisely and completely for the measured memory span" (Doshier & Ma, p. 316), concluding that "temporal limits during recall offer an alternative, and possibly superior explanation of span performance" (Doshier & Ma, p. 328). Their study involved immediate serial recall of sequences sampled from sets of digits, consonants, or short words. Each set comprised 10 items, and subjects were given sufficient prior practice to ensure that all three sets of material resulted in equivalent performance when tested using the classic Sternberg (1966) probed recognition procedure. Over many trials, they studied immediate serial recall, together with recall output time, observing that span approximates to the number of items that can be recalled in about 4 to 6 s—considerably slower than the rate for speeded rehearsal, which typically approximates about 2 s (Baddeley et al., 1975; Schweickert & Boruff, 1986). The basis of Doshier and Ma's suggestion that the word length effect is purely an output phenomenon is essentially correlational, reflecting the observation that correct trials are associated with shorter output times. Doshier and Ma do, however, accept the possibility that the direction of causality that they propose is open to challenge, but they opt for an interpretation of their results purely in terms of output delay on the grounds of greater simplicity.

Despite the attractions of the simplicity of assuming only one factor, however, there is evidence to suggest that articulation time itself is likely to be determined by at least two factors, one involving retrieval of the spoken form and the other the time taken to articulate it. Cowan and colleagues (Cowan, 1992; Cowan et al., 1994, 1998) suggest that at least two factors influence articulation time, one being indexed by the time taken to produce the word itself, and the other reflecting the inter-word pause duration during spoken recall. In a subsequent study, Jarrold, Hewes, and Baddeley (2000) extended this to the study of pauses during the speeded articulation of sub-span sequences. They found that spoken duration of words and the length of the pauses between them predicted separate and non-overlapping variance in performance, arguing that the impact on recall of spoken duration came through time-based or item-based forgetting, in contrast to pause duration, which reflected individual differences in phonological retrieval or speech planning.

In the Doshier and Ma (1998) study, one might therefore expect three factors to contribute to the time to retrieve and output a remembered sequence. The first of these would be the time to articulate each given item, the variable that is assumed to play the principal role in standard word length experiments. The second would be the time to retrieve the correct item from memory, where memory trace strength might reasonably be assumed to be associated with speed of retrieval. The third factor would be the time to access the relevant phonological form necessary for producing the verbal response. The fact that digits, letters, and words had been equated by practice for performance on the Sternberg recognition task (Sternberg, 1966) does not of course mean that the three are likely to be equated in time to recall the phonological representation. Hence, although Doshier and Ma observed a high degree of correlation between recall time and recall accuracy, the essentially correlational nature of their results suggests a need for further investigation before ruling out the possibility that one component of the word length effect operates through rehearsal rather than output. The experiments that follow attempt to obtain more direct evidence of the potential importance of rehearsal, by using a paced serial recognition procedure in order to equate the time between presentation and test across different word lengths. Before going on to describe this approach, however, we should clarify our interpretation of the rehearsal process, contrasting it with an alternative approach to explaining the word length effect.

As mentioned earlier, we assume that rehearsal is a process of maintaining information over a delay, and that in the case of the phonological loop, it involves a form of subvocal articulation, a process that is fundamentally equivalent to that used in spoken verbal output. In the initial Baddeley and Hitch (1974) working-memory model, and in other computationally more developed and explicit models (Burgess & Hitch, 1992, 1999; Henson, 1998; Page & Norris, 1998), the effect of word length operates through time-dependent trace decay. The trace decay assumption is not, however, crucial. An interpretation based on interference would give essentially equivalent predictions, provided the degree of interference was assumed to be directly related to word length or associated complexity (Service, 1998).

An alternative explanation of the word effect has been presented by G.D.P. Brown and Hulme (1995), who assume that long words are more vulnerable to forgetting because they have more components. This view is also adopted by Neath and Nairne (1995), who present an account of word length based on interference theory. Brown and Hulme assume a further process of redintegration, whereby prior linguistic knowledge can be used to compensate for loss of components of longer words. Although these views do not require the assumption of a rehearsal process, the observation of a word length effect on rehearsal would not be problematic, as it is presumably necessary to retrieve the constituent items before they can be rehearsed, with long words presumably being more subject to error than are short words.

The experiments to be described used a serial recognition task in which a sequence of items is presented, followed by either an identical sequence, or one in which the order of two adjacent items has been switched (Allport, 1984; Gathercole, Service, Hitch, Adams, & Martin, 1999). As both learning and test sequences are paced, the delay between presentation and test is likely to be same for long as for short words. Furthermore, recent work by Gathercole, Pickering, Hall, and Peaker (2001) has shown that this method is highly resistant to the effects of prior language status, such that memory for sequences of nonwords is virtually the same as

that for familiar words. This has been interpreted as suggesting that storage depends upon a system that is relatively uninfluenced by lexicality, whereas recall, and presumably articulation, is very language sensitive (Baddeley, in press). This paradigm has the advantage that it is less likely to disturbance by slight differences in the characteristics of the long and short words. More important, by equating presentation and test rate between the two sets of words and obviating any requirement to produce the words overtly, any differences in performance would seem to implicate sub-vocal rehearsal, as proposed by Baddeley et al. (1975).

We therefore designed a pilot experiment, which had two components. The first involved a recall paradigm based on that used by Baddeley et al. (1975); we also included an articulatory suppression condition. This should remove the word length effect; if it did not, then it would suggest that our two word samples may differ on variables other than length, making interpretation of our results problematic. The second condition involved using the same material to test recognition performance, using the procedure described by Gathercole et al. (2001), with spoken word sequences being followed by either an identical sequence or one in which the order of two words was reversed. To allow for possible differences in overall level of performance resulting in ceiling or floor effects, sequences of five, six, seven, and eight items were used.

As expected, the recall condition showed a clear effect of word length, which was removed by articulatory suppression. Recognition performance, however, was less clear cut. As expected, level of performance decreased with list length, but although there was a suggestion of a word length effect for four of the five list lengths, this effect did not reach statistical significance. A possible reason for this was suggested when subjects were debriefed after the experiment. A number of subjects reported rehearsing the initial letters in the recognition task. Although the presence of a small number of repeated initial letters would make this a less than perfect strategy, it would only fail under conditions when a mis-match sequence happened to have been created by switching two adjacent items that both happen to begin with the same initial letter, and even then only half of the time. In the case of recall, of course, the initial letter strategy is likely to be problematic for any sequence in which one of the repeated letters occurs, and in addition it would require the subject to generate the appropriate words from the cue letters during recall, a relatively demanding task that would be likely to increase forgetting. We therefore decided to modify our pilot study using a separate set of material designed to eliminate the first letter strategy in two ways. First of all, we had several common initial letters in each group. Second, we chose words that had initial letters that had names that were phonologically similar to each other, hence making the first letter strategy even less attractive.

EXPERIMENT 1

We constructed pools of 10 short and 10 long words. Each word pool consisted of two pairs of items beginning with the letters *c* and *e*, and two triplets of words beginning with the letters *t* and *s*. Thus, in addition to increasing the degree of overlap in initial letter between items, three of the initial letters were phonologically similar (*c*, *e*, and *d*).

Method

Subjects

The subjects were 16 undergraduate students from the University of Bristol and were paid or received course credits for participation. In this and subsequent experiments, all subjects were native or fluent English speakers.

Materials

Two pools of 10 items, one consisting of single-syllable words, the other comprising three-syllable words, were constructed. The word pools were matched for age of acquisition, imagability, frequency, familiarity, and visual complexity using the Morrison, Chappell, and Ellis (1997) norms. The short-word pool comprised *crab*, *cheese*, *toad*, *torch*, *tent*, *ear*, *eye*, *ski*, *stool*, and *sun*, and the long-word pool consisted of *caravan*, *celery*, *tomato*, *typewriter*, *telephone*, *elephant*, *envelope*, *motorbike*, *strawberry*, *submarine*, and *screwdriver*. Analysis using a *t* test showed that there was no statistical difference between the short- and long-word pools for any of the dimensions listed.

The items, spoken in a neutral tone by a female, were captured by a DAT recorder and subsequently digitized using speech analysis software. They were then segmented into individual sound files, and a computer program was constructed for the presentation of item sequences in a pre-specified rate and order.

For the measurement of memory span, two sets of short words and two sets of long words were constructed by randomly selecting items from the appropriate pool without replacement as follows: Each set commenced with four sequences of 2 items, followed by four sequences of 3 items, and so on to a maximum of 10 items.

For the measurement of serial recognition, four sets of sequences of short words and four sets of long-word sequences were constructed as follows: For each word pool, 10 sequences of five, six, seven, and eight items were selected at random and without replacement. Half of the sequences at each length were followed by identical sequences, and the remaining half were followed by a different second sequence, achieved by transposing a single adjacent item pair at random, with the constraint that the first and last item in the sequence were never changed.

Procedure

First, the subjects were given a list of the items used in the experiment and were requested to write the first three letters for each word. Subjects were required to respond in this way so that the written output demand would be the same for long and short words, as in the original Baddeley et al. (1975) study. Next, half of the subjects were tested with the recall task first, followed by the recognition task; the order was reversed for the remaining subjects.

For the recall task, the sequences were presented via headphones and monitored simultaneously by the experimenter via a loudspeaker. First, a tone of 1-s duration was presented, followed by the presentation of the sequence at the rate of one item per second. After the presentation of the last item, a tone prompted the subject to recall the sequence in serial order by writing the first three letters of each item on pre-prepared booklets. After recall was completed, the experimenter presented the next sequence and so forth. For the articulatory suppression conditions, the subjects were instructed to articulate the word *the* audibly at the rate of two utterances per second in time to a metronome throughout presentation and recall.

Testing continued until three of the four sequences were incorrectly recalled at a given length. Memory span was operationalized at the longest length in which a minimum of two sequences were recalled correctly with a score of 0.25 added for every correct trial at this sequence length. The item sequences

were blocked by condition and counterbalancing, achieved through the application of a Williams Latin square design.

For the recognition task, the sequence pairs were presented via headphones and monitored simultaneously by the experimenter via a loudspeaker. First, a tone of 1-s duration was presented, followed by the presentation of the first sequence in the pair at the rate of one item per second. Once the first sequence terminated, the second sequence was presented after a 1-s delay. After the presentation of the second sequence a tone prompted the subject to decide whether the sequence pairs were the same or different by writing *S* or *D* on a pre-prepared sheet. The 10 trials in each word length condition were blocked by sequence length, and within each sequence length block the five same and five different lists were presented randomly. All the subjects commenced with the set of five-item sequences and progressed systematically to the eight-item sequences.

Results

The data from the recall task are summarized in Figure 1. They were subjected to a two-way analysis of variance (ANOVA) in which word length (short or long) and condition (control or articulatory suppression) were both within-subject factors. There was a significant effect of word length, $F(1, 15) = 12.75, p < .01$, with a memory span for short words (4.94) greater than that for long words (4.32), and a significant effect of condition, $F(1, 15) = 15.54, p < .01$, with a memory span greater in the control condition (4.98) than under articulatory suppression

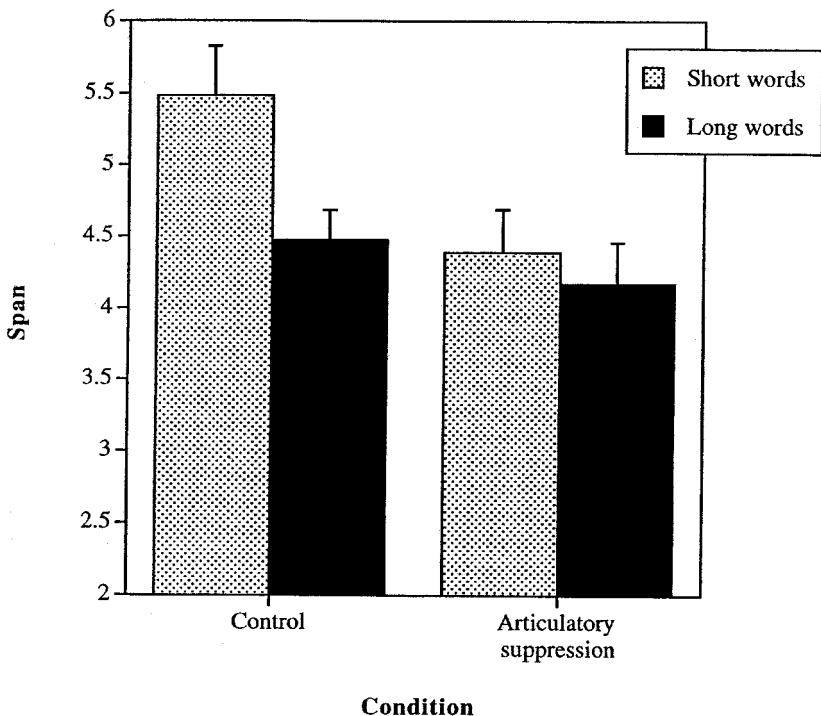


Figure 1. Mean memory span for short and long words under control and articulatory suppression conditions in Experiment 1.

(4.28). The interaction between word length and suppression was also reliable, $F(1, 15) = 8.63$, $p < .05$.

Analysis of the Word Length $\times$ Suppression interaction indicated that the memory span difference between short and long words in the control condition, $F(1, 15) = 17.26$, $p < .001$, $M = 5.48$ and 4.39 , respectively, was abolished under articulatory suppression, $F(1, 15) < 1$, $M = 4.39$ and 4.17 , respectively. Articulatory suppression produced a decrement in memory span for both short words, $t(15) = 4.28$, $p < .0001$, and long words, $t(15) = 2.36$, $p < .05$.

The data from the recognition task were converted into percentage of correct scores¹ and are represented in Figure 2. These data were then subjected to a two-way analysis of variance in which word length (short or long) and list length (5–8) were both within-subjects factors. This showed a reliable main effect of word length, $F(1, 15) = 10.16$, $p < .01$, $M =$ short 73.1%, long 62.0%, and a reliable main effect of list length, $F(1, 15) = 11.55$, $p < .001$. Analysis of these and subsequent recognition data using a guessing correction of hits minus false alarms gave equivalent results.

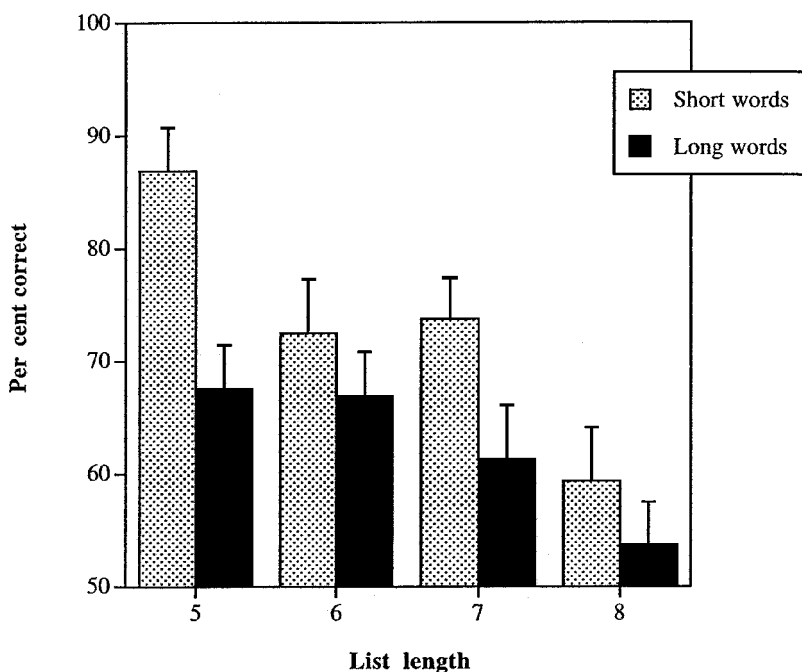


Figure 2. Mean percentage of recognition accuracy for short and long words as a function of sequence length in Experiment 1.

¹We chose percentage correct as the simplest sensible measure of performance. The number and probable distribution of responses made the use of signal detection theory questionable. Non-parametric equivalents such as R (J. Brown, 1974), or A' (Norman, 1964), are under the present circumstances essentially equivalent to using hits minus false alarms, a measure that leaves our results unchanged.

Discussion

The results of Experiment 1 suggested that word length is an important mediator in both recall and recognition tasks when the selection of items reduces the opportunity for the use of an initial-letter coding strategy. However, the magnitude of the effect was somewhat less for recognition than for recall (19.9% vs. 15.2%). However, it could be argued that this is not a fair comparison, as the employment within the recognition paradigm of lists up to a length of eight meant that subjects were typically dealing with longer sequences and were producing a higher error rate. We therefore decided to attempt to replicate our findings, with list length held constant between recall and recognition. We selected a list length of six as being likely to avoid floor and ceiling effects in both recall and recognition conditions. We introduced a second variable, namely delay, which we hoped would ensure that rehearsal occurred, at least in the delayed recall condition.

EXPERIMENT 2

Method

Subjects

The subjects were 16 undergraduate students from the University of Bristol who were paid or received course credits for participation.

Materials

Two sets of 20 six-item sequences of short words were selected at random and without replacement from the pools described in Experiment 1, one set for testing recall and one for recognition. A further two sets of 20 six-item long-words sequences were used to test long-word recall and recognition.

Procedure

The procedure for the measurement of both serial recall and serial recognition was as that described for Experiment 1 with the following exceptions. In Experiment 2 all of the recall and recognition sequences consisted of six items. In the case of the recall task, a tone was presented 1-s after the presentation of the last item in the sequence in half of the trials, and in the remaining half of the trials the tone was presented 5 s after the sequence terminated. In both cases the tone prompted the subject to recall the sequence in serial order. For the recognition task, the delay between the presentation of the first sequence and presentation of the second was 1 s for half of the trials and 5 s for the remaining half of the trials.

The experimenter monitored performance closely throughout the session to ensure that subjects did not pre-empt the tone in the 5-s delay condition. On the negligible number of occasions in which subjects recalled the sequence prior to the presentation of the prompt, the data for that trial were discarded. In both the recall and the recognition tasks the trials were blocked by word length, and the order of testing the 1-s and 5-s delay trials was randomized.

Results

The data from the recall task were subjected to a corresponding ANOVA and are summarized in Figure 3. This showed reliable main effects of word length, $F(1, 15) = 77.37, p < .0001, M =$ short 65.6%, long 51.2%, and condition, $F(1, 15) = 7.09, p < .05, M =$ immediate 60.6%,

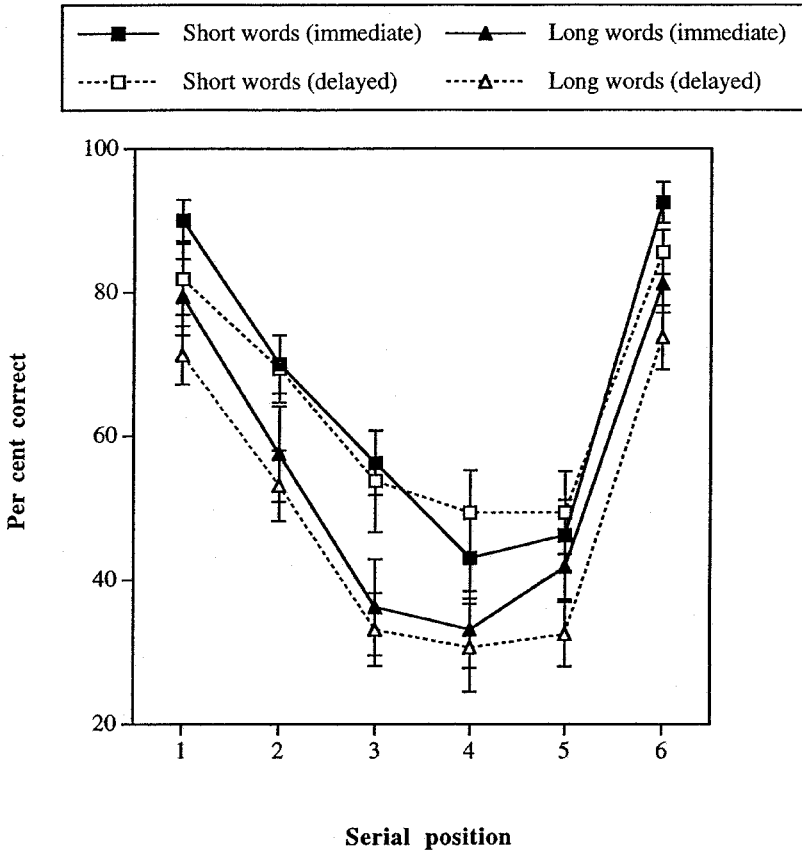


Figure 3. Mean percentage of correct immediate (1-s) and delayed (5-s) serial recall for short and long words in Experiment 2.

delay 57.0%. The main effect of serial position was also reliable, $F(1, 15) = 43.33, p < .0001$, but did not interact with either word length or condition, $F(5, 75) < 1$.

The data from the recognition task were subjected to a corresponding analysis in which word length (short or long) and condition (immediate or delayed) were both within-subjects factors and are summarized in Figure 4. The main effect of word length was reliable, $F(1, 15) = 6.09, p < .05$, with a greater recognition accuracy for short words (71.1%) than for long words (64.5%); the main effect of condition was marginally reliable, $F(1, 15) = 3.88, p = .068$, with a higher degree of accuracy for the immediate condition (70.2%) than for the delayed condition (65.5%). The interaction between the effects of word length and delay was not satisfactorily significant, $F(1, 15) < 1$.

Discussion

Two findings in Experiment 2 are of particular interest. First, an effect of delay was present in both the recall and the recognition tasks such that serial recall and recognition were poorer

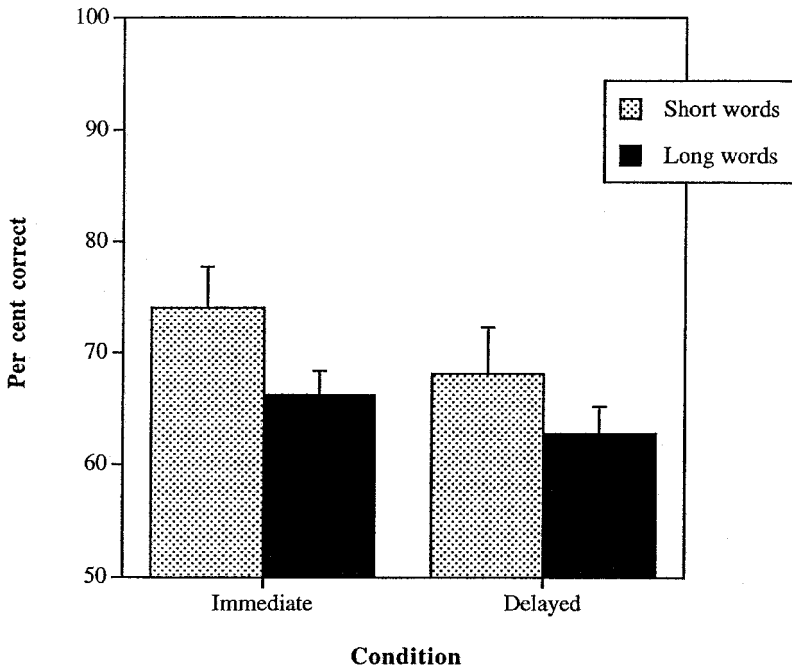


Figure 4. Mean percentage of correct immediate (1-s) and delayed (5-s) serial recognition for short and long words in Experiment 2.

under circumstances in which the time between the end of presentation and commencement of recall and the length of the inter-sequence delay, respectively, was greater. The second finding was that an effect of word length was again noted in both the recall and the recognition task. Once again, the magnitude of the effect is greater for recall than for recognition (21.1% vs. 9.3%). There was, however, no interaction between delay and word length.

The presence of a word length for both recall and recognition is consistent with our initial hypothesis that word length may influence the efficiency of rehearsal. The fact that delay impairs performance suggests that such rehearsal is not entirely successful in avoiding forgetting. One might at first sight, however, expect an interaction, such that long words suffered more from delay than did short words. However, if one assumes that rehearsal involves a time-limited chunk of words lasting, for example, something like 2 s, then this package could well be determined at the start of the retention interval and maintained across the delay. In short, the penalty for holding additional words will already have been paid by the start of the rehearsal interval. Such an interpretation is of course purely speculative and would need to be replicated and explored much more carefully before being accepted.

We have therefore demonstrated that word length influences immediate memory as tested by the serial recognition procedure, an effect that we have successfully replicated. As the two sets of words are matched on age of acquisition, frequency, familiarity, imagability, and visual complexity, and as the length effect is eliminated by articulatory suppression, it seems unlikely that our findings are attributable to differences in the lexical status of the two sets of words. It

would therefore appear to be the case that the word length effect influences ongoing rehearsal as well as output effects.

There is, however, one possible alternative interpretation. It is conceivable that subjects may have continued to rehearse the items during the process of output, even though generation of the words is not necessary to perform the task. We therefore designed a third experiment in which subjects were required to suppress articulation throughout the recognition period, hence eliminating the capacity for continued rehearsal. If the word length effect remains, it would further strengthen the case for attributing at least part of the word length effect to ongoing rehearsal. Finally, to increase the generality of our results and eliminate the possibility that they might reflect some form of auditory memory, rather than the phonological loop, we chose visual presentation for our final experiment.

EXPERIMENT 3

Method

Subjects

The subjects were 16 undergraduate students from the University of Bristol who received course credits for participation.

Design

This experiment utilized a repeated measures design with two experimental conditions: word length (short or long words), and condition (control or output interference).

Materials

Four sets of 20 five-item sequences of short words were generated from the pool of 10 short words used in Experiment 1. Each word appeared 10 times in total in each 20 five-item sequence, twice in each serial position.

Procedure

For the control condition, each word list was presented twice. In the initial presentation the words appeared in capital letters in Times font, size 100. In the second presentation the words appeared in lower case Times font, size 120. This difference was assumed to reduce any influence of visual coding. Each word was presented in the centre of the computer screen for 1 s with an inter-stimulus interval of 500 ms. After presentation of the entire list, a 5-s maintenance interval ensued before presentation of the second list. On half the experimental trials, List 2 contained the same words in the same order ("same" trials), and for the remaining trials the same words appeared in the second list, but the serial order of the words was altered by transposing two of the items ("different" trials). At the end of the presentation of the second list, participants were required to indicate whether the two lists were the same or different. In the output interference condition, the lists were presented in the same way, except that a visual cue 3.25-s after initial list presentation instructed participants to begin articulatory suppression by repeating the word "the" at the rate of two per second. Suppression continued throughout the presentation of the recognition test list. The experiment was undertaken as a single randomized block of trials.

Results

Recognition accuracy was calculated by summing all correct responses for each of the experimental variables. These data were subjected to a two-way ANOVA in which word length (short or long) and condition (control or output interference) were both within-subject factors; results are summarized in Figure 5.

There was a significant main effect of word length, $F(1, 15) = 39.82, p < .001$, with short words being more accurately recognized than long words. We also found a significant main effect of condition, $F(1, 15) = 11.10, p < .005$, with performance impaired in the output interference relative to the control condition. However, there was no significant word length by condition interaction, $F(1, 16) < 1$.

Discussion

Experiment 3 once again showed a clear effect of word length on verbal recognition memory. The fact that we obtained the result using visual presentation extends the generality of our earlier findings, indicating that they are not dependent upon the greater durability of auditory over visual short-term memory. The resistance of the word length effect to articulatory suppression during the test phase rules out an interpretation in terms of subvocal recall at output, as suppression should remove this option. Suppression does indeed impair performance, but the effect is equivalent for sequences of long and short words. Such a result is entirely consistent with an interpretation of the word length effect in terms of its impact on subvocal

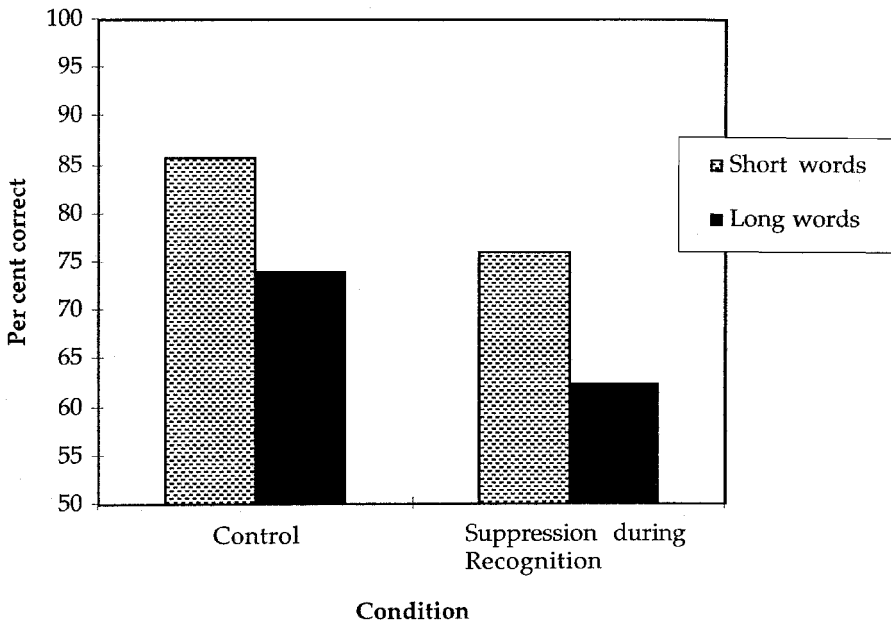


Figure 5. Mean percentage of recognition accuracy for short and long words: the effect of articulatory suppression during the recognition test.

rehearsal; once such rehearsal is prevented, performance will decline, but not differentially as the factors that lead to forgetting are assumed to be no greater for long than for short words.

GENERAL DISCUSSION

Although our three experiments were targeted at investigating the role of rehearsal, they do have broader implications for the explanation of the word length effect and hence for theories of verbal short-term memory. As discussed earlier, interpretations of the word length effect fall into two broad categories. The initial interpretation offered within the Baddeley and Hitch (1974) working-memory framework attributes the effect of word length to greater forgetting during rehearsal and/or recall, with the forgetting resulting from either trace decay (Baddeley et al., 1975; Cowan, Nugent, & Elliott, 2000; Cowan, Nugent, Elliott, & Geer, 2000), or some form of interference (Service, 1998, 2000). The second approach argues that long words are rather like long lists in that they have more components and are thus more vulnerable to forgetting (G.D.A. Brown & Hulme, 1995; Neath & Nairne, 1995).

Our results suggest that the following phenomena require explanation within these theoretical frameworks. First of all, it is clear that the word length effect is found with both recall and recognition. Second, our work and that of Gathercole et al. (2001) suggest that the effect is substantially larger when tested by recall. The final potentially theoretically important feature of our data stems from the observation that suppression during recognition impairs performance to the same extent for long and short words.

The phonological loop hypothesis offers a simple and straightforward interpretation of this pattern of data. Longer words are assumed to take more time to articulate, hence allowing a greater degree of forgetting, either from trace decay or from interference. Such articulation is assumed to be involved in both subvocal rehearsal and overt recall. The recognition paradigm will reflect the occurrence of such differential forgetting during rehearsal, but not during the recognition process, which our own data and that of Gathercole et al. (2001) suggest does not involve subvocalization. The extent of the word length effect in recognition will thus be less than that observed in recall, which presumably reflects both subvocal rehearsal and forgetting during output. The requirement to suppress articulation during output leads to a significant decline in performance, possibly because the signal to begin subvocalization allows less time for rehearsal, or possibly because the simple requirement to perform two tasks simultaneously impairs performance. However, this does not differentially impair long-word retention, as the requirement for articulatory suppression that impairs performance also removes the factor that mediates the word length effect, namely differences in articulation time. We assume that suppression during recognition does not, however, disrupt the phonological store on which the recognition task is based. This store is assumed to hold a less adequate trace of long words, because they have benefited less from rehearsal; when further rehearsal is prevented, further differential decline ceases.

Interpretation of our results using the alternative word length explanations of Neath and Nairne (1995) and G.D.P. Brown and Hulme (1995) are less straightforward, partly because the relevant models are themselves somewhat complex. In the case of the Brown and Hulme model this stems from the additional factor of redintegration, whereas the Neath and Nairne feature model involves a relatively complex set of parameters, as a result of its attempt to cover

a wide range of phenomena. Furthermore, although assumptions about the processes involved in rehearsal are not crucial to the word length effect in either model, it seems likely that neither would deny the existence of some kind of subvocal rehearsal, or of interference effects during the process of recall. Hence, it would presumably be open to them to suggest an interpretation that does not differ markedly from that already proposed, namely that word length effects influence both rehearsal and recall, although the proposed origin of word length effects would of course be different. Thus rehearsing sequences of long words would be more liable to error because of the greater number of components involved, a process that might also be assumed to occur during recall.

The failure of the word length effect to interact with articulatory suppression in Experiment 3 might seem to present a problem for these models, which one might assume would predict that anything that increases forgetting will further disadvantage the retrieval of long words, although one cannot be certain of this without running a full simulation on the relevant models.

In the meantime, we would suggest that the phonological loop hypothesis continues to provide a clear and simple account of our current results and of the rest of the increasingly rich pattern of results in this field. As such, we suggest that the phonological loop hypothesis should be preferred to more complex accounts.

REFERENCES

- Allport, D.A. (1984). Auditory-verbal short-term memory and conduction aphasia. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 313–326). Hove, UK: Lawrence Erlbaum Associates Ltd.
- Avons, S.E., Wright, K.L., & Pammer, K. (1994). The word length effect in probed and serial recall. *Quarterly Journal of Experimental Psychology*, 47A, 207–231.
- Baddeley, A.D. (in press). Levels of working memory. In M. Naveh-Benjamin, M. Moscovitch, & H.L. Roediger (Eds.), *Perspectives on human memory and cognitive ageing: Essays in honor of Fergus Craik*. Hove, UK: Psychology Press.
- Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In G.A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.
- Baddeley, A.D., & Hull, A.J. (1979). Prefix and suffix effects: Do they have a common basis? *Journal of Verbal Learning & Verbal Behavior*, 18, 129–140.
- 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.
- Brown, G.D.A., & Hulme, C. (1995). Modelling item length in memory span: No rehearsal needed? *Journal of Memory and Language*, 34, 594–621.
- Brown, J. (1974). Recognition assessed by rating and ranking. *British Journal of Psychology*, 65, 13–22.
- Burgess, N., & Hitch, G.J. (1992). Towards a network model of the articulatory loop. *Journal of Memory and Language*, 31, 429–460.
- Burgess, N., & Hitch, G.J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, 106, 551–581.
- Caplan, D., & Waters, G.S. (1994). Articulatory length and phonological similarity in span tasks: A reply to Baddeley and Andrade. *Quarterly Journal of Experimental Psychology*, 47A, 1055–1062.
- Coltheart, V., & Langdon, R. (1998). Recall of short word lists presented visually at fast rates: Effects of phonological similarity and word length. *Memory and Cognition*, 26, 330–342.
- Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language*, 31, 668–684.

- Cowan, N., Day, L., Sauls, J.S., Keller, T.A., Johnston, T., & Flores, L. (1992). The role of verbal output time in the effects of word-length on immediate memory. *Journal of Memory and Language*, 31, 1–17.
- Cowan, N., Keller, T.A., Hulme, C., Roodenrys, S., McDougall, S., & Rack, J. (1994). Verbal memory span in children—speech timing clues to the mechanisms underlying age and word-length effects. *Journal of Memory of Language*, 33, 234–250.
- Cowan, N., Nugent, L.D., & Elliott, E.M. (2000). Memory search and rehearsal processes and the word length effect in immediate recall: A synthesis in reply to Service. *Quarterly Journal of Experimental Psychology*, 53A, 666–670.
- Cowan, N., Nugent, L.D., Elliott, E.M., & Geer, T. (2000). Is there a temporal basis of the word length effect? A response to Service (1998). *Quarterly Journal of Experimental Psychology*, 53A, 647–660.
- Cowan, N., Wood, N.L., Nugent, L.D., & Treisman, M. (1997). There are two word length effects in verbal short-term memory: Opposed effects of duration and complexity. *Psychological Science*, 8, 290–295.
- Cowan, N., Wood, N.L., Wood, P.K., Keller, T.A., Nugent, L.D., & Keller, C.V. (1998). Two separate verbal processing rates contributing to short-term memory span. *Journal of Experimental Psychology: General*, 127, 141–160.
- Doshier, B.A., & Ma, J.J. (1998). Output loss or rehearsal loop? Output-time versus pronunciation-time limits in immediate recall for forgetting-matched materials. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 24, 316–335.
- Gathercole, S.E., Pickering, S., Hall, M., & Peaker, S. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *Quarterly Journal of Experimental Psychology*, 54A, 1–30.
- Gathercole, S.E., Service, E., Hitch, G.J., Adams, A.M., & Martin, A.J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology*, 13, 65–77.
- Henry, L.A. (1991). The effect of word length and phonemic similarity in young children's short term memory. *Quarterly Journal of Experimental Psychology*, 43A, 35–52.
- Henson, R.N.A. (1998). Short-term memory for serial order. The Start–End model. *Cognitive Psychology*, 36, 73–137.
- Hulme, C., Maughan, S., & Brown, G.C.A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short term memory span. *Journal of Memory and Language*, 30, 685–701.
- Hulme, C., Thomson, N., Muir, C., & Lawrence, W.A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology*, 38, 241–253.
- Jarrold, C., Hewes, A., & Baddeley, A.D. (2000). Two separate speech measures constrain verbal short-term memory in children. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 26, 1626–1637.
- Longoni, A.M., Richardson, J.T.E., & Aiello, A. (1993). Articulatory rehearsal and phonological storage in working memory. *Memory and Cognition*, 21, 11–22.
- Lovatt, P., Avons, S.E., & Masterson, J. (2000). The word length effect and disyllabic words. *Quarterly Journal of Experimental Psychology*, 53A, 1–22.
- Mackworth, J.F. (1963). The duration of the visual image. *Canadian Journal of Psychology*, 17, 62–81.
- Morrison, C.M., Chappell, T.D., & Ellis, A.W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, 50A, 528–559.
- Murray, D.J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, 78, 679–684.
- Neath, I., & Nairne, J.S. (1995). Word-length effects in immediate memory: Overwriting trace decay. *Psychonomic Bulletin & Review*, 2, 429–441.
- Norman, D.A. (1964). A comparison of data obtained with different false-alarm rates. *Psychological Review*, 71, 243–246.
- Page, M.P.A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105, 761–781.
- Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory & Cognition*, 12, 419–425.
- Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology*, 51A, 283–304.

- Service, E. (2000). Phonological complexity and word duration in immediate recall: Different paradigms answer different questions. A comment on Cowan, Nugent, Elliott, and Geer. *Quarterly Journal of Experimental Psychology*, 53A, 661–665.
- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, 153, 652–654.

Original manuscript received 20 September 1999

Accepted revision received 17 July 2001