THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2000, *53A* (3), 647–660

Is there a temporal basis of the word length effect? A response to Service (1998)

Nelson Cowan, Lara D. Nugent, Emily M. Elliott, and Tara Geer
University of Missouri, Columbia, Missouri, U.S.A.

Service (1998) carried out a study of the word length effect with Finnish pseudowords in which short and long pseudowords were identical except for the inclusion of certain phonemes differing only in pronunciation length, a manipulation that is impossible in English. She obtained an effect of phonemic complexity but little or no word duration effect per se—a discrepancy from the expectations generated by the well-known working memory model of Baddeley (1986). In the present study using English words, we controlled for phonemic complexity differences by using the same words for the short- and long-word sets, but with instructions inducing shorter or longer pronunciation of the words. We obtained substantial word duration effects. Concerns raised by Service are addressed, and we conclude that both duration and complexity are likely to contribute to the word length effect in serial recall.

Working memory could be defined as the temporarily heightened accessibility of information to be used to carry out a task such as language comprehension, reasoning, or problem solving. It is generally agreed that working-memory capacity is limited. The leading model of working memory (Baddeley, 1986) proposes central executive control of at least two subsidiary systems: a phonological loop (which is the focus of the present paper) and a visuospatial sketchpad. Some key evidence of the nature of the phonological loop's operation is the word length effect investigated by Baddeley, Thomson, and Buchanan (1975) and numerous subsequent studies. Baddeley et al. found that the serial recall of word lists is superior when the words can be pronounced quickly. According to their theoretical account, lists of shorter words are recalled better because more of those words can be silently rehearsed, and thereby reactivated, before they fade from temporary storage. In a slightly different account that is still fundamentally compatible with Baddeley's storage-and-decay model, lists of shorter words are recalled better because more of them can be pronounced in overt verbal recall before they fade from temporary storage (Brown & Hulme, 1995; Cowan et al., 1992; Henry, 1991).

Requests for reprints should be sent to Nelson Cowan, Department of Psychology, 210 McAlester Hall, University of Missouri, Columbia, MO 65211, U.S.A. Email: CowanN@missouri.edu

This project was supported by NICHD grant RO1 HD–21338 awarded to NC. We thank Nate Fristoe, Mike Page, J. Scott Saults, Elisabet Service, Michel Treisman, and an anonymous reviewer for valuable comments, and Simone Ispa-Landa for some of the timing measurements.

http://www.tandf.co.uk/journals/pp/02724987.html

In a competing type of account it is not the spoken duration of words per se that is critical in producing the word length effect, but rather some other characteristic of the words. Caplan, Rochon, and Waters (1992) and Caplan and Waters (1994) suggested that it is the phonemic complexity rather than the duration of words that makes a difference, and that the word length effect arises because longer words typically are more complex phonemically.[1]

For the purpose of the present paper, we define phonemic complexity as the number of transitions from one phoneme to a different phoneme. If both the time and the phonological complexity of words are important for immediate serial recall, then only a moderate degree of modification of the working-memory theory is needed. However, a more extreme modification would be needed if only phonological complexity were important and there were no effect of time at all. Exactly this case was argued by Service (1998) on the basis of serial recall experiments in Finnish. In that language, vowels have both a short-duration and a long-duration form, and consequently the duration of words could be manipulated without altering the phonological complexity of the word, by specifying which form of the vowel should be used. The duration of vowels in the words had at best a very small effect on recall, whereas the complexity of the words had a large effect. Service therefore suggested that there may be no time-based word length effect in immediate serial recall.

In reaching this conclusion, Service (1998) attempted to explain away discrepant results obtained by Cowan, Wood, Nugent, and Treisman (1997, Experiment 2). Cowan et al. used English words in a procedure in which the subject was to pronounce each word with the same duration as shown by a series of asterisks that could last for 300 ms (short, S) or 600 ms (long, L), both during the visual presentation of a list and during spoken recall. There were six words in a list, and the duration cue was manipulated separately for the first three words and for the last three words, and the duration cue for any particular word was the same during presentation and recall. Thus, there were four configurations of word duration cues for a list: SS, SL, LS, and LL. The number of syllables in each word (one or two) also was manipulated independently in each half-list, and there were some filler trials in which the cue durations or number of syllables varied randomly from word to word. (Backward recall was used, for a theoretical reason that is unimportant here. In particular, it was an attempt to require recall first for items that had been available for rehearsal for the shortest time.)

Cowan et al. (1997) found that there was an advantage for lists composed of short words, and this short-word advantage grew larger across output positions, so that it was largest at the first few input serial positions of the list. However, there also was a disadvantage for switching timing cues in the middle of a list, so that performance was poorer overall for SL and LS lists than for SS and LL lists. Finally, there was an advantage for disyllables over monosyllables and an advantage especially for lists com-

[1] Neath and Nairne (1995) provided a theoretically more neutral account, in the form of a mathematical model in which it is the larger number of “segments” in longer words that produces the word length effect. These segments were not defined and could refer either to units of time taken up by a speech sound or to distinct phonic identities, one per segment.

posed of three words of each kind (three monosyllables and three disyllables or vice versa).

Service (1998) had several criticisms of this study. First, she noted: "The study showed a relatively small word-length effect affecting the three items recalled last in the six-item lists presented, although the difference in spoken durations of the items as well as the pauses between the items was over 80%. This suggests that the effect was limited to a forgetting mechanism active during output" (p. 285). Second, she added: "Performance was impaired when subjects were made to recall half-lists at different pronunciation speeds. This suggests that at least re-programming output pace during recall is a secondary task that requires working memory capacity." (p. 285). Third, she went on to suggest the following: "Whether maintaining a single pronunciation speed also works as an additional interfering task is unclear at the moment. If it does, it can be expected to become effective over a time lag in a manner similar to the interfering task used in the Brown–Peterson paradigm. Thus, the results from an experiment based on subjects consciously controlling output durations still leave questions unanswered, and the assumption that the origin of the word-length effect is the spoken duration of words remains in question" (p. 285). Fourth, she noted: "In the study by Cowan et al. (1997), in which subjects consciously regulated their output pace, both items and pauses between them were over 80% longer in the long condition, yet the word-duration effect remained small, of the order of 5%. Thus, in their study, a much greater difference in output duration resulted in an effect of roughly the same size as the difference in recall between short and simple long items in this experiment [her Experiment 3] in which the recall times differed by less than 20%" (p. 297).

If these points made by Service (1998) are valid, the potential impact of that paper for current working-memory models is notable. There have already been modifications of the working-memory model, but these modifications generally still make use of a time factor. For example, Hulme, Maughan, and Brown (1991) showed that not all factors that affect recall also affect the duration of pronunciation. They found that a word length manipulation resulted in a linear relation between speech rate and memory span, as expected from the results of Baddeley et al. (1975), but that the use of word versus nonword stimuli (in their Experiment 1) altered the intercept rather than the slope of the linear relation between speech rate and memory span (although the use of Italian words before and after training on those words, in their Experiment 2, resulted in a change in both the intercept and the slope). On the basis of that paper, one could postulate a model like that of Baddeley (1986), but with an added contribution of lexical knowledge to recall. In contrast, Service (1998) presumably would assume that the relation between speech rate and memory span is itself spurious and occurs only because the longer words used by Hulme et al. (1991), and in all other experiments demonstrating the word length effect, were phonologically more complex than the shorter words. If there is no time factor per se in the word length effect, as Service claimed in the title to her article, then the apparent time-based effect must occur because of a confounding between time and phonological complexity. Moreover, that confounding would not be inevitable, as is clear from the fact that in Service's own stimuli, lists of long two-syllable words and lists of long three-syllable words could be read in approximately the same time.

The purpose of the present paper is to reassess these strong theoretical claims of Service (1998) and her reasons for dismissing the contradictory findings of Cowan et al. (1997). Some of the concerns voiced by Service were related to perceived weaknesses in the findings of Cowan et al. (1997, Experiment 2). However, as mentioned earlier, the Cowan et al. experiment was conducted using backward recall.

Some prior research suggests that forward and backward recall relies upon different retrieval strategies, such as a greater reliance on auditory–verbal codes in forward recall versus visual-spatial or semantic codes in backward recall (Hulme, et al. 1997; Li & Lewandowsky, 1995). Perhaps this is the basis of some of the weakness in the results of Cowan et al. (1997) pointed out by Service (1998). If backward recall results in a use of codes or strategies that is atypical for word length studies, it may produce atypically small (though still significant) word length effects. This could occur, for example, if there are two bases of the word length effect, one related to the duration of covert rehearsal and the other related to the duration of overt pronunciation (Cowan et al., 1992; Henry, 1991). Backward recall may discourage the use of covert rehearsal, eliminating one of two sources of the effect and therefore minimizing its magnitude. To redress this problem, we report here results from an experiment identical to that of Cowan et al. (1997, Experiment 2), in which identical words were pronounced in a short versus a long manner to produce short versus long conditions; except that forward instead of backward recall was used here. The similarity in methods to those used by Cowan et al. (1997) provides continuity inasmuch as cross-study comparisons can be suggested. With the results of this forward recall experiment in hand, we will re-examine the status of the logical arguments made by Service.

Method

Subjects

A total of 26 University of Missouri college students (10 male, 16 female) received course credit for their participation in this experiment.

Apparatus, stimuli, and procedure

The apparatus, stimuli, and procedure were identical to those used by Cowan et al. (1997) except that forward recall was used here. As in Cowan et al.'s experiment, the monosyllabic words were *dome, gate, lamp, nest, rice*, and *tank*, and the disyllabic words were *bedroom, cocktail, football, hardware, platform,* and *sunlight*. These sets are approximately matched in frequency (Kucera & Francis, 1967) and imageability (Quinlan, 1992). No word was used more than once in a list, and all lists included six words. Subjects read each item aloud at input and recalled it aloud, at a fast or slow rate.

The experiment took place in a quiet room, and stimuli appeared on the screen of an IBM-compatible, 486-type computer in the ordinary screen font, in white lettering on a black background. There was an initial, brief training phase in which the subject learned to control pronunciation duration (six trials) and to use this skill during recall (three trials). Each memory trial began with the words "Get Ready" presented for 2 s in the centre of the screen. This was followed by a timing cue (a row of asterisks approximately in the centre of the screen, growing longer for 300 or 600 ms). At the end of the cued period, the timing cue disappeared from the screen. One second after the onset of the timing cue, a word came on the screen and lasted for 1 s. Each word was followed immediately by the next timing cue. After the last word in the list, the timing cue for the recall of the first word was immediately presented. The subject had 2.5 s to recall each word, but each 2.5-s recall interval started with a 300- or 600-msec timing cue, the same one that had been used for the to-be-recalled word when it was presented. During the presentation of this timing cue, the subject was instructed not to speak. For the remainder of the recall period for a word, a three-symbol string (*?*) was instead presented, as a recall signal. As the subject spoke each word in recall, the experimenter, sitting slightly behind, recorded the responses in a pre-arranged manner, requiring only first-letter notation for each word in the set, on a sheet attached to a clipboard out of the subject's view. Each word was counted correct only if it was recalled in the correct serial position, and subjects were instructed to guess if they did not know a word or to say the word "blank" if they could not even guess. The final recall period was followed by the printed word "Stop" and then, after the "enter" key was pressed three times, a vertically arranged feedback list of the correct answers. Later, the subjects' answers were transcribed for computer analysis.

Most of the stimuli appeared at the centre of the computer screen. However, the first recall signal (*?*) appeared offset toward the upper left, and the recall signals for successive words on the list appeared lower and further to the right on the screen, to support forward recall. This progression was the opposite of what occurred in the experiment of Cowan et al. (1997), in which the signals progressed from lower right to upper left to support backward recall.

Six test trial blocks were run, each of which included one trial of each of 16 trial types, in a new random order each time, for a total of 96 test trials. The 16 trial types can be explained by describing the words in the order "first half list–second half list". Using that notation, there were the following trial types: 4 in which items in both halves of the list had the same number of syllables and the same spoken durations (1/Short–1/Short, 1/Long–1/Long, 2/Short–2/Short, and 2/Long–2/Long); 4 in which the number of syllables changed in the middle of the list, whereas the durations remained the same (1/Short–2/Short, 2/Short–1/Short, 1/Long–2/Long, and 2/Long–1/Long); 4 in which the pronunciation duration cues changed in mid-list, whereas the number of syllables remained fixed (1/Short–1/Long, 1/Long–1/Short, 2/Short–2/Long, and 2/Long–2/Short); and 4 filler trial types in which either the number of syllables or the duration cues varied randomly within the trial (1/Random–1/Random, 2/Random–2/Random, Random/Short–Random/Short, and Random/Long–Random/Long).

The proportion correct was scored in each condition, with items counted as correct only if reported in the correct serial position. Random filler trials were excluded from the analyses. The durations of all available word repetitions during presentation and of words spoken in their correct serial positions during recall were measured from tape recordings using a speech waveform editor (SoundEdit 16 by Macromedia, Inc., San Francisco, California, U.S.A.) on a Macintosh computer. Tape recordings of the words produced by the participant were digitally recorded in the computer and examined oscillographically. The program allowed any highlighted segment of the sound stream to be heard and displayed its duration to the nearest millisecond. Both visual and acoustic cues were used to ensure that each word's duration was measured accurately.

Results

As in Cowan et al. (1997), the duration of one trial of each of the 12 conditions (which do not include random filler trials) was randomly sampled for each participant. Due to difficulties in the electronic setup, timing measurements were available for only 14 participants. In all, 996 words from 166 lists during input and 441 words from 152 lists during recall were measured. The pattern of mean pronunciation durations for these participants, shown in Table 1, makes it clear that the duration manipulation worked in much the same

TABLE 1
Mean duration of word pronunciations in 14 subjects for input and output phases of a trial, separately for each half-list

		Number of syllables and pronunciation instructions			
	Measure	*1–Long*	*1–Short*	*2–Long*	*2–Short*
Input, List-half 1	Mean	0.92	0.50	0.99	0.59
	SD	0.17	0.13	0.19	0.13
Input, List-half 2	Mean	0.85	0.51	0.91	0.63
	SD	0.13	0.11	0.19	0.14
Output, List-half 1	Mean	0.89	0.53	0.95	0.62
	SD	0.21	0.16	0.20	0.12
Output, List-half 2	Mean	0.92	0.55	1.01	0.64
	SD	0.26	0.12	0.14	0.13

Note: Means refer to the mean durations of words as pronounced by the participants while reading the stimuli aloud during list presentation (input) and recall (output).

way as in Cowan et al. (1997, Experiment 2). To compare the durations of words under short and long pronunciation instructions, eight separate *t* tests were conducted (for one- and two-syllable words, from the first and the second list halves, pronounced during presentation or during recall). All eight tests yielded *p*s $< .002$.

Mean performance levels on each of the 12 trial types in each serial position appear in Table 2. The table is arranged in such a way that pairs of conditions differ only in the duration cues. Among the first four pairs of conditions, the first condition in the pair has the short duration cue and the second condition has the long duration cue. One can see that there was an overall advantage for the short duration cue in each case (as shown in the rows labelled "difference"). A 95% confidence interval for the mean difference between conditions (rightmost column of the table) indicates that the short duration cue advantage was reliable for the first, third, and fourth comparisons (i.e., the confidence interval does not include 0), and nearly so for the second comparison (the confidence interval does not go below 0). Thus, the duration-based word length effect obtained by Cowan et al. (1997) with backward recall occurs also in forward recall. Here it was, on the average, an 8.5% effect. Notice also that when a list contained half short words and half long words (the fifth and sixth comparisons), it did not matter which half had the short words.

In Figure 1, all conditions with a particular series of duration cues (short–short, short–long, long–short, and long–long) are combined. The scores corresponding to those means were subjected to an analysis of variance (ANOVA) with duration condition (4) and serial position (6) as factors. This analysis produced effects of the duration cue condition, $F(3, 75) = 14.10$, $MSE = 0.03$, $p < .001$, and serial position, $F(5, 125) = 45.13$, $MSE = 0.04$, $p < .001$, and their interaction, $F(15, 375) = 1.99$, $MSE = 0.01$, $p < .02$. Explaining the main effect of duration cue condition, post hoc Newman–Keuls tests showed that the proportion correct on short–short lists (.53) was higher than that on

TABLE 2
The proportion correct in each condition and serial position

	Serial position							
Condition	*1*	*2*	*3*	*4*	*5*	*6*	*Mean*	*95% C.I.*
1/S–1/S	0.84	0.71	0.50	0.44	0.35	0.37	0.54	
1/L–1/L	0.68	0.50	0.38	0.31	0.28	0.32	0.41	
Difference	0.16	0.21	0.12	0.12	0.07	0.05	0.12	[0.07–0.17]*
2/S–2/S	0.72	0.56	0.39	0.37	0.29	0.31	0.44	
2/L–2/L	0.68	0.45	0.37	0.31	0.23	0.26	0.38	
Difference	0.04	0.11	0.02	0.06	0.06	0.06	0.06	[0.00–0.12]
1/S–2/S	0.80	0.65	0.57	0.50	0.38	0.43	0.56	
1/L–2/L	0.67	0.53	0.43	0.48	0.31	0.41	0.47	
Difference	0.13	0.13	0.14	0.02	0.06	0.02	0.08	[0.03–0.13]*
2/S–1/S	0.79	0.68	0.55	0.49	0.46	0.47	0.57	
2/L–1/L	0.70	0.51	0.45	0.49	0.37	0.43	0.49	
Difference	0.09	0.17	0.10	0.00	0.08	0.04	0.08	[0.02–0.14]*
1/S–1/L	0.66	0.56	0.42	0.43	0.26	0.31	0.44	
1/L–1/S	0.60	0.51	0.38	0.41	0.31	0.34	0.42	
Difference	0.06	0.05	0.04	0.02	−0.05	−0.03	0.02	[−0.03–0.07]
2/S–2/L	0.62	0.40	0.36	0.42	0.33	0.24	0.39	
2/L–2/S	0.60	0.40	0.40	0.38	0.27	0.30	0.39	
Difference	0.01	0.00	−0.04	0.04	0.06	−0.06	0.00	[−0.06–0.06]

Note: Asterisks indicate that the 95% confidence interval for the difference averaged across serial positions does not include 0. To illustrate the notation used in this table, 2/S–2/L refers to lists in which the first three words were disyllables pronounced in a short manner and the last three words were disyllables pronounced in a long manner.

short–long (.42), long–short (.41), or long–long (.44) lists (*p*s $< .001$), which did not differ from one another. Figure 1 shows the nature of the serial position effect and interaction. Regarding the latter, notice that the advantage for the short–short condition over other conditions occurred across serial positions but was largest at the early serial positions, which were also the positions recalled first.

Figure 2 shows the effect of the number of syllables, combined for all lists in which there were 1–1, 1–2, 2–1, or 2–2 syllables in the first and second half lists, respectively. To examine the effects of syllables, an ANOVA was conducted with syllabic condition (4) and serial position (6) as factors. This analysis produced main effects of the number of syllables, $F(3, 75) = 18.61$, $MSE = 0.03$, $p < .001$, and serial position, $F(5, 125) = 48.65$, $MSE = 0.04$, $p < .001$, but no interaction, $F(15, 375) = 1.08$, $MSE = 0.01$, $p = .38$. A Newman–Keuls test of the syllabic conditions showed that performance levels in the 1–2 condition (.51) did not differ from the 2–1 condition (.53) but that both of them were higher than the 1–1 condition (.45) and the 2–2 condition (.40), *p*s $< .003$.

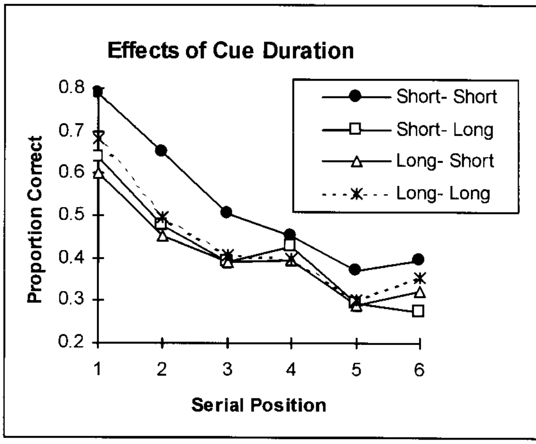

Figure 1. Effects of pronunciation cue duration. "Short–Long" = trials in which the first- and second-half-list cues were short (300 ms) and long (600 ms), respectively; and so on.

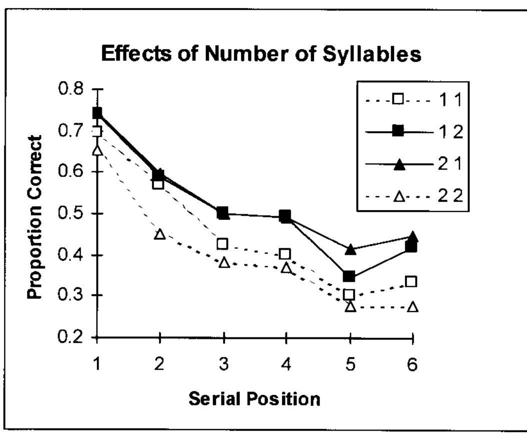

Figure 2. Effects of number of syllables per word. "1 2" = trials in which the first- and second-half-list words were monosyllabic and disyllabic, respectively; and so on.

This advantage of lists with mixed syllabic lengths over lists with uniform syllabic lengths appears to reflect the ease of grouping a mixed list into monosyllabic and disyllabic list halves. Support for this interpretation comes from position errors made on words from Serial Positions 3 and 4, the items at the boundaries of the supposed groups in the mixed conditions. Transposition errors betwen Serial Positions 3 and 4 occurred in 4% of all trials in the uniform syllabic conditions, but in only 1% of all trials in the mixed syllabic conditions, a significant difference (Table 3, top row). Also, as shown in the next two rows of Table 3, when the item in Serial Position 3 migrated to Position 4 in the response or vice versa (whether as part of a transposition error or not), the pattern of migration differed for the uniform versus mixed syllabic conditions. Specifically, the tendency for the erroneously recalled item to migrate into the wrong half of the list in the output protocol was much stronger for the uniform syllabic condition than for the mixed syllabic conditions. The bottom two rows of Table 3 show that this difference between conditions was not simply the result of more migration errors of any kind in the uniform syllables condition, inasmuch as migrations to the adjacent item in the same half list occurred at least as often in the mixed as in the uniform syllables condition. Thus, the uniform syllables condition produces more migration errors specifically across list halves, presumably because such lists tend not to be parsed into subgroups of 3 and 3 items, in contrast to lists in the mixed syllables condition.

A Newman–Keuls test showed that performance was higher for the 1–1 condition than for the 2–2 condition, $p < .02$, indicating a disadvantage for phonological complexity. This result differs from what Cowan et al. (1997) found in backward recall, in which there was an advantage for disyllables over monosyllables. However, the present finding of a disadvantage for phonological complexity matches what has been reported by other investigators using forward recall (e.g., Caplan et al., 1992; Service, 1998).

Our other present results seem similar to those of Cowan et al. (1997) obtained with backward recall, although the effects of word duration cues were considerably larger in the present experiment using forward recall. In fact, notice from Figure 1 that, at Serial Position 2, the magnitude of the word duration effect was about 20%.

Discussion

Service (1998) questioned the basis of the word length effect and concluded that the duration of words played little, if any, role. Instead, she suggested that the duration of words is only a correlate of the causal factor, the phonological complexity of words. In this study we have shown that there is a substantial effect of word duration obtained using the same words within the short and long sets, with instructions that induced participants to produce pronunciations of short versus long durations. The procedure was identical to one used previously by Cowan et al. (1997) except that forward rather than backward

TABLE 3
Mean proportion of uniform- and mixed-syllable trials with transposition and migration errors involving serial Positions 3 and 4

	Uniform syllables		*Mixed syllables*		$t(25)$	p
Serial Position error type	*M*	*SD*	*M*	*SD*		
Transposition, Serial Positions 3 & 4	.04	.03	.01	.02	3.43	.002
Migration, Serial Position 3 to Output 4	.08	.05	.03	.04	4.92	.000
Migration, Serial Position 4 to Output 3	.14	.07	.08	.07	4.29	.000
Migration, Serial Position 3 to Output 2	.12	.05	.14	.08	−1.55	ns
Migration, Serial Position 4 to Output 5	.07	.04	.17	.04	−3.17	.005

Note: Migration errors from Positions 3 to 4 and vice versa include the transposition errors.

recall was used. The word length effect proved to be much larger in forward recall. We also found an advantage for phonologically simpler words (monosyllables) over more complex words (disyllables) that is in keeping with Caplan et al. (1992) and Service (1998).

Given our result, the points raised by Service (1998) in criticism of Cowan et al. (1997) can be re-evaluated. The first point is the suggestion that because the pronunciation duration effect obtained by Cowan et al. was limited to the three words recalled last, the result suggested that "the effect was limited to a forgetting mechanism active during output" (p. 285). Although this was the favoured interpretation of Cowan et al. for their backward recall data, it does not fit the present, forward recall data. The present Figure 1 shows that the word duration effect occurred throughout the list and was largest for the words recalled first. This result is more consistent with a mechanism involving rehearsal during the reception of the list, similar to what Baddeley (1986) originally proposed, possibly in conjunction with output effects.

Cowan et al. (1992, Experiment 3) used short- and long-word sets matched for the number of phonemes and syllables, and they obtained word length effects that were largest for the words recalled last in both forward and backward recall, unlike the present finding. However, it must be pointed out that Cowan et al. presented a cue to recall in the forward or backward order only after the list was presented. In experiments in which the subject knew that recall would proceed in the forward order (Cowan et al., 1992, Experiments 1 & 2), the word length effect was more ubiquitous across the list, as in the present data (except for Serial Position 1 of these experiments by Cowan et al., where a ceiling effect precluded a word length effect). The reason could be that participants are more likely to use articulatory rehearsal when they know that forward recall will be required.

Second, Service (1998) suggested that, inasmuch as performance was impaired for lists in which the halves received different timing cues, "at least re-programming output pace during recall is a secondary task that requires working memory capacity" (p. 285). The present data appear to weaken that statement inasmuch as very similar levels of performance were obtained for short–long, long–short, and long–long trials (all of which were inferior to performance on short–short trials). It is possible that, in the long–long trials, a disadvantage of having all long words (as opposed to only half long words) offset an advantage of not having to switch timing in the middle of the list. However, this possibility still includes the assumption that, all else being equal, subjects recall more when recalling short-duration words than when recalling long-duration words with the same phonological content.

Third, Service (1998) suggested that "Whether maintaining a single pronunciation speed also works as an additional interfering task is unclear at the moment. If it does, it can be expected to become effective over a time lag in a manner similar to the interfering task used in the Brown–Peterson paradigm. Thus, the results from an experiment based on subjects consciously controlling output durations still leave questions unanswered" (p. 285). We agree that this is an argument showing that we do not have evidence for a strong version of decay theory in which items fade from storage within a few seconds even if no mental events interfere. Time-based interference in which information is lost because other information is being rehearsed or spoken is subtly different from Baddeley's (1986) pure-decay description of forgetting. However, time-based interference could be substituted for pure decay with no further modification of the working-memory theory. Service's main premise, that no variety of time-based interference occurs except when time is an incidental correlate of phonological complexity, would require a more drastic modification in Baddeley's theory; but the present data contradict that suggestion, demonstrating a non-negligible time-based word length effect in addition to, and separable from, the phonological complexity effect.

Fourth, Service (1998) noted that in Cowan et al. (1997), "both items and pauses between them were over 80% longer in the long condition, yet the word-duration effect remained small, of the order of 5%. Thus, in their study, a much greater difference in output duration resulted in an effect of roughly the same size as the difference in recall between short and simple long items in this experiment [Service's Experiment 3] in which the recall times differed by less than 20%" (p. 297). Her comment was a comparison of a forward recall experiment (Service, 1998, Experiment 3) with a backward recall experiment (Cowan et al., 1997, Experiment 2). In the present forward recall data, however, the word length effect was an 8.5% effect overall and varied with specific conditions. For lists of monosyllabic words spoken in a short versus a long manner, it reached 21% in Serial Position 2, and it was larger than 10% in a number of specific conditions (as shown in Table 2). Thus, the very small magnitude of the duration-based word length effect should no longer be an issue. It probably does not account for all of the original word length effect noted by Baddeley et al. (1975), but it may well account for a non-negligible proportion of it.

It is not yet clear exactly why the present, substantial word length effects were obtained. It is relevant to keep in mind that the input and output phases of the trials were paced and that the word length effect was largest for the words presented and recalled first. If output delay or output interference were the reason for the effect, it should be largest for the words recalled last. Two possibilities arise that can be explained with reference to the phonological loop model of Baddeley (1986). One possibility is that participants form mental images of the items as they are spoken and that the imagined temporal durations of the words affect the duration of each rehearsal cycle. Another possibility is that words that are spoken more quickly leave more time for rehearsal during the silent periods between these words (in both the input and the output portions of the trial), given the paced nature of the task. In future work it might be possible to distinguish between these accounts if a satisfactory way could be found to manipulate the durations of words and yet eliminate both repetition by the participant during input and pacing during recall, as these affect the accounts differently.

One priority for future research is to explain a discrepancy in results obtained with the present type of procedure, in which word durations are manipulated through instructions, using forward recall versus backward recall. The present forward recall procedure resulted in a relatively large word duration effect and in an advantage for monosyllabic words, whereas the backward recall procedure of Cowan et al. (1997) resulted in a relatively small word duration effect and an advantage for disyllabic words. One possible explanation is that recall depends upon not only the processes described within the phonological loop theory of Baddeley (1986) but also the redintegration processes described by Schweickert (1993). In redintegration, information in long-term memory is used to identify a phonological memory trace that has become degraded through forgetting. Redintegration is thought to be easier to accomplish for phonologically longer (in the present terms, more complex) words because such words contain more phonological cues (Brown & Hulme, 1995; Cowan et al., 1997). It is quite possible that the pattern of recall depends on the balance between phonological loop and redintegration processes. Phonological loop processes could be more important in forward recall, producing the larger short-word advantage and the advantage for monosyllables found with that procedure in the present study; whereas redintegration processes could be more important in backward recall, producing the smaller short-word advantage and the advantage for disyllabic words found with that procedure by Cowan et al. (1997). The basic reason for this difference between the balance of processes in forward versus backward recall could be that the phonological trace of the sequence that is constructed by the participant during reception of the list can be used without transformation in forward recall only.

The aforementioned account of forward and backward recall appears to be at odds with the finding of Hulme et al. (1997) that word frequency effects were larger in forward recall than in backward recall, a finding that appears to suggest that redintegration is more important in forward recall. However, this finding of Hulme et al. is consistent with our suggested account if the success of redintegration depends on different factors in forward and backward recall. In forward recall, where the phonological trace can be used most easily, the success of redintegration may depend most critically on the ease of access to the underlying lexical representation (hence the word frequency effect of Hulme et al.). In contrast, in backward recall, the success of redintegration may depend more critically upon the cues provided by the phonological representation, given that it has to be used in a non-continuous manner (hence the advantage for disyllables obtained by Cowan et al., 1997). Research on this topic seems essential for a determination of how the phonological loop must be modified.

The present results also are consistent with the view that a portion of the word length effect is caused by output delay or output interference effects, as was found previously (Cowan et al., 1992, 1997). It is possible that output delay or interference always plays a role in the word length effect, but that covert articulatory rehearsal adds to such effects only under certain circumstances. Rehearsal may not occur if the participant knows that recall is to be backward (e.g., Cowan et al., 1997) or if the forward versus backward direction of recall is unknown to the participant until after the list is presented (e.g., Cowan et al., 1992, Experiment 3). Under such circumstances, one finds a relatively small word length effect limited to the items recalled last. In contrast, when the participant knows that recall will be in the forward direction (e.g., Cowan et al. 1992, Experiments 1 and 2, and in the present experiment), one finds a larger word length effect that is more ubiquitous across serial positions, presumably because of the contribution of rehearsal.

Perhaps the most important remaining puzzle is why duration-based effects were so small in the Finnish study of Service (1998), in contrast to the present study. Some factors could be interpreted to suggest that Service's procedure is the most appropriate, and other factors could be interpreted to suggest that the present procedure is the most appropriate. On one hand, Service (1998) suggested that the present requirement to recall words at a particular rate may act as a processing load that influences performance. The present task demands also require use of the temporal properties of the stimuli, which are self-produced acoustically. On the other hand, it seems possible that when spoken duration distinguishes between one phoneme and another, as in the Finnish lists that Service used, participants refrain from rehearsal because it would interfere with information about phonemic duration. That would not be the case in the present study because phonemic information is carried by sound qualities and remains independent of the spoken duration. Another difference between the studies is that the present study required mimicry of an absolute duration whereas Service's task only required the production of relatively short versus long vowel durations. There also could be English versus Finnish language differences that play a role. Thus, it is an open question whether the test procedure of Service (1998) or of the present study best matches what participants do in a typical serial recall task in English.

In sum, the present results do not negate the phonological complexity effects observed by Service (1998) and do not prove a pure-decay model of short-term memory; but they do rebut the notion that stimulus duration per se is unimportant in recall. Thus, although the present data point toward the need for revisions of the working-memory model of Baddeley (1986), such as the revisions offered by Brown and Hulme (1995), Cowan (1995), and Hulme et al. (1991), they do not warrant abandonment of a time-related forgetting proposed within that model.

REFERENCES

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

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). Modeling item length effects in memory span: No rehearsal needed? *Journal of Memory & Language*, *34*, 594–621.

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

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.

Cowan, N. (1995). Attention and memory: An integrated framework. New York: Oxford University Press.

Cowan, N., Day, L., Saults, J.S., Keller, 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., 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: Opposed effects of duration and complexity. *Psychological Science*, *8*, 290–295.

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

Hulme, C., Maughan, S., & Brown, G.D.A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory & Language*, *30*, 685–701.

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

Kucera, H., & Francis, W.N. (1967). Computational analysis of present day English. Providence, RI: Brown University Press.

Li, S.-C., & Lewandowsky, S. (1995). Forward and backward recall: Different retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 837–847.

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

Quinlan, P.T. (1992). *Oxford psycholinguistic database*. Oxford, UK: Oxford University Press.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, *21*, 168–175.

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.

Original manuscript received 2 November 1998
Accepted revision received 2 April 1999