THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2000, *53A* (1), 1–22

The Word-length Effect and Disyllabic Words

Peter Lovatt, S.E. Avons, and Jackie Masterson
University of Essex, U.K.

Three experiments compared immediate serial recall of disyllabic words that differed on spoken duration. Two sets of long- and short-duration words were selected, in each case maximizing duration differences but matching for frequency, familiarity, phonological similarity, and number of phonemes, and controlling for semantic associations. Serial recall measures were obtained using auditory and visual presentation and spoken and picture-pointing recall. In Experiments 1a and 1b, using the first set of items, long words were better recalled than short words. In Experiments 2a and 2b, using the second set of items, no difference was found between long and short disyllabic words. Experiment 3 confirmed the large advantage for short-duration words in the word set originally selected by Baddeley, Thomson, and Buchanan (1975). These findings suggest that there is no reliable advantage for short-duration disyllables in span tasks, and that previous accounts of a word-length effect in disyllables are based on accidental differences between list items. The failure to find an effect of word duration casts doubt on theories that propose that the capacity of memory span is determined by the duration of list items or the decay rate of phonological information in short-term memory.

A cornerstone of the memory span literature is the finding that lists made up of long words are harder to recall than lists of short words. Baddeley, Thomson, and Buchanan (1975) investigated the effect of the syllabic length of list items on immediate serial recall and found that as the number of syllables in a word increased serial recall performance decreased. In addition, Baddeley et al. (1975) found that reading rate and articulation rate correlated positively with serial recall across subjects and materials. This suggested that the limiting factor in serial recall was the time taken to articulate the list items rather than the number of syllables. In order to separate the effects of syllabic length and articulation rate, Baddeley et al. selected two sets of disyllabic words, which were matched for frequency and number of phonemes but which differed in terms of their articulatory duration. In tests of immediate serial recall the lists made from words with a short articulatory duration were better recalled than those made from long words. This finding confirmed that the articulatory duration of words determined memory span. The slope of

Requests for reprints should be sent to Peter Lovatt, Research Centre for English and Applied Linguistics, University of Cambridge, Keynes House, Trumpington Street, Cambridge CB2 1QA, U.K. Email: PJL31@cam.ac.uk
Peter Lovatt is now at the University of Cambridge.

the relationship between serial recall and articulation rate suggested that the capacity of serial recall was the number of items that could be spoken in about 1.8 sec (see also Schweickert & Boruff, 1986). Based on these results Baddeley et al. argued for a short-term memory system, the "phonological loop", consisting of a phonological store with a fixed decay rate supported by a separate rehearsal mechanism. The word-length effect arises because more short words than long words can be rehearsed before decay occurs in the phonological store (Baddeley, 1986). Rehearsal was assumed to be a process resembling covert speech, and hence speech rate was adopted as an indicator of rehearsal rate (Landauer, 1962). We will refer to this as the standard model of the phonological loop.

The model predicts within-subject effects of speech rate on memory span, such that memory span will be greater for materials that can be spoken faster. This has been confirmed in many studies that have demonstrated effects of syllabic length in adults (e.g. Avons, Wright, & Pammer, 1994; LaPointe & Engle, 1990; Standing, Bond, Smith, & Isely, 1980), and children (e.g. Henry, 1991a; Hitch, Halliday, & Littler, 1989; Hulme, Thomson, Muir, & Lawrence, 1984). Other studies have demonstrated differences in span across languages, which reflect articulation rates (e.g. Chen & Stevenson, 1988; Ellis & Hennelly, 1980; Naveh-Benjamin & Ayres, 1986). The model also predicts a strong correlation between speech rate and memory span across individuals, a relationship that has been reported in a number of studies (e.g. Baddeley et al., 1975; Gathercole, Adams, & Hitch, 1994; Standing & Curtis, 1989) but see Smyth and Scholey (1996) for an alternative interpretation. A further prediction of the model is that memory span should increase during development, because children's speech rates increase as they get older. This developmental relationship has been widely confirmed (Cohen & Heath, 1990; Hitch, Halliday, Dodd, & Littler, 1989; Hitch, Halliday, & Littler, 1989; Hulme et al., 1984; Hulme & Tordoff, 1989; Nicolson, 1981; Standing & Curtis, 1989). Thus there is a considerable body of evidence from different sources in support of the standard model of the phonological loop.

Developmental studies have shown that children show word-length effects from as early as 4 years old (e.g. Hitch, Halliday, Schaafstal, & Heffernan, 1991; Hulme et al., 1984; Johnston, Johnson, & Gray, 1987). This poses a problem for the rehearsal-based account, as it is questionable whether very young children engage in rehearsal (e.g. Bebko & McKinnon, 1990; Flavell, Beach, & Chinsky, 1966; Henry & Millar, 1993). Henry (1991b) suggested that decay during output may account for the word-length effect found in young children, and reported a corresponding absence of word-length effects when using probed recall. Avons et al. (1994) showed that the word-length effect in adults varied with output requirements, again suggesting some involvement of output processes. Cowan et al. (1992) studied the recall of mixed lists of long and short disyllabic words and found that the probability of recalling later words was critically determined by the type of word appearing earlier in the list. Recall of later items was poorer following recall of long words. All these results can be interpreted in terms of decay during list output. As such, this requires little modification to the standard model: Serial recall performance is limited by the number of items that can be rehearsed or output within the decay time of the loop.

A further difficulty with the standard model of the word-length effect arises from the presence of marked word-length effects in severe dysarthrics, patients with severe disturbances of articulation (Baddeley & Wilson, 1985; Bishop & Robson, 1989; Vallar &

Cappa, 1987). If rehearsal is producing the effect in these patients, it must be envisaged as a central process operating with abstract speech codes, rather than feedback from articulatory programs (Baddeley, 1986). Such a proposal weakens the theoretical grounds for a close linkage between rehearsal and overt speech, as reflected in the speech rate/span correlations. The extent to which rehearsal is *necessary* to explain the word-length effect has also been questioned. Some theorists have recently argued that word-length effects can be produced by passive decay without rehearsal (e.g. Brown & Hulme, 1995; Neath & Nairne, 1995).

Taking an alternative position, Caplan, Rochon, and Waters (1992) propose that the output planning requirements of long and short words are the cause of the word-length effect. Caplan et al. accept that dysarthric patients have normal memory spans and show normal word-length effects. However, they point out that patients with apraxia of speech (a high-level articulatory disturbance resulting in an inability to plan motor speech gestures) do not show word-length effects in immediate serial recall tasks. Caplan et al. suggest that differences in the complexity of planning motor speech gestures of 1- and 3-syllable words and not the articulatory duration of words are responsible for the production of the word-length effect (see also Klapp, 1974, 1976; Service, 1998; Waters & Caplan, 1995; Waters, Rochon, & Caplan, 1992).

To test this hypothesis, Caplan et al. investigated serial recall of word lists that were matched for number of syllables and number of phonemes and differed in spoken duration. Subjects responded by pointing to pictures depicting the words presented. The short-word list was made up of words containing principally lax vowels (e.g. carrot, ladder, devil), with a mean spoken duration of 546 msec, whereas the long-word list was made up of words containing principally tense vowels (e.g. sirloin, balloon, tower) with a relatively long mean spoken duration of 720 msec. In contrast to the findings of Baddeley et al. (1975), recall was higher for the long-duration words.

Baddeley and Andrade (1994) criticized this study on grounds that the measurement of speech rate, based on the duration of individual words rather than words uttered in combination, was inappropriate and that the two lists differed in terms of phonological similarity. Using the technique of speeded repetition of list pairs, Baddeley and Andrade reported only a small, and in one case non-significant, difference in duration between Caplan et al.'s (1992) long and short words. They also found that subjects rated the short items as more phonologically similar than the long items, suggesting that the recall advantage for long words was confounded by phonological similarity. Baddeley and Hitch (1994) further suggested that some of Caplan et al.'s long words might be truncated during rapid rehearsal (e.g. balloon to b'loon) thus diminishing differences in word duration.

In response to Baddeley and Andrade's (1994) criticisms, Caplan and Waters (1994) reported further serial recall results and new speech rate measures on the same stimuli. Subjects were first assessed on serial recall using 10 lists for each of the long and short word sets. They were then timed as they read aloud each list of words as quickly as they could. The same subjects then rated pairs of words, from the long- and short-word lists, for phonological similarity. Contrary to the findings of Baddeley and Andrade, they found: (1) a significant difference between the mean durations of the long and short lists, (2) no significant difference in ratings of phonological similarity, and (3) no difference in recalls of the long- and short-duration lists. They attributed the disparity between their own and Baddeley and Andrade's results for phonological similarity to variations in dialect.

Service (1998) investigated memory span for nonwords, making use of the phonological structure of Finnish to vary duration and complexity independently. In the Finnish language, there are long and short versions of vowels and consonants. Linguistically, long phonemes are regarded as equivalent to a repeated phoneme. In terms of articulation, Service maintains that items containing long phonemes are no more complex than those containing equivalent short phonemes. Hence increasing the duration of phonemes in a word (/tepa/ /te: p: a/) increases duration but leaves complexity constant, whereas increasing the number of phonemes (/tiempa/) increases both complexity and duration. Service showed that increasing duration by lengthening the duration of the phonemes did not decrease memory span. Zhang and Feng (1990) also reported no difference in the serial recall of short- and long-duration Chinese disyllables. They selected 288 words varying in duration, frequency, and graphic complexity, providing 18 separate conditions. Two lists of 8 words were presented under each condition, so each word was shown only once in the experiment. With this rather unusual design they reported that serial recall varied as a function of word frequency and graphic complexity but was not affected by word duration.

Thus, the present evidence is inconsistent concerning the existence of a word duration effect in memory span. Reliable span differences have been reported between mono- and polysyllabic words, but these words differ on several factors, including number of phonemic units, output planning requirements, prosody, intonation, complexity, concreteness, and linguistic origin. The critical evidence for decay in the phonological store rests on comparisons between disyllabic words that differ only in duration.

The failure of Caplan et al. (1992), Caplan and Waters (1994), Service (1998) and Zhang and Feng (1990) to find duration effects might be explained by a Type II error if their short and long items did not differ substantially in duration. In fact the percentage increases in duration of long words over short words reported by Caplan et al. (32%) and Service (31%) were smaller than the value reported by Baddeley et al. (1975) (67%). However, different methods were used to obtain these estimates, which are discussed later. Another possibility is that in Caplan et al. (1992) the word-length effect was attenuated because they used a picture-pointing task rather than spoken recall. Cowan and Kail (1996) suggest that word-duration effects arising at output may be nullified in such circumstances, because pointing response time will be independent of word duration. However, this objection will not apply if covert speech is used to select the response items. Caplan et al. argue that the production of word-length effects should not be affected by different recall methods as long as they make use of similar output-planning processes, and hence picture-pointing recall should be as sensitive as spoken recall to phonological complexity.

The present study investigated the effect of word duration and recall method on serial recall of disyllables. To do this, two new sets of disyllabic words were selected, which differed in spoken duration. Experiments 1a and 1b used the first set of words and two output techniques. Subjects responded either by pointing to pictures of the list items that had been presented or by recalling the list items verbally. In selecting the words for the present study strict criteria were employed in matching the words for frequency, familiarity, number of phonemes, and phonological similarity, and two measures of word duration were obtained. According to the standard model (Baddeley, 1986), duration differences operating through rehearsal should produce a word-duration effect regardless of recall method. Accounts based on output decay (Cowan & Kail, 1996; Cowan et al. 1992) also predict a word-duration effect but argue that this may be attenuated in pointing recall. In contrast, the recent findings of Caplan et al. (1992) and Service (1998) argue that serial recall is not affected by word duration, and hence for well-matched lists there should be no consistent advantage for short words, irrespective of recall method.

EXPERIMENT 1

Following Caplan et al. (1992) we tested each subject's serial recall at the span size at which they recalled approximately 60% of monosyllables correctly. A pilot study conducted on 10 subjects demonstrated that spoken and pointing recall showed the same level of performance for lists of five, seven, and nine items. Therefore, the same span length was adopted for each subject in both recall modes. In Experiment 1a individual word duration was measured for a large pool of words, and a set of eight short-duration and eight long-duration words was selected. These were then presented to 18 subjects using auditory presentation in two sessions: Session 1 involved measuring each subject's span and taking additional word-duration measures; Session 2 involved serial recall with long and short disyllabic word lists at the span size at which subjects scored 60% correct in Session 1. Each subject was tested separately, using a repeated measures design.

EXPERIMENT 1A

Method

Subjects

A total of 18 subjects (11 female and 7 male) were recruited from the University of Essex Subject Panel and were paid for their participation. The subjects ranged in age from 19 to 28 years (M = 21.3 years). All the subjects were native English speakers.

Materials

Word Span. A pool of 97 monosyllabic picturable nouns were selected from Snodgrass and Vanderwart (1980). These monosyllabic words were within the same frequency range as that of the 16 disyllabic words used in the serial recall experiments (outlined in the next section). A total of 10 lists were generated for each list length of four, five, six, seven, eight, and nine items, giving 60 lists in total. No word appeared more than once in each list.

Serial Recall. A total of 16 words were selected from a pool of 152 disyllabic imageable nouns, which had been selected from Carroll, Davies, and Richman (1971). The 16 words were selected such that 8 of them had a relatively short spoken duration and 8 had a relatively long spoken duration. Measures of duration were obtained for the words by asking six subjects from the University of Essex Subject Panel to read aloud the full set of 152 words, one at a time, at a normal pace. Subjects were asked to tap a pencil on the desk between each word so that separate recordings of each word could be obtained. Word duration was measured using the digital sound editing application SoundEdit run on a Macintosh LCII computer.

The final 16 words were selected with the following constraints:

1. The short and long words differed in terms of their spoken word durations. The mean duration of the short words was 605 msec ($SD = 31.4$), and the mean duration of the long words was 793 msec ($SD = 61$).
2. The long- and short-word sets were matched for standard frequency index using the Carroll et al. (1971) norms. The mean frequency of the short words was 53.3 (SD - 2.98), and that of the long words was 52 ($SD = 1.44$).
3. The words were matched for familiarity. Familiarity ratings were obtained by embedding the selected words in a list of 58 words. Fifteen subjects rated each word for familiarity following the instructions provided by Hirsh and Funnell (1995), on a 5-point Likert scale (1 = very unfamiliar, 5 = very familiar). The mean rating was 2.67 ($SD = 0.42$) for all short-word pairs and 2.98 ($SD = 0.59$) for long-word pairs.
4. The words were matched for number of phonemes. The mean number of phonemes was 5.3 ($SD = 0.18$) in the short words and 5.7 ($SD = 0.25$) in the long words.
5. The words were matched for phonological similarity. Phonological similarity ratings were obtained by pairing each selected word with every other word to produce 120 word pairs. Fifteen subjects rated each word pair on a 5-point Likert scale (1 = not similar, 5 = very similar). The mean rating across subjects was 1.6 ($SD = 0.79$) for short words and 1.53 ($SD = 0.81$) for long words. The short words were: *button*, *tractor*, *whistle*, *spider*, *pencil*, *pocket*, *shovel*, and *candle,* and the long words were: *pebbles*, *curtains*, *station*, *needle*, *branches*, *canoes*, *necklace*, and *robot*. The words used in each pool, together with their frequency values, familiarity ratings, number of phonemes and length of spoken duration can be found in Appendix I. The two sets of 8 words did not differ significantly on frequency, $t(14) < 1.0$, familiarity, $t(14) < 1.0$, number of phonemes, $t(14) < 1.0$, or phonological similarity, $t(14) < 1.0$. However, there was a significant difference in spoken duration, $t(14) = 7.74$, $p < .001$.

A total of 80 lists of words were generated for each span size at which subjects were tested; 40 lists each of either short or long disyllables were randomly generated (using randomized Latin squares). For 20 lists from each word pool a corresponding array of pictures was generated for picture-pointing recall. Each array contained a picture corresponding to each word in a particular list and one other (distractor) picture, plus a black square printed in the middle of the page. A different array was made up for each list. The position of pictures in each array was determined randomly for each word list, with the constraint that no picture should appear in the same position on consecutive trials. The arrays were printed on A4 paper.

Duration Measures. Three different random lists were constructed from each of the short- and long-word sets, and these were used to provide duration estimates. Two additional lists of eight words each contained a randomized combination of words from both word sets. These mixed-length lists were used for practice.

Procedure

Word Span. Lists of monosyllabic words were presented auditorily at a rate of one word per second, read by a male speaker, followed by immediate spoken recall. Subjects were instructed to say the word "blank" if they forgot a word in the list. Testing began with 10 lists of four words, followed by 10 lists of five words, and so on, incrementing the list length after every 10 trials. Subjects were informed when the lists were incremented. Tests ended when subjects failed to recall 60% of words at a given list length. Span was estimated as the maximum list length at which subjects could recall 60% of list items.

Duration Measures. Subjects were asked to read each list of disyllabic words first silently and then as quickly and as clearly as they could. This was repeated for all six lists. The two mixed-length lists were used as practice lists.

Serial Recall. There were four conditions made by combining word duration (short vs. long) with recall method (spoken vs. pointing). For each condition, subjects were first familiarized with the stimulus words (spoken recall) or pictures (picture-pointing recall). Subjects were told in advance how many words there would be in each presentation list and which recall method to use. The lists were read out with the words being spoken at a rate of one per second. Subjects were instructed that at the end of the presentation period they were to recall the words in the correct serial order either by pointing to pictures that corresponded to the words or by speaking the words out loud. Conditions were blocked using four blocks of 20 trials. The order of block presentation was determined by a randomized Latin square. With the picture-pointing recall method subjects were shown the array of pictures that corresponded to the list they had just heard. If they could not remember a word they were instructed to point to the black square in the middle of the array. In the spoken recall condition, subjects were instructed to begin recalling the words in the list when a sheet of paper containing four question marks was placed in front of them. If subjects were unable to recall a word they were instructed to say the word "blank". Once the appropriate number of responses had been made, in both the spoken and pointing recall conditions, the response sheet was removed and the next trial initiated. Subjects' responses were videotaped.

Results and Discussion

Duration Measures

For each subject, the spoken duration for each list was found. The mean reading duration was 2.94 sec ($SD = 0.24$) for lists of short words and 3.20 sec ($SD = 0.37$) for lists of long words. A related t test showed a significant difference in the list reading times for short and long words, $t(17) = 3.428$, $p < .01$.

Serial Recall

The percentage of correct recall for each condition and the standard deviations are given in Table 1.

Contrary to the conventional view that memory span is limited by word duration, more long words were recalled than short words. The results also showed an advantage for picture-pointing recall over spoken recall. Typical U-shaped serial position curves were observed at each list length. A 2 $\times$ 2 within-subjects analysis of variance (ANOVA) showed a significant main effect of word duration, $F(1, 17) = 5.6$, $p < .05$, and a significant main effect of recall method, $F(1, 17) = 26.1$, $p < .001$. The interaction was not significant, $F(1, 17) < 1.0$.

TABLE 1
Mean percentage of correct recall as a function of word length and recall method in Experiments 1a and 1b

	Recall Method	Short Words Mean	Short Words SD	Long Words Mean	Long Words SD
Experiment 1a	spoken	60.74	10.39	65.05	13.46
	pointing	68.41	13.32	73.23	10.44
Experiment 1b	spoken	56.91	14.50	59.55	12.68
	pointing	57.60	17.55	63.20	16.06

The results of Experiment 1a differed from those of Baddeley et al. (1975) in that long words were better recalled than short words with auditory presentation. The finding of a recall superiority for long words was unexpected. We decided to repeat the experiment using the same long- and short-duration words in serial recall in order to see if the results could be replicated. Items in Experiment 1a were read out by the experimenter, and it is possible that variability in presentation of the lists may have contributed to the results. In Experiment 1b we changed from auditory presentation to computer-controlled visual presentation of the stimuli in order to eliminate variations in presentation. An additional aim of Experiment 1b was to confirm that the same result could be obtained with a different presentation modality.

EXPERIMENT 1B

As the majority of the subjects in Experiment 1a had memory spans of six items, a list length of six items was adopted for serial recall for all subjects in this experiment.

Method

Subjects

A total of 18 subjects (9 male and 9 female) were recruited from the University of Essex Subject Panel and were paid for their participation. The subjects ranged in aged from 19 to 36 years (M = 24.1 years). All of the subjects were native English speakers. None of the subjects had taken part in Experiment 1a.

Materials

The same disyllabic lists and recall techniques were used as for serial recall in Experiment 1a, except that list length was fixed at six items. The stimuli were presented on a computer screen in 36-pt upper-case Geneva font.

Procedure

Following the presentation of a fixation point in the centre of the computer screen for 500 msec, the words in each list were shown at a rate of one word per second. Each word was displayed for 500 msec with a 500-msec interstimulus interval. Subjects were instructed that once all six words from a list had been presented they were to recall the words in the correct serial order either by pointing to pictures that corresponded to the words or by speaking the words out loud. Subjects were told in advance which recall method to use. The pictures were presented on paper, and in all other respects the procedure was the same as that for Experiment 1a.

Results and Discussion

The percentage of correct recall for each condition and the standard deviations are given in Table 1. The results show that long words were once again better recalled than short words, but in this experiment there was no marked difference for the two recall procedures. A 2 × 2 within-subjects ANOVA showed a significant main effect of word duration, $F(1, 17) = 6.1$, $p < .05$, but no main effect of recall method, $F(1, 17) < 1.0$. The interaction was not significant, $F(1, 17) < 1.0$.

Experiments 1a and 1b showed that when lists of words differing in spoken duration were presented either auditorily or visually, a recall advantage was found for long-duration words. There was no indication of any interaction between word duration and recall method. The results confirm those obtained by Caplan et al. (1992) who found no evidence for superiority of short-duration words in the recall of disyllables. Moreover, we used both spoken and pointing recall, making it unlikely that the failure of Caplan et al. to detect a superiority for the recall of short words was simply a consequence of their use of a picture-pointing recall technique, as suggested by Cowan and Kail (1996).

It is necessary to consider whether the results obtained in Experiments 1a and 1b may have been caused by factors other than word duration. Two potential problems with Experiments 1a and 1b can be identified. First, there may be relevant characteristics of the word sets that were not controlled, and second, there may be limitations in the way that word duration was measured. Concerning the first possibility, although the present word sets were controlled for word frequency, familiarity, phonological similarity, and number of phonemes, there were differences between the word sets. Some of the long words were plurals (*canoes*, *pebbles*, *branches*, *curtains*), whereas none of the short words were. In addition, *necklace* in the long-word set may be considered a compound word, whereas there were no compound words in the short-duration set. A further concern with the current findings is that the duration difference between the word sets may not have been sufficient to produce reliable differences in serial recall. We therefore attempted to maximize the duration difference between two new sets of short and long words, while attempting to control for the morphemic differences that existed in the sets used in Experiment 1.

EXPERIMENT 2

Experiments 2a and 2b were replications of Experiments 1a and 1b using new sets of disyllabic words. Experiment 2a used auditory presentation and both spoken serial recall and picture-pointing recall. Experiment 2b used visual presentation and both spoken serial recall and picture-pointing recall.

Selection of New Word Pools

The Oxford Psycholinguistic Database (Quinlan, 1992) was searched for disyllabic concrete nouns containing tense or lax vowels, with a printed word frequency value within the range 0–40 and a familiarity rating between 2 and 5.5. Unsuitable words (e.g. plurals and polymorphemic words) were rejected. This search generated a candidate set of 93 disyllabic words. Four subjects from the University of Essex Subject Panel (two male, two female) were asked to read each of these 93 words at a normal rate into a tape recorder. The spoken duration of each word was measured using the digital wave form analysis program SampleEdit on a Macintosh computer. The 93 words were then rank ordered according to their mean spoken duration, and the 30 longest words and shortest words were selected. From the 30 words in each group, 8 were selected. The mean spoken word duration of the short words was 509 msec (SD = 58.9), and that of the long words was 735 msec (SD = 51.1). The printed word frequency values for the words were obtained from the Hofland and Johansson (1982) printed word frequency count. The mean frequency of the short words was 7.5 (SD = 12.5), and that of the long words was 6.62 (SD = 13.5). Familiarity ratings were obtained for the words using the same procedure as that in Experiment 1. The mean familiarity rating was 3.62 (SD = 0.45) for short words and 3.53 (SD = 1.13) for long words. The mean number of phonemes in the short words was 5.12 (SD = 0.3), and that in the long words was 5.25 (SD = 0.7). Phonological similarity ratings were obtained for the words using the same procedure as that in Experiment 1. The mean rating was 1.26 (SD = 0.27) for short words and 1.27 (SD = 0.26) for long words. The eight short words that were selected were: *oblong, lemon*, *puppet*, *wizard*, *camel*, *tablet*, *coffin*, and *hammock*, and the eight long words were: *gazelle*, *brochure*, *sari*, *cartoon*, *mustang*, *cottage*, *protein*, and *cistern*. The words used in each pool, together with their frequency values, familiarity ratings, number of phonemes, and length of spoken duration can be seen in Appendix II. The two sets of eight words did not differ significantly on frequency, $t(14) < 1.0$, familiarity, $t(14) < 1.0$, number of phonemes, $t(14) < 1.0$, or phonological similarity, $t(14) < 1.0$. However, there was a significant difference in spoken duration, $t(14) = 8.158$, $p < .001$.

Having selected the long- and short-word sets, their duration was measured independently using three techniques. Ten native English speakers from the University of Essex Subject Panel (five female, five male) volunteered to take part. The duration of the individual words was measured by asking the subjects to read aloud each word individually at a normal rate. The subjects' speech was recorded on audio tape. The duration of each word was measured using SampleEdit, run on a Macintosh computer. For each subject the duration of each word was recorded, and the mean was calculated. The mean duration averaged across subjects for short words was 519 msec (SD = 18.5) and the mean duration for the long words was 676 msec (SD = 18.2), an increase of 30.2%. A related t test showed the durations of long and short words to be significantly different, $t(9) = 14.61$, $p < .001$.

To estimate reading rate, three different lists were made from word pools by randomizing the order of the words separately for the long- and short-word lists. The words in the six lists were printed in 30-pt lower-case Times font, double-spaced and centred on a sheet of A4 paper. Subjects were asked to read each list first silently, and then as quickly and as clearly as they could. After reading two practice lists, subjects read the six test lists. A different random order of lists was used for each subject. The spoken duration of each list was measured using a stop watch. The mean list duration was 2.22 sec ($SD = 0.29$) for short words and 2.87 sec ($SD = 0.56$) for the long words, an increase of 29.2%. A related t test showed the durations of long and short lists to be significantly different, $t(9) = 6.1$, $p < .001$.

The third measure taken was the spoken duration of word pairs. Each word in each set of long and short words was paired with each other word from the same set. Word pairs were printed on individual slips of paper in 36-pt lower-case Times font. Each subject was presented with eight word pairs, four pairs from each set of long and short words, such that subjects saw each word only once. On each trial subjects were asked to familiarize themselves with the word pair on the slip of paper. The experimenter said the word "Go", which was the signal for the subject to start reading the word pair as quickly as they could. When the subject had read the word pair 10 times the experimenter said "Stop". Word-pair repetition was timed using a stop watch. Two practice trials, using the word pairs *penny–bucket* and *apple–rabbit*, were given. For each subject the mean word-pair duration, measured from the beginning of the utterance of the 1st word to the final utterance of the 20th word, was found. The mean word-pair duration of the short words was 6.57 sec ($SD = 0.72$), and that of the long words was 7.78 sec (SD 0.7), an increase of 18.4%. A related t test showed the duration for short- and long-word pairs to be significantly different, $t(9) = 12.28$, $p < .001$.

EXPERIMENT 2A

Method

Subjects

A total of 18 undergraduate students from the University of Essex Psychology Department (14 female and 4 male) participated in this experiment for course credit. The subjects ranged in age from 18 to 42 years ($M = 27.3$ years). All the subjects were native English speakers. None had participated in Experiment 1.

Materials

Materials were the sets of short and long disyllabic words described earlier. The picture-pointing recall sheets contained pictures corresponding to these words.

Procedure

The procedure used was identical to that in Experiment 1a. All subjects were tested using six-item lists as in Experiment 1b.

Results and Discussion

For each subject the percentage of correct recall was measured at each word length using both spoken and pointing recall. These values are given in Table 2.

The recall of short words was not different from the recall of long words, but pointing recall again produced higher levels of recall than did spoken recall. A 2 × 2 within-subjects ANOVA showed no significant main effect of word length, $F(1, 17) < 1.0$, whereas there was a significant main effect of recall method, $F(1, 17) = 11.1$, $p < .001$. The interaction was not significant, $F(1, 17) < 1.0$.

EXPERIMENT 2B

Experiment 2b used visual presentation and both spoken serial recall and picture-pointing recall.

Method

Subjects

A total of 18 undergraduate subjects from the University of Essex Psychology Department (14 female and 4 male) participated for course credit. The subjects ranged in age from 18 to 46 years (M = 24.4 years). All the subjects were native English speakers. None had participated in the earlier experiments.

Procedure

The procedure used was identical to that in Experiment 1b.

Results and Discussion

For each subject the percentage of correct recall was found for each recall method for each word length. Overall mean performance, in terms of percentage of correct recall, and standard deviations are given in Table 2. There was no difference in the recall of long and

TABLE 2
Mean percentage of correct recall as a function of word length and recall method in Experiments 2a and 2b

		Short Words		*Long Words*	
	Recall Method	*Mean*	*SD*	*Mean*	*SD*
Experiment 2a	spoken	58.71	14.72	58.48	13.04
	pointing	63.80	12.80	65.51	14.79
Experiment 2b	spoken	56.26	13.28	57.09	12.71
	pointing	61.68	13.28	63.16	14.60

short words, and again pointing recall produced higher levels of recall than did spoken recall. A 2 × 2 within-subjects ANOVA showed no significant main effect of word length, $F(1, 17) < 1.0$, whereas there was a significant main effect of recall method, $F(1, 17) = 8.9$, $p < .001$. The interaction was not significant, $F(1, 17) < 1.0$.

The results of Experiments 2a and 2b confirm the findings of Experiments 1a and 1b and the findings of Caplan et al. (1992) in failing to find an effect of spoken duration for disyllabic words matched for number of phonemes and frequency. As in Experiment 1, we found no evidence that output modality affected the results.

It is possible that in some way our method was insensitive to word-duration effects. If this were the case then we would expect to find no effect of word duration using our procedures when word lists are used that have previously been found to show significant word-length effects. In order to test this possibility, a further experiment was conducted using the disyllabic words from Baddeley et al. (1975, Experiment 3). The same population of subjects was tested as in Experiments 1 and 2. Not all the words that Baddeley et al. used were concrete nouns, and therefore it was not possible to use a picture-pointing recall paradigm. We thus restricted the recall technique that we used in Experiment 3 to spoken recall. Our failure to detect consistent word-duration effects in serial recall in Experiments 1 and 2 might be because the duration differences between the word pools were small compared to those in the studies by Baddeley et al. In order to examine this further possibility, Experiment 3 also examined the duration of the words used by Baddeley et al., using the three measurement techniques described in Experiment 2.

EXPERIMENT 3

Word-Duration Measures

Method

The same three measures of word duration were taken of the disyllabic words used by Baddeley et al. (1975, Experiment 3) as in our Experiment 2. The short words were: *bishop*, *pectin*, *ember*, *wicket*, *wiggle*, *pewter*, *tipple*, *hackle*, *decor*, and *phallic*. The long words were: *Friday*, *coerce*, *humane*, *harpoon*, *nitrate*, *cyclone*, *morphine*, *tycoon*, *voodoo*, and *zygote*. Measurements were obtained from the same 10 subjects who took part in Experiment 2. The mean individual word duration averaged across subjects was 530 msec ($SD = 20$) for the short words, and 693 msec ($SD = 16.8$) for the long words, a significant increase of 30.7%, $t(9) = 15.08$, $p < .001$. The mean word list reading-rate measure averaged across subjects was 3.26 sec ($SD = 0.56$) for the short words and 3.8 sec ($SD = 0.61$) for the long words, a significant increase of 16.5%, $t(9) = 3.67$, $p < .01$. The mean word-pair duration averaged across subjects was 6.85 sec ($SD = 0.64$) for the short words and 8.07 sec ($SD = 0.86$) for the long words, a significant increase of 17.8%, $t(9) = 7.02$, $p < .001$.

Serial Recall

Method

Subjects

A total of 18 subjects (10 male and 8 female) were recruited from the University of Essex Subject Panel and were paid for their participation. Subjects ranged in age from 17 to 53 ($M = 28.8$ years). All of the subjects were native English speakers. None had participated in the earlier experiments.

Materials

A total of 40 lists of disyllabic words were generated from the long and short pools of words: 20 lists from each pool. The words were selected for each list without replacement, such that each word appeared in each serial list position twice. No word appeared in the same serial list position on consecutive lists.

Procedure

The procedure was the same as that in Experiments 1b and 2b, with the modification that only spoken recall was used following visual presentation.

Results and Discussion

The percentage of correct recall was calculated for each subject for long- and short-duration words. The mean percentage of correct recall for short disyllabic words, across all the subjects and serial positions, was 70.7 ($SD = 21.2$), and that for long disyllabic words was 65.5 ($SD = 20.9$). A 2×5 within-subject ANOVA showed a significant main effect of word length, $F(1, 17) = 8.7$, $p < .01$, and of serial position, $F(4, 68) = 47.2$, $p < .01$. The interaction was not significant, $F(4, 68) < 1.0$.

The main effects reported here are strikingly similar to the effects reported by Baddeley et al. (1975, Experiment 3) using the same word sets, suggesting that the current methodology is sensitive enough to detect whatever is causing the difference in recall for this particular set of short and long disyllabic words.

Independent estimates of the duration of the words used by Baddeley et al. (1975) and the words used in Experiment 2, using the same three techniques, have shown similar percentage differences in duration between the sets of words. Therefore, our replication of a disyllabic word-length effect using the Baddeley et al. stimuli cannot be attributed to greater differences in word duration between those words and our own.

GENERAL DISCUSSION

According to Baddeley's (1986) theory of the phonological loop, the traces of recently presented items decay and become unavailable unless rehearsed (or output) within the decay time. The theory proposes that the limit of phonological short-term memory is set by the decay rate of list items. As rehearsal is assumed to be a real-time process of covert speech, or something similar, the theory also predicts a linear relationship between serial recall performance and rate of speech, such that the number of items available is the product of the decay rate and the rehearsal rate. Hence for any two sets of items differing in their spoken duration, serial recall should be superior for the set of items with shorter spoken duration.

The present series of experiments has repeatedly failed to find evidence of superior serial recall for lists of short-duration words than for lists with longer spoken duration. In these experiments, the word lists were constructed from disyllabic words so as to maximize differences in spoken duration, whilst keeping the lists matched for potentially confounding variables such as word frequency, familiarity, and number of phonemes. This absence of duration effects was observed for both spoken recall and picture-pointing recall (Experiments 1 and 2). The one exception was Experiment 3, which used word lists culled from the original study by Baddeley et al. (1975). These findings call into question Baddeley et al.'s claim that spoken duration determines serial recall in disyllabic-word lists. They suggest that there is no general effect of word duration on disyllabic-word recall, and that the differences originally observed arose as an accident of item selection. As the disyllabic-word-duration effect is one of the major pieces of evidences in support of a decay-based capacity limit in short-term memory, the results of the present study cast doubt not only on the duration-based explanation of the word-length effect, but more generally on the standard phonological loop model, which is centred on a decaying phonological store supported by rehearsal (Baddeley, 1986).

A problem for the present and past studies is that word-duration effects are typically measured by contrasting recall of two restricted sets of words. Small sets of eight words were used in this study, and similarly small sets were used by Baddeley et al. (1975), by Cowan et al. (1992) who used a subset of the words used by Baddeley et al., and by Caplan et al. (1992). Small set sizes are an inevitable consequence of carefully matching lists for critical properties. The problem with using such small sets of words is that items will be sampled that vary in their characteristics. The presence of one or more unusually difficult items in a set may be enough to produce a significant difference in performance across word sets. As there is no way of testing for item effects when most or all items appear in each list, it is essential to examine the generality of the results across independently selected groups of words. Historically, this has not been done, and the original items have been retained and re-utilized in a number of studies. To our knowledge, it is only the original Baddeley et al. (1975) word sets, and subsets of these words (Cowan et al. 1992; Longoni, Richardson, & Aiello, 1993), that have shown a clear advantage for short-duration disyllables. We emphasize that no study in this regard can be definitive, including our own, and we urge appropriate caution. The results presented here are supported by other studies that failed to find word-duration effects in items controlled by syllabic length, such as Caplan et al. (1992), Caplan and Waters (1994), Service (1998), and Zhang and Feng (1990). Across these studies, and our own experiments, there has been a failure to detect disyllabic-word-length effects using five different sets of disyllabic words. In a complementary experiment, Henry and Millar (1991) devised word sets that varied in syllabic length but were matched for spoken duration. They reported superior recall for words with fewer syllables, and concluded that word duration was not the only determinant of serial recall (see also Hulme et al., 1997).

The relation between speech rate and serial recall has been studied in a number of ways. In addition to the study of short- and long-duration disyllables, which we call into question, there are differences: (1) across words varying in the number of syllables, (2) across equivalent words in different languages, and (3) across individuals.

Syllabic Word-length Effects

The first of these is the common observation that word span decreases as the syllabic length of the items increases. The criticisms raised here do not apply to the syllabic word-length effect, which has been obtained many times with different word sets in adults and children (e.g. Henry, 1991b; Hulme et al., 1984), with French words (e.g. Belleville, Peretz, & Arguin, 1992), with Italian words (e.g. Longoni et al., 1993), with Finnish-based nonwords (Service, 1998), and in Hebrew (e.g. Birnboim & Share, 1995). As syllabic length increases, so the phonological description of each item increases in complexity. Therefore increasing syllabic length will increase the load on any phonological memory system. Many of these studies have also shown that serial recall correlates with speech rate across word lengths when pooled across subjects. Again, this is hardly surprising, as speech rate is limited by the number of syllables that can be output in unit time and so must decrease as a function of syllabic length. On this basis there is no necessity to postulate a causal relationship between speech rate and memory span across words varying in syllabic length, as proposed by the rehearsal theory. Indeed, syllabic word-length effects can be explained by the increased output planning requirements with the number of syllables in a word (see Waters, Rochon, & Caplan, 1992).

Cross-linguistic Comparisons

Several studies have compared digit span across languages, and have shown that span varies as a function of speech rate. In most cases, the languages differ in terms of the number of syllables in the digit names, so that the span/speech rate relationship may be determined by syllabic word length (e.g. Chincotta & Hoosain, 1995; da Costa Pinto, 1991). The notable exceptions are Welsh and Chinese (in Mandarin or Cantonese dialects), for which the digit names are mainly monosyllabic. Ellis and Hennelly (1980) reported that Welsh–English bilinguals showed superior digit span in English, and they claimed that this was due to a word-duration difference between the digit names in the two languages. However, although Arabic numerals were read more slowly in Welsh, there were no differences in reading rates for lists of digit names in Welsh and English. It is therefore questionable whether the two languages differ in terms of digit articulation rates, and the differences in digit span may arise from a preference for processing digits and performing arithmetic in English.

Several studies have demonstrated faster articulation rates and superior memory span in Chinese than in English (e.g. Chincotta & Hoosain, 1995; Stigler, Lee, & Stevenson, 1986). Again there are several studies that weaken the evidence. In studies of bilinguals, for example, lower familiarity with the English language may account for or accentuate the differences between languages (Cheung & Kemper, 1993; Chincotta & Hoosain, 1995; Zhang & Simon, 1985). In comparisons of native monolinguals, problems of sampling, cultural and educational differences may contribute (see Stigler et al., 1986). Two findings are in direct conflict with the standard model regarding memory span and speech rate. First, in a study of young children Chen and Stevenson (1988) found that Chinese–English differences in digit span remained even after regression on list duration. Second, Cheung and Kemper (1993) found that the gradient of the span/speech rate function was much steeper in Chinese adults than in English. Cheung and Kemper ascribed the change in slope to the disproportionately high span scores for Chinese monosyllables, and they pointed out that Chinese monosyllables have low phonological complexity (syllables are CV or CVC), in contrast to English monosyllables, which may contain consonant clusters. Thus, although it is appealing to explain cross-linguistic differences in word and digit span in terms of word duration, the supporting evidence is not conclusive.

Individual Differences

Many studies have shown that individuals who speak faster generally have increased memory spans (e.g. Baddeley et al., 1975; Ellis & Hennelly, 1980; Hulme et al., 1984; Schweickert & Boruff, 1986). However, this could be explained by more efficient speech buffering (e.g. Monsell, 1987) or speech planning (Caplan et al., 1992) leading to enhanced speech rates, rather than by speech rate itself determining memory capacity. In a recent study Smyth and Scholey (1996) reported a significant relationship between articulation rate and digit span, but they also found that articulation rate predicted performance equally well on a number of visual and spatial memory tasks, some of them quite unlike serial recall. They proposed that speech rate is a general measure of processing speed, and as such is related to many cognitive skills.

Thus it is possible to account for the relationship between speech rate and memory span across syllabic length, languages, and individuals without invoking an explanation in terms of phonological decay offset by rehearsal or output rates. The most direct evidence for an effect of word duration, and hence of decay in phonological short-term memory, has been the effect of word duration across controlled sets of disyllables, an effect which is questionable, given the evidence presented here. One way to sustain the decay theory would be to propose that rehearsal rates are not reflected in speech rates, possibly because they represent a more abstract code. This proposal conflicts with subjective reports of rehearsal, and it would considerably weaken the utility of the theory, as speech rate could no longer be used to predict span. Nevertheless, a theory of this type could account for the lack of a relationship between duration and span in the present findings, and it could provide a general account of word-length effects if rehearsal was governed by syllabic length.

Several recent models of the word-length effect have not included rehearsal assumptions (e.g. Anderson & Matessa, 1997; Brown & Hulme, 1995; Neath & Nairne, 1995). However, all of these models suggest that the amount of decay or interference that occurs is determined by the duration of items, whereby longer duration items suffer more decay or interference. Therefore, these models make the same prediction as does the standard model (Baddeley, 1986) concerning word-duration effects in disyllables.

The present findings have implications for those studies that manipulated lists or recall methods to control output delay (Avons et al., 1994; Cowan et al., 1992; Cowan, Wood, &

Borne, 1994; Cowan, Wood, Nugent, & Treisman, 1997; Henry, 1991b). Henry (1991b) and Avons et al. (1994) compared probed and serial recall. As subjects only have to make one response per trial in probed recall, the output delay before making later serial position responses should be less for probed recall than for serial recall. Both Henry and Avons et al. found that the syllabic word-length effect was greater in serial recall than in probed recall. Although this is consistent with decay during output, the effects of recall procedure may arise from output interference or output buffer requirements. All these accounts suggest that serial recall for long words should decline towards the end of the list, an effect which is prominent in lists of polysyllabic words (Avons et al., 1994).

Cowan et al. (1992, 1994, 1997) manipulated output delay by controlling the duration of the words recalled first, and they found that recalling longer words at the start of output led to poorer recall of the remaining items, consistent with decay during output. However, of these studies Cowan et al. (1992) used a subset of the disyllabic words originally chosen by Baddeley et al. (1975), and the present paper questions the extent to which span for these words results from duration differences. A further difficulty is that, if Cowan et al.'s (1992) claim is correct, the serial position curves for lists of short and long words should diverge across serial positions. Although this pattern of results was found by Cowan et al. (1992, Experiment 1), no such effects were found in the present study (Experiment 3), and, in sharp contrast, Baddeley et al. (1975, Experiment 3) found that the serial position curves for long and short disyllables converged in later serial positions. Such results clearly do not support the output delay hypothesis.

The evidence for decay and duration effects in memory span rests heavily on the word-length effect for disyllabic words, and the serial position manipulations described by Cowan et al. (1992). This evidence was obtained from one set of words, originally chosen by Baddeley et al. (1975). We have confirmed the finding, earlier reported by Caplan et al. (1992), that the word-length effect for disyllabic words does not generalize to other word sets selected for word duration.

REFERENCES

Anderson, J.R., & Matessa, M. (1997). A production system theory of serial memory. *Psychological Review*, *104*, 728–748.

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. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A.D., & Andrade, J. (1994). Reversing the word-length effect: A comment on Caplan, Rochon, and Waters. *The Quarterly Journal of Experimental Psychology*, *47A*, 1047–1054.

Baddeley, A.D., & Hitch, G.J. (1994). Developments in the concept of working memory. *Neuropsychology*, *8*, 485–493.

Baddeley, A.D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short term memory. *Journal of Verbal Learning and Verbal Behaviour*, *14*, 575–589.

Baddeley, A.D., & Wilson, B. (1985). Phonological coding and short term memory in patients without speech. *Journal of Memory and Language*, *24*, 490–502.

Bebko, J.M., & McKinnon, E.E. (1990). The language experience of deaf children: Its relation to spontaneous rehearsal in a memory task. *Child Development*, *61*, 1744–1752.

Belleville, S., Peretz, I., & Arguin, M. (1992). Contribution of articulatory rehearsal to short-term memory: Evidence from a case of selective disruption. *Brain and Language*, *43*, 713–746.

Birnboim, S.L., & Share, D.L. (1995). Surface dyslexia in Hebrew: A case study. *Cognitive Neurophsychology*, *12*, 825–846.

Bishop, D.V.M., & Robson, J. (1989). Unimpaired short-term memory and rhyme judgment in congenitally speechless individuals: Implications for the notion of "articulatory coding". *Quarterly Journal of Experimental Psychology*, *41A*, 123–140.

Brown, G.D.A., & Hulme, C. (1995). Modelling item length effects in memory span: No rehearsal needed? *Journal of Memory and Language*, *34*, 594–621.

Caplan, D., Rochon, E., & Waters, G. (1992). Articulatory and phonological determinants of word-length effects in span tasks. *Quarterly Journal of Experimental Psychology*, *45A*, 177–192.

Caplan, D., & Waters, G. (1994). Articulatory length and phonological similarity in span tasks: A reply to Baddeley and Andrade. *Quarterly Journal of Experimental Psychology*, *47A*, 1055–1062.

Carroll, J.B., Davies, P., & Richman, A. (1971). *Word frequency book*. Boston, USA: Houghton Mifflin Company.

Chen, C., & Stevenson, H.W. (1988). Cross-linguistic differences in digit span of preschool children. *Journal of Experimental Child Psychology*, *46*, 150–158.

Cheung, H., & Kemper, S. (1993). Recall and articulation of English and Chinese words by Chinese–English bilinguals. *Memory and Cognition*, *21*, 666–670.

Chincotta, D., & Hoosain, R. (1995). Reading rate, articulatory suppression and bilingual digit span. *European Journal of Cognitive Psychology*, *7*, 201–211.

Cohen, R.L., & Heath, M. (1990). The development of serial short-term memory and the articulatory loop hypothesis. *Intelligence*, *14*, 151–171.

Cowan, N., Day, L., Saults, J.S., Kellar, T.A., Johnson, 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., & Kail, R. (1996). Covert processes and their development in short-term memory. In S.E. Gathercole (Ed.), *Models of short-term memory*. Hove, U.K.: Psychology Press.

Cowan, N., Wood, N.L., & Borne, D.N. (1994). Reconfirmation of the short-term storage concept. *Psychological Science*, *5*, 103–106.

Cowan, N., Wood, N.L., Nugent, L.D., & Treisman, M. (1997). There are two word-length effects in verbal short-term memory: Opposing effects of duration and complexity. *Psychological Science*, *8*, 290–295.

da Costa Pinto, A. (1991). Reading rates and digit span in bilinguals: The superiority of mother tongue. *International Journal of Psychology*, *26*, 471–483.

Ellis, N.C., & Hennelly, R.A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology*, *71*, 43–51.

Flavell, J.H., Beach, D.R., & Chinsky, J.M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development*, *37*, 283–299.

Gathercole, S.E., Adams, A.-M., & Hitch, G.J. (1994). Do young children rehearse? An individual-differences analysis. *Memory & Cognition*, *22*, 201–207.

Henry, L.A. (1991a). Development of auditory memory span: The role of rehearsal. *British Journal of Developmental Psychology*, *9*, 493–511.

Henry, L.A. (1991b). The effects of word-length and phonemic similarity in young children's short-term memory. *The Quarterly Journal of Experimental Psychology*, *43A*, 35–52.

Henry, L.A., & Millar, S. (1991). Memory span increase with age: A test of two hypotheses. *Journal of Experimental Child Psychology*, *51*, 459–484.

Henry, L.A., & Millar, S. (1993). Why does memory span improve with age? A review of the evidence for two current hypotheses. *European Journal of Cognitive Psychology*, *5*, 241–287.

Hirsh, K.W., & Funnell, E. (1995). Those old, familiar things: Age of acquisition, familiarity and lexical access in progressive aphasia. *Journal of Neurolinguistics*, *9*, 23–32.

Hitch, G.J., Halliday, M.S., Dodd, A., & Littler, J.E. (1989). Development of rehearsal in short-term memory: Differences between pictorial and spoken stimuli. *British Journal of Developmental Psychology*, *7*, 347–362.

Hitch, G.J., Halliday, M.S., & Littler, J.E. (1989). Item identification time and rehearsal rate as predictors of memory span in children. *Memory and Cognition*, *16*, 120–132.

Hitch, G.J., Halliday, M.S., Schaafstal, A.M., & Heffernan, T.M. (1991). Speech, "inner speech", and the development of short-term memory: Effects of picture-labeling on recall. *Journal of Experimental Child Psychology*, *51*, 220–234.

Hofland, K., & Johansson, S. (1982). *Word frequency in British and American English*. Bergen: Norwegian Computing Centre for Humanities.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1217–1232.

Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech rate and the development of short-term memory. *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.

Johnston, R.S., Johnson, C., & Gray, C. (1987). The word-length effect in young children: The effects of overt and covert rehearsal. *British Journal of Developmental Psychology*, *5*, 243–248.

Klapp, S.T. (1974). Syllable-dependent pronunciation latencies in number naming: A replication. *Journal of Experimental Psychology*, *102*, 1138–1140.

Klapp, S.T. (1976). Short-term memory as a response preparation state. *Memory and Cognition*, *4*, 721–729.

Landauer, T.K. (1962). Rate of implicit speech. *Perceptual and Motor Skills*, *15*, 646.

LaPointe, L.B., & Engle, R.W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*, 1118–1133.

Longoni, A.M., Richardson, J.T.E., & Aiello, A. (1993). Articulatory rehearsal and phonological storage in working memory. *Memory and Cognition*, *21*, 11–22.

Monsell, S. (1987). On the relation between input and output pathways for speech. In A. Allport, D. Mackay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Relations between listening, speaking, reading and writing*. London: Academic Press.

Naveh-Benjamin, M., & Ayres, T.J. (1986). Digit span, reading rate, and linguistic relativity. *Quarterly Journal of Experimental Psychology*, *38A*, 739–751.

Neath, I., & Nairne, J.S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin and Review*, *2*, 429–441.

Nicholson, R., (1981). The relationship between memory span and processing speed. In M.P. Freidman, J.P. Das, & N. O'Connor (Eds.), *Intelligence and learning*. New York: Plenum Press.

Quinlan, P.T. (1992). *Oxford Psycholinguistic Database*. Oxford, U.K.: Oxford University Press.

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.

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.

Smyth, M.M., & Scholey, K.A. (1996). The relationship between articulation time and memory performance in verbal and visuospatial tasks. *British Journal of Psychology*, *87*, 179–191.

Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

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.

Standing, L., & Curtis, L. (1989). Subvocalization rate versus other predictors of the memory span. *Psychological Reports*, *65*, 487–495.

Stigler, J.W., Lee, S., & Stevenson, H.W. (1986). Digit memory in Chinese and English: Evidence for a temporally limited store. *Cognition*, *23*, 1–20.

Vallar, G., & Cappa, S.F. (1987). Articulation and verbal short-term memory: Evidence from Anarthria. *Cognitive Neuropsychology*, *4*, 55–78.

Waters, G.S., & Caplan, D. (1995). What the study of patients with speech disorders and of normal speakers tells us about the nature of rehearsal. In R. Campbell & M. Conway (Eds.), *Broken memories, Case studies in memory impairment*. Oxford, U.K.: Blackwell.

Waters, G.S., Rochon, E., & Caplan, D. (1992). The role of high level speech planning in rehearsal: Evidence from patients with apraxia of speech. *Journal of Memory and Language*, *31*, 54–73.
Zhang, W., & Feng, L. (1990). The visual recognition and capacity of STM for Chinese disyllabic words. *Acta Psychologica Sinica*, *22*, 383–390.
Zhang, G., & Simon, H.A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory and Cognition*, *13*, 193–201.

Original manuscript received 8 August 1997
Accepted revision received 18 December 1998

APPENDIX

Appendix I
Standard frequency index, familiarity, number of phonemes, and mean spoken duration of the short- and long-duration disyllabic words used in Experiments 1a and 1b

Duration of Words	*Word*	*SFI*	*Familiarity*	*Phonemes*	*Duration*[a]
Short	button	53.3	2.71	5	567
	tractor	48.5	2.43	6	635
	whistle	53.9	2.43	5	583
	spider	54.8	2.43	5	593
	pencil	56.9	3.57	6	603
	pocket	56.5	2.57	5	604
	shovel	49.6	2.28	5	590
	candle	53.2	3.00	6	666
Long	pebbles	49.6	2.71	6	775
	curtains	48.7	4.00	6	830
	station	57.8	3.00	6	715
	needle	55.0	3.28	5	698
	branches	57.8	2.14	7	783
	canoes	49.9	2.85	5	845
	necklace	48.8	3.45	6	850
	robot	48.9	2.43	5	850

[a]Duration given in msec.

Appendix II
Familiarity, frequency, number of phonemes, and mean spoken duration of the short- and long-duration disyllabic words used in Experiments 2a and 2b

Duration of Words	*Word*	*Familiarity*	*Frequency*	*Phonemes*	*Duration (msec)*
Short	oblong	3.12	3	5	603
	lemon	4.32	38	5	549
	puppet	3.56	1	5	483
	wizard	3.48	2	5	416
	camel	3.12	5	5	506
	tablet	4.12	4	6	506
	coffin	3.92	7	5	554
	hammock	3.36	0	5	459
Long	protein	4.08	5	6	783
	brochure	4.28	1	5	701
	sari	2.60	2	4	698
	cartoon	5.36	2	5	816
	mustang	2.76	0	6	697
	cottage	4.36	40	5	718
	gazelle	2.12	1	5	783
	cistern	2.72	2	6	681