THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2003, *56A* (8), 1307–1334

Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis

Richard Henson, Tom Hartley, and Neil Burgess

Institute of Cognitive Neuroscience, University College London, London, UK

Graham Hitch and Brenda Flude

Lancaster University, Lancaster, UK

Many recent computational models of verbal short-term memory postulate a separation between processes supporting memory for the identity of items and processes supporting memory for their serial order. Furthermore, some of these models assume that memory for serial order is supported by a timing signal. We report an attempt to find evidence for such a timing signal by comparing an "item probe" task, requiring memory for items, with a "list probe" task, requiring memory for serial order. Four experiments investigated effects of irrelevant speech, articulatory suppression, temporal grouping, and paced finger tapping on these two tasks. In Experiments 1 and 2, irrelevant speech and articulatory suppression had a greater detrimental effect on the list probe task than on the item probe task. Reaction time data indicated that the list probe task, but not the item probe task, induced serial rehearsal of items. Phonological similarity effects confirmed that both probe tasks induced phonological recoding of visual inputs. Experiment 3 showed that temporal grouping of items during list presentation improved performance on the list probe task more than on the item probe task. In Experiment 4, paced tapping had a greater detrimental effect on the list probe task than on the item probe task. However, there was no differential effect of whether tapping was to a simple or a complex rhythm. Overall, the data illustrate the utility of the item probe/list probe paradigm and provide support for models that assume memory for serial order and memory for items involve separate processes. Results are generally consistent with the timing-signal hypothesis but suggest further factors that need to be explored to distinguish it from other accounts.

Requests for reprints should be sent to Richard Henson, Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, UK. Email: r.henson@ucl.ac.uk

This work was supported by a BBSRC project grant to NB and GH. RH is supported by the Wellcome Trust. NB is supported by the Medical Research Council. We thank Nelson Cowan, Murray Maybery and an anonymous referee for their helpful comments.

Graham Hitch is now at the Department of Psychology, University of York, York YO10 5DD, UK.

http://www.tandf.co.uk/journals/pp/02724987.html DOI:10.1080/02724980244000747

Several computational models of verbal short-term memory have been developed recently (e.g., Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1992, 1999; Henson, 1998; Page & Norris, 1998). These models go beyond previous theories, such as the phonological loop (Baddeley, 1986; Baddeley & Hitch, 1974), by simulating serial position curves, error patterns, and effects of characteristic variables such as phonological similarity, word length, and presentation modality. Moreover, the models promote further theorizing and empirical predictions. For example, the models of Burgess and Hitch (1999) and Brown et al. (2000) postulate the existence of a timing signal that is used to represent serial order. These models assume that the coding of order information is separate from the coding of item information, consistent with previous research (e.g., Bjork & Healy, 1974; Healy, 1974; McNichol, 1971; Murdock, 1976). According to Burgess and Hitch (1999) and Brown et al. (2000), the timing signal derives from a set of internal oscillators and enables the coding of the positions of items within a sequence (for which there is considerable evidence, Henson, 1999a). More generally, the involvement of oscillators in the encoding and retrieval of verbal material can help to explain error patterns in tasks involving phonological output (Hartley, 1995, 2002; Hartley & Houghton, 1996; Vousden, Brown, & Harley, 2000). However, though the concept of a timing signal has proved useful in explaining some aspects of short-term memory, including effects of temporal grouping (Burgess & Hitch, 1996), temporal distinctiveness (Brown et al., 2000), and relative positional coding (Henson & Burgess, 1997), direct empirical evidence is lacking. We report a first attempt at finding such evidence, by looking for variables that affect the operation of the hypothetical timing signal. To support the claim that effects were specific to serial ordering, we devised two probed recall tasks that differed in the degree to which they required maintenance of serial order.

Item probe task

In the item probe (IP) task, participants see a list of items presented sequentially, followed by a single probe item, and judge whether or not the probe item was in the list (Figure 1a). This task was pioneered by Sternberg (1969), who suggested that performance was based on exhaustive serial scanning of the list items in search of the probe item. This conclusion was based chiefly on the finding that reaction times (RTs) increased linearly with list length. Subsequent data, however, have disputed this claim (e.g., Ashby, Tein, & Balakrishnan, 1993; Baddeley & Ecob, 1973; Monsell, 1978). More sophisticated analyses by McElree and Dosher (1989) showed that performance is better explained by direct access than by serial scanning. In fact, their favoured model of the IP task was one based on decaying "strengths" of item representations (e.g., Wickelgren & Norman, 1996), consistent with the recency effect found in accuracy and RT as a function of probe position. Providing presentation rates are reasonably fast and the retention interval is minimal (to minimize explicit rehearsal, Monsell, 1978), the IP task therefore seems likely to index short-term memory for item information in the absence of serial scanning or rehearsal.

List probe task

We considered various probe recognition tasks that might require access to serial order information. The most obvious is a relative order task (e.g., a probe "FK" prompting whether "F"

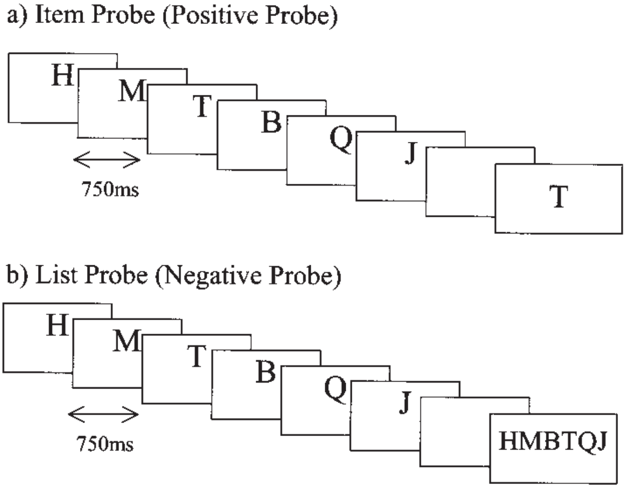

Figure 1. Schematic of (a) item probe (IP) and (b) list probe (LP) tasks.

and "K" were in the same or a different order in the sequence). However, there is evidence that this task can also be performed on the basis of the relative levels of item strength in memory, and possibly backward search, and so does not require maintenance of serial order per se (Hacker, 1980; McElree & Dosher, 1993). An alternative method is a position-item probe task (e.g., a probe "F3" prompting whether or not "F" occurred in the third position of the sequence). However, this task suffers from the additional processing requirement to decode the numeric representation of position. A more promising alternative is a list probe (LP) task, similar to tasks used previously by Allport (1984; see also Gathercole, Service, Hitch, Adams, & Martin, 1999).

In the LP task, a list is presented sequentially as in the IP task, but the probe is a second simultaneously presented list that participants judge as the same or different from the first (Figure 1b). The probe list always contains the same items as the original list and, when it differs, it does so only in the transposition of two adjacent items. We assumed that the probe format would encourage forward serial processing, such that participants would compare successive items in the probe against their memory for the list. Thus, whereas the IP task is primarily a test of item information, the LP task is primarily a test of order information (Murdock, 1976). Moreover, according to the "oscillator models" of Burgess and Hitch (1999) and Brown et al. (2000), the LP task should involve utilization of a timing signal whereas the IP task should not.

We have already used these tasks in a functional neuroimaging study in an attempt to isolate brain regions associated with storage, rehearsal, and temporal grouping in verbal short-term memory (Henson, Burgess, & Frith, 2000). Relative to the IP task, the LP task activated regions that included the left dorsal premotor cortex. Furthermore, a temporally grouped version of the LP task produced less activation in premotor cortex than an ungrouped version. Dorsal premotor cortex has been activated previously when comparing sequential with repetitive finger movements (Catalan, Honda, Weeks, Cohen, & Hallet, 1998) and damage to this region impairs reproduction of rhythmic motor sequences (Halsband, Ito, Tanji, & Freund, 1993). Left dorsal premotor cortex may also be recruited in the perception of rhythmic auditory stimuli whose temporal properties (amplitude modulation) are consistent with speech (Giraud et al., 2000). These imaging data are thus consistent with a role for left dorsal premotor cortex in the processing of serial order and rhythm, making this region a plausible site for the timing signal assumed by oscillator models.

Our assumption behind the present series of experiments was that any variable that affects the operation of the timing signal will have a larger effect on the LP than on the IP task. We chose to investigate a diverse set of variables that each possessed a temporal component while differing in other ways. These were irrelevant sound, articulatory suppression, temporal grouping, and rhythmic finger tapping. If a common timing signal is responsible for maintaining memory for serial order, we would expect to find a similar pattern of effects across all four variables. Experiment 1 examined the effect of irrelevant sound.

EXPERIMENT 1
Irrelevant sound

The presence of background sound, which people are told to ignore, can nonetheless impair short-term memory. This effect holds whether the sound is irrelevant speech (Salamé & Baddeley, 1982), or even simple tones, provided the sound exhibits some degree of change over time (Jones & Macken, 1995). One argument for this "irrelevant sound effect" reflecting more than simple attentional distraction, and the reason for its present interest, is that the effect is generally found only for tasks that require memory for serial order. For example, Salamé and Baddeley (1990) found an irrelevant sound effect with serial recall, but not free recall, and Jones and Macken (1993) found the effect with serial recall, but not with a missing-item task. For this reason, proposals have been made that irrelevant sound interferes with order information, seen variously as the formation of inter-object links (Jones, 1993), coding of the position of items with a sequence (Henson, 1998), or the relative order of item strengths in memory (Page & Norris, 1998). Alternatively, in the present context, irrelevant sound might disrupt a timing signal.

Nonetheless, LeCompte (1996) recently reported effects of irrelevant sound in recognition, paired-associate learning, and free recall: tasks that prima facie do not require maintenance of serial order. This observation is consistent with the view that irrelevant sound alters representations of item content but does not affect information about serial order, as in Neath's feature model (Neath, 2000). However, though the instructions for the tasks used by LeCompte (1996) do not require attention to serial order, participants may nonetheless rehearse items serially, in an attempt to aid retention. When Beaman and Jones (1997) minimized the use of such rehearsal strategies by requiring participants to suppress articulation, irrelevant sound had little effect on recognition or paired-associate learning. However, one problem with the tasks used by LeCompte, Beaman, Jones, and colleagues is that they differ along several dimensions besides the degree of seriality involved, thus precluding direct comparisons. Because the present IP and LP tasks involve the same method of list presentation and yes–no responding, performance levels are commensurable, and the specificity of the irrelevant sound effect to serial order can be tested directly via a task by irrelevant-sound interaction.

Method

Participants

The 20 volunteers replying to adverts in the UCL Psychology Department consisted of 12 men and 8 women (mean age of 27.6 years), and they were paid for participating.

Materials

Lists of 5, 6, or 7 letters were generated by random selection without replacement from the consonants BDGPTVHMQRYZ. Half of the items were phonologically confusable (BDGPTV), and half were phonologically nonconfusable (HMQRYZ). Four blocks of 60 lists were constructed for each participant, with 16 lists of five items, 20 lists of six items, and 24 lists of seven items.

Half of the lists were followed by a positive probe and half by a negative probe. For the IP task, a positive probe was an item from the list, whereas a negative probe was a vocabulary item that was not in the list. Each position 1 to $N - 1$ in lists of N items was probed positively twice. For the LP task, a positive probe was a probe list that matched the stimulus list in order, whereas a negative probe was a probe list that did not. Each position 1 to $N - 1$ was probed negatively twice, the probe for position i being a probe list in which item i and item i +1 were transposed.

Stimuli and responses were controlled by an IBM PC, using software written by R. Henson and S. Zielinski at the UCL Psychology Department. The irrelevant speech consisted of 10 sentences, spoken by different American voices, obtained from the Haskins Laboratory webpage (with kind permission of R. Remez). The sentences were repeated in fixed order and played through headphones. The volume was set reasonably loud, though participants could reduce the volume if they found it uncomfortable. The precise volume was not thought relevant, given that sound levels have no detectable effect on the magnitude of the irrelevant sound effect (Tremblay & Jones, 1999).

Procedure

Each consonant was presented in the centre of a VDU, 0.5 inches high, at a rate of one every 750 ms (500 ms on, 250 ms off). Participants were instructed to read each letter in silence. Each list began with a warning signal ("!") and was followed by a recall cue ("?") presented in the same manner as the letters. After the cue, the probe appeared centre-screen, where it remained until participants pressed one of two keys marked "y" and "n". Trials were self-paced by the space-bar. Instructions emphasized both speed and accuracy. In the LP task, participants were also told that, if the probe list differed from the stimulus list, it would only ever differ in the order of two adjacent items. In the irrelevant-sound conditions, participants were instructed to ignore the background speech and were reassured that their memory for the speech would not be tested.

The order of lists within blocks was randomized, such that list length on any given trial was unpredictable. This was to discourage participants from grouping the items (Henson, 1996; see also Experiment 3). The order of irrelevant sound conditions was counterbalanced across participants, with the constraint that the first and second, and third and fourth blocks involved the same probe task (to minimize confusion between tasks). Participants received 10 practice trials, and the experiment took approximately 30 minutes.

Data analysis

Serial position effects were analysed in repeated measures analyses of variance (ANOVAs) on the arcsin of proportion correct and the logarithm of RT. RTs outside the range 200–9000 ms were removed from analyses; missing RT values were set to the mean for that cell. Nonsphericity in ANOVAs was accommodated by Greenhouse–Geisser correction.

Results

The proportions of correct responses in each condition are shown in the upper panel of Figure 2 (chance = .5). Though irrelevant sound impaired performance on both tasks, the impairment on the LP task ($M = 0.08$, $SD = 0.09$) was greater than that on the IP task ($M = 0.02$, $SD = 0.03$). Indeed, a 2 (irrelevant sound) × 2 (probe task) × 2 (task order) ANOVA showed a significant interaction between irrelevant sound and probe task, $F(1, 18) = 8.97, p < .01$. No effects of task order approached significance, Fs $< 1.43, p > .25$.

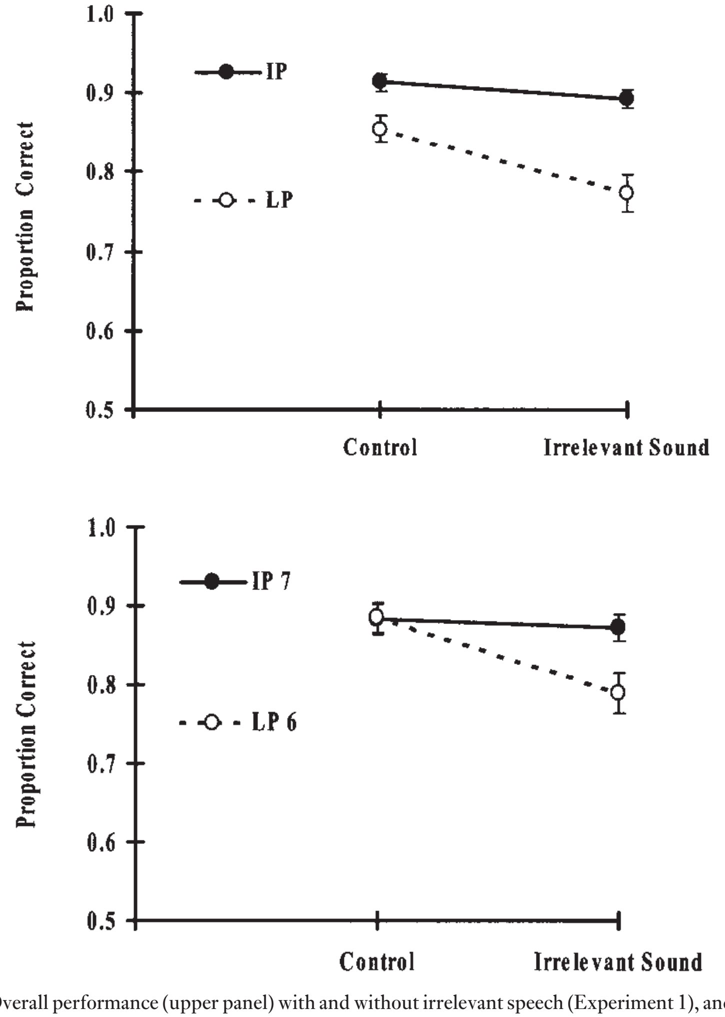

Figure 2. Overall performance (upper panel) with and without irrelevant speech (Experiment 1), and performance when control performance in IP and LP tasks equated by different list lengths (lower panel). Error bars show standard error of mean (between participants).

Post hoc *t*-tests showed significant effects of irrelevant sound on both the IP task, $t(20) = 2.77$, $p < .05$, and the LP task, $t(20) = 3.95$, $p < .01$.

One might argue that the above interaction reflects a ceiling effect, given that control performance in the IP task was over 90% correct and in the LP task (85%) was significantly less, $t(20) = 3.81$, $p < .01$. This argument was examined by comparing six-item lists in the LP task with seven-item lists in the IP task (Figure 2, lower panel). Control performance on the two tasks no longer differed significantly, $t(20) = 0.02$, $p = .98$, and yet a two-way ANOVA still showed an interaction between probe task and irrelevant sound, $F(1, 19) = 9.54$, $p < .01$.

Phonological similarity and probe type effects

To test whether both probe tasks were accessing phonological short-term memory, performance on confusable and nonconfusable probes was compared. For the IP task, a confusable probe was one in which there was at least one other item in the list that was phonologically similar to the probe (see Materials section). For the LP task, a confusable probe was one in which the transposed letters were phonologically similar to each other (see Figure 3). Approximately one half of IP trials and one quarter of LP trials contained a confusable probe. Analyses were performed on the arcsin transform of the proportions of confusable and nonconfusable trials that had correct responses. It is important to emphasize that interest was focused on phonological similarity effects within each task. Given the different definitions of similarity, it is not meaningful to compare effects across tasks.

The proportions of responses for confusable and nonconfusable positive and negative probes are shown in Figure 4. In the IP task, a small disruptive effect of phonological similarity

Probe Type	Probe Conf.	Example List	Example Probe
	IP Task		
Pos	Con	H T B M	T?
Pos	Non	H T Q M	Q?
Neg	Con	H T Q M	V?
Neg	Non	H T B M	Q?
	LP Task		
Pos	Con	H T B M	H T B M?
Pos	Non	H T Q M	H T Q M?
Neg	Con	H T B M	H B T M?
Neg	Non	H T Q M	H Q T M?

Figure 3. Example probe types and probe confusabilities. Pos = positive probe; Neg = negative probe; Con = confusable probe; Non = nonconfusable probe.

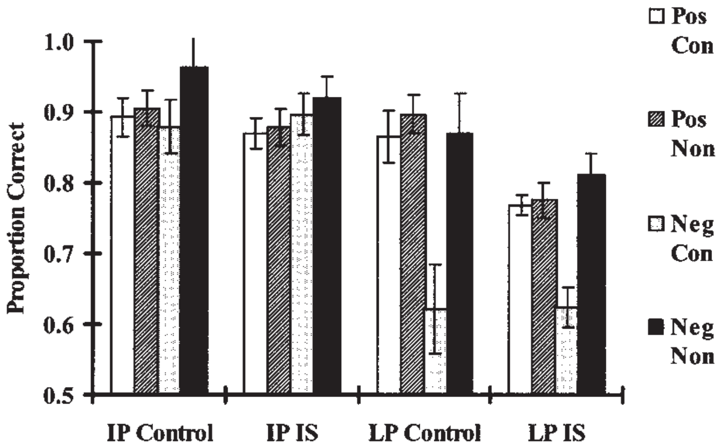

Figure 4. Accuracy as a function of probe type and probe confusability with and without irrelevant speech (Experiment 1). IS = Irrelevant sound. See Figure 3 for more details.

was apparent. This was confirmed by a 2 (irrelevant sound) × 2 (probe type) × 2 (probe confusability) ANOVA, which showed a significant effect of probe confusability, $F(1, 19) = 7.79, p < .05$. The phonological similarity effect appeared larger for negative than for positive probes, though the interaction only approached significance, $F(1, 19) = 3.21, p = .09$.

For the LP task, phonological similarity exerted a large effect in both control and irrelevant-sound conditions, again mainly by reducing performance on negative probes. Indeed, an ANOVA showed a significant interaction between probe confusability and probe type, $F(1, 19) = 18.76, p < .001$. In addition, the interaction between probe type and irrelevant sound approached significance, $F(1, 19) = 3.88, p = .06$, suggesting that irrelevant sound produced a greater disruption for positive than for negative probes.

Serial position effects

Initial examination of serial position curves for IP and LP tasks did not suggest any interaction between probe position and irrelevant sound. Data were therefore collapsed across the irrelevant-sound manipulation in order to increase the numbers of observations.

Percentage correct as a function of probe position for each probe task and list length is shown in the upper panel of Figure 5. For the IP task, Positions 1–6 represent positions of positive probes, and Position 0 refers to negative probes. For the LP task, Positions 1–6 represent positions of negative probes, and Position 0 refers to positive probes. Data from Position 0 were excluded from all analyses of position effects. The IP task showed recency and primacy effects, whereas performance in the LP task generally decreased with probe position. This was confirmed by significant interactions between probe task and probe position in seven-item lists, $F(4.1, 77.7) = 2.43$, $MSE = 0.16$, $p = .05$, and six-item lists, $F(3.1, 58.8) = 3.87$, $MSE = 0.15$, $p < .05$, though not five-item lists, $F(2.6, 49.5) = 1.54$, $MSE = 0.13$, $p = .22$ (the latter probably reflecting ceiling effects).

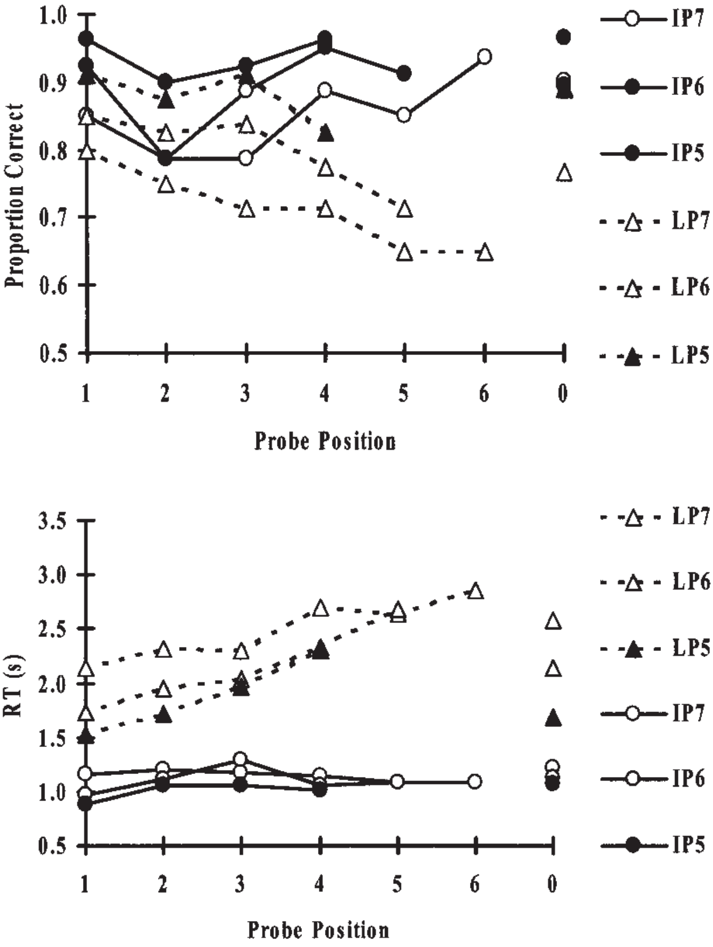

Figure 5. Accuracy and reaction times (RT) as a function of probe position for each probe task and list length collapsed across irrelevant speech (Experiment 1). Position 0 reflects negative probes for the IP task and positive probes for the LP task.

The mean of the median correct RTs across participants is shown as a function of probe position in the lower panel of Figure 5. Responses in the LP task were generally slower than those in the IP task, but whereas the IP task showed little effect of probe position, the LP task showed a general increase in RT across Positions 1–6. This was again confirmed by significant interactions between probe task and probe position, this time for all list lengths, *F*s > 5.15, *MSE* < 0.01, $p < .001$. The coefficients of the best fitting linear regressions of RT on probe position for list lengths 5–7 were 259, 230, and 141 ms/position, respectively, for the LP task. RTs to positive probes (Position 0) in the LP task were approximately equal to the average RTs for negative probes.

Discussion

There was a greater detrimental effect of background speech on the LP task than on the IP task, as predicted by the hypothesis that irrelevant sound is particularly disruptive of tasks that require maintenance of serial order. These results bolster those of previous studies (Beaman & Jones, 1997; Jones & Macken, 1993; Salamé & Baddeley, 1990), and go beyond them by demonstrating a significant interaction between irrelevant-sound and memory tasks that were matched closely in respects other than the requirement for serial order. The observation that both tasks were performed less accurately when probes were confusable provides useful confirmation that they were accessing phonological short-term memory. The finding that the additional errors induced by background speech in the LP task were mainly to positive probes suggests that irrelevant sound increases transpositions in the order of items in memory. Such errors would cause participants to make incorrect "no" responses to list probes that nonetheless matched the original sequence, whereas errors that happened to counteract the transposition in a negative list probe would be far less likely.

Independent support for our hypothesis that the LP and IP tasks differ in their serial order requirements comes from the analysis of RTs as a function of probe position. The RT–position function in the IP task was generally flat, consistent with direct-access theories of this task (e.g., McElree & Dosher, 1989). In the LP task, the RT–position function had an average slope of approximately 210 ms/item, which is close to the rate of subvocal rehearsal for familiar monosyllables (Baddeley, 1986). It is also close to the rate found in Sternberg's successor naming task (Sternberg, 1967), another task believed to engage serial rehearsal. Furthermore, the slope is too slow to be attributed to the visual scanning of list probes from left to right, as visual search rates for lists of letters are of the order of 50 ms/item (Bisanz & Resnick, 1978; Pashler & Badgio, 1985). Taken together, these RT analyses provide an independent measure of the degree of serial rehearsal in each task, addressing the problem raised by Beaman and Jones (1997).

It is important to note that, in addition to the interaction between irrelevant sound and probe task, there was a small but significant residual effect of background speech on the IP task. One interpretation is that irrelevant sound has more than one interfering effect: a general distraction of attention, which affects both the IP and LP tasks, and a specific disruption of serial ordering, which affects only the LP task. An effect of irrelevant sound on any task requiring attention and memory is consistent with LeCompte (1996) and Neath (2000).

One puzzle concerning RTs in the LP task is that mean correct RT for positive probes (Probe Position 0) was faster than RTs for later negative probe positions. If the LP task is performed via serial rehearsal from the start of the sequence, participants should only be able to respond "yes" to a positive probe when they have finished rehearsing the whole sequence. That is, such responses should be at least as slow as responses to a negative probe involving transposition of the last two items. One possible explanation is that participants have a tendency to guess "yes" when unsure (consistent with the greater incidence of "yes" responses in the LP task). If guesses were relatively fast, then correct responses to positive probes would be shorter than those predicted by the serial rehearsal hypothesis.

In summary, comparison of performance on the IP and LP tasks supports the view that irrelevant speech is primarily disruptive of memory for serial order, though it may also have a weaker effect on memory for items. The LP and IP tasks were also differentiated by RT data consistent with the use of subvocal rehearsal in retrieving serial order information.

EXPERIMENT 2
Concurrent articulatory suppression

In order to assess whether subvocalization plays a different role in the LP and IP tasks, the interference manipulation in Experiment 2 was the presence or absence of articulatory suppression. Concurrent repetition of an irrelevant vocalization has long been known to impair verbal short-term memory (Murray, 1967), and its interaction with the word length effect (Baddeley, Thomson, & Buchanan, 1975) has been interpreted as its prevention of subvocal rehearsal. Furthermore, changing-state articulation (e.g., "one–two–three–four–one . . .") produces a greater impairment than repetitive articulation (e.g., "the–the–the–the–the . . ."; Macken & Jones, 1995). This effect of changing state parallels that found with irrelevant sound (Jones & Macken, 1995). If the LP task differs from the IP task in requiring subvocalization to maintain and retrieve information about serial order, it should be more sensitive to disruption from articulatory suppression.

Method

Participants

The 16 participants replying to adverts in the UCL Psychology Department consisted of 7 men and 9 women (mean age of 26.5 years), and they were paid for participating.

Materials and procedure

The same materials were used as those in Experiment 1, with participants attempting one probe task per session, consisting of two blocks, one with suppression and one without. Given the tiring nature of articulatory suppression, the two sessions were held on separate days. This also reduced the likelihood of participants using the same strategies (e.g., serial rehearsal) for the two probe tasks. The order of probe tasks and the order of suppression conditions were counterbalanced across participants.

In the suppression condition, participants were instructed to repeat the sequence "one–two–three–four" as quickly as possible to themselves, loudly enough to hear their own voice. They were instructed to continue suppression throughout the block of trials, though if they wanted a rest, they could pause between the self-paced trials.

Results

Suppression caused considerable impairment on both tasks, but the impairment was greater on the LP task ($M = 0.17$, $SD = 0.11$) than on the IP task ($M = 0.09$, $SD = 0.08$; upper panel of Figure 6). A 2 (suppression) × 2 (probe task) × 2 (session order) ANOVA showed a significant interaction between suppression and probe task, $F(1, 14) = 9.84$, $p < .01$. However, there was also a significant effect of session order, $F(1, 14) = 6.29$, $p < .05$, and the three-way interaction approached significance, $F(1, 14) = 4.48$, $p = .05$. There was a general improvement in performance in the second session, and the interaction between suppression and probe task was more noticeable in the second session than in the first. It seems that combining tasks was difficult and that a certain amount of practice may increase selective interference effects.

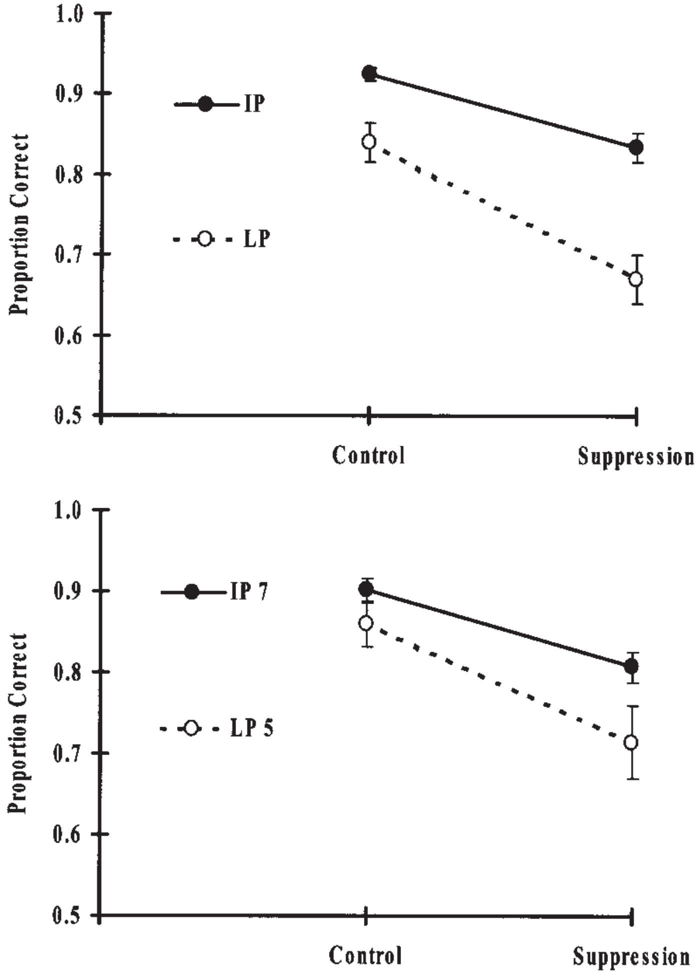

Figure 6. Overall performance (upper panel) with and without articulatory suppression (Experiment 2) and when control performance in IP and LP tasks equated (lower panel). See Figure 2 for more details.

As in Experiment 1, one might argue that the Interference × Probe Task interaction reflects a ceiling effect, given that control performance in the IP task is over 90% correct and significantly greater than that in the LP task, $t(16) = 4.11$, $p < .001$. Performance on the two tasks was better matched by comparing seven-item lists in the IP task with five-item lists in the LP task (lower panel of Figure 6), such that control performance on the tasks no longer differed significantly, $t(16) = 1.31$, $p = .21$. Both tasks still showed a significant effect of suppression, $t(16) > 4.48$, $p < .001$, and the interaction between probe task and suppression approached significance in a two-way ANOVA, $F(1, 15) = 4.00$, $p = .06$. Given that a one-tailed criterion is more appropriate here, we conclude that the corresponding interaction in the main analysis was not due to a ceiling effect.

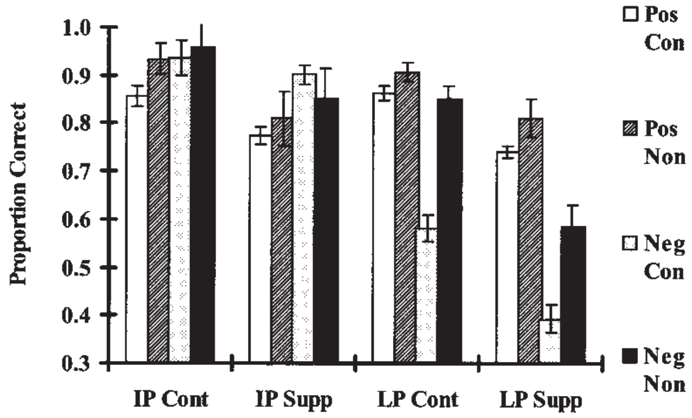

Figure 7. Accuracy as a function of probe type and probe confusability with and without articulatory suppression (Experiment 2). Cont = control; Supp = suppression. See Figure 4 for more details.

Phonological similarity and probe type effects

The proportions of correct responses to confusable and nonconfusable positive and negative probes (as designated in Experiment 1) are shown in Figure 7. In the IP task, the effect of probe confusability was greater in the control condition than in the suppression condition. Indeed, a 2 (suppression condition) × 2 (probe type) × 2 (probe confusability) ANOVA showed a significant interaction between suppression and probe confusability, $F(1, 15) = 6.94$, $p < .05$. Pairwise tests showed a significant effect of confusability in the control, $t(16) = 4.74$, $p < .001$, but not in the suppression condition, $t(16) = 0.33$ (collapsing across probe type in both cases). These data are consistent with the proposal that articulatory suppression prevents phonological recoding (Baddeley, 1986). There was also a main effect of probe type, $F(1, 15) = 15.4$, $MSE = 0.07$, $p < .001$, with negative probes producing higher accuracy.

For the LP task, phonological similarity reduced correct responses to negative probes in both control and suppression conditions. The was confirmed by a significant interaction between probe confusability and probe type, $F(1, 15) = 5.87$, $p < .05$. Unlike the IP task, however, suppression did not interact with probe confusability. Indeed, an effect of probe confusability remained under suppression, $t(16) = 4.63$, $p < .001$ (collapsing across probe type).

Serial position effects

Due to relatively small numbers of observations, data were collapsed across list length. The upper panel of Figure 8 shows accuracy as a function of probe position and probe task. Interpretation of serial position effects is compromised by the fact that Positions 1–4 are collapsed over all list lengths, whereas Position 5 is only collapsed over list lengths 6 and 7, and Position 6 is only from list length 7. However, interest concerns interactions with probe position, which are not compromised. A 2 (probe task) × 2 (suppression condition) × 6 (probe position) ANOVA revealed a significant interaction between probe task and probe position,

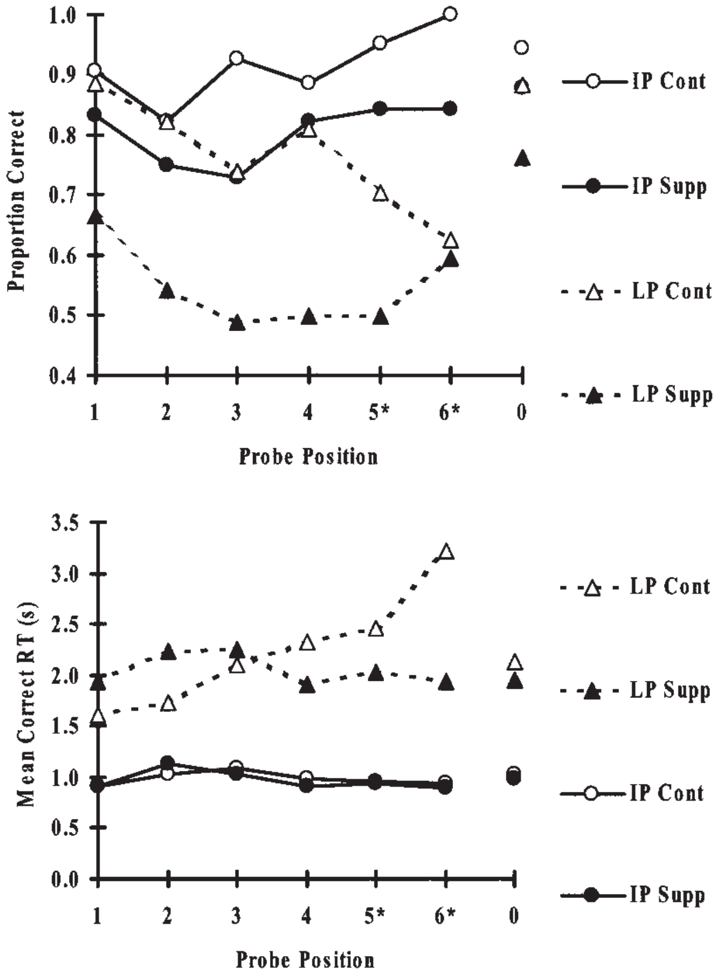

Figure 8. Accuracy and reaction times (RT) as a function of probe position with and without articulatory suppression (Experiment 2). Cont = control; Supp = suppression. Positions marked with an asterix are derived from fewer lists (see text). See Figure 5 for more details.

$F(3.45, 51.78) = 3.57$, $MSE = 0.14$, $p < .01$, reflecting greater recency in the IP than LP task, as in Experiment 1. No other interactions reached significance, however, suggesting that suppression had a uniform effect over probe position.

Reaction times in the control condition of the LP task increased monotonically with probe position, as in Experiment 1 (see lower panel of Figure 8). However, this was not true in the suppression condition, suggesting that suppression was effective in preventing serial rehearsal. This was confirmed by a significant three-way interaction between probe task, suppression, and probe position, $F(3.05, 45.8) = 6.82$, $MSE = 0.01$, $p < .001$, indicating a greater interaction between suppression and probe position in the LP task than in the IP task. Indeed, suppression appeared to have little effect on the RT functions in the IP task.

Discussion

The present experiment supported the hypothesis that the LP task involves a greater degree of serial rehearsal than does the IP task. Concurrent articulatory suppression, which is assumed to prevent rehearsal (e.g., Baddeley, 1986), had a larger detrimental effect on the LP than on the IP task. Independent support for the assumption that suppression prevented subvocal rehearsal was obtained from the finding that it flattened RT profiles as a function of probe position in the LP task. It is interesting to note the sharp contrast between this latter effect of suppression and the absence of a corresponding effect of irrelevant speech on RTs in Experiment 1. We suggest that while both manipulations degrade memory for serial order, irrelevant speech does not prevent participants from using subvocalization to respond to list probes because, unlike suppression, it does not capture the articulatory system.

If suppression prevents serial rehearsal, how can people perform the LP task above chance under suppression? One possibility is that sequential presentation of list items gives rise to a visuospatially organized orthographic representation of the list, as well as a temporally organized phonological representation (see, e.g., Logie, Della Sala, Wynn, & Baddeley, 2000). Under normal conditions, people may compare the visuospatial representation against the list probe, in parallel. Only if a mismatch is detected might they compare the phonological representation to the list probe via sequential, subvocal articulatory rehearsal. (This is another potential explanation for why "yes" responses are faster than "no" responses for later probe positions; see Discussion to Experiment 1.) Under suppression, however, people must rely on the visuospatial representation, reducing but not abolishing performance. The parallel nature of the visuospatial comparison would then explain the flatter probe position curves and faster RTs (at least for later probe positions).

Suppression also exerted a significant detrimental effect on the IP task. This may be a consequence of articulatory suppression preventing phonological recoding of the list items and probe (Baddeley, 1986). This interpretation receives some support from the fact that suppression removed the effect of phonological similarity in the IP task. In this case, correct performance of the IP task under suppression might be attributable to a form of visual memory like that described above. However, unlike the IP task, the LP task continued to show an effect of phonological similarity under suppression. This suggests that suppression may not have prevented phonological recoding (or serial rehearsal) completely—for example, participants may have attempted to recode list items in between irrelevant articulations. Thus, interpretation of data obtained with suppression is complicated by the fact that it may have multiple interference effects. A more selective interference with a timing signal was attempted in Experiment 3, by introducing temporal grouping into list presentation.

EXPERIMENT 3
Temporal grouping

Temporal grouping of a sequence by the insertion of a pause every few items is well known to improve serial recall (e.g., Ryan, 1969). This temporal grouping effect is independent of word length and phonological similarity (Hitch, Burgess, Towse, & Culpin, 1996). Temporal grouping is also independent of articulatory suppression when the list items are presented auditorily, but not when presented visually (Hitch et al., 1996). According to several models

(Brown et al., 2000; Burgess & Hitch, 1999; Henson & Burgess, 1997; Hitch et al., 1996), such grouping effects reflect a change in the nature of the timing signal underlying serial recall. More specifically, grouping results in a differentiation of the timing signal into two components: one tracking the timing of items within groups and one tracking the timing of items (or groups) within lists. If memory for serial order depends on such a timing signal, the LP task should show better performance for grouped than for ungrouped lists. To the extent that the IP task does not depend on memory for serial order, its performance should be relatively insensitive to the temporal rhythm of list presentation.

Method

Participants

The 18 volunteers from Lancaster University, 6 male and 12 female, were paid to participate in the experiment.

Materials

A total of 120 lists of consonants were constructed using the stimuli described in Experiment 1. Of these, 60 contained six items for presentation as sequences of two groups of three items. The remaining 60 lists were assigned for ungrouped presentation. Of these, 15 lists contained five items, 30 contained six items, and 15 contained seven items. The order of list lengths within the ungrouped set was randomized in order to discourage the use of subjective grouping. Selection of positive and negative probes for the IP task and LP task were as described previously.

Timing of stimuli was as that in Experiment 1 for the ungrouped condition. For grouped presentation there was a 750-ms pause corresponding to the presentation of an extra "blank" item between Items 3 and 4. Participants were told to use the pause to group the letters in threes in the same way that they might do for a telephone number.

Procedure

Presentation of stimuli and instructions were as those in Experiment 1 except that trials were continuous rather than self-paced. All participants were given the ungrouped lists in the first of two test sessions, with half performing the IP task first and half performing the LP task first. Participants returned for their second session some hours or days later, when they were given the grouped lists. The order of probe tasks was again counterbalanced. Grouped lists were not presented in the first session in order to avoid unwanted transfer of grouping strategies to the ungrouped lists. While this meant that any improvement associated with grouping may have reflected general practice effects, we were primarily interested in the interaction between grouping and task. Each session lasted for about 50 minutes.

Results

Overall performance

To facilitate comparisons, data from only the six-item lists in the ungrouped condition are reported. The proportions of correct responses for grouped and ungrouped list presentation are shown in Figure 9. Grouping improved performance by .03 (SD = 0.08) in the IP task, and by .09 (SD = 0.09) in the LP task. A 2 (grouping) × 2 (probe task) ANOVA showed significant effects of grouping, $F(1, 17) = 27.03$, $p < .001$, and probe task, $F(1, 17) = 12.88$, $p < .05$. The

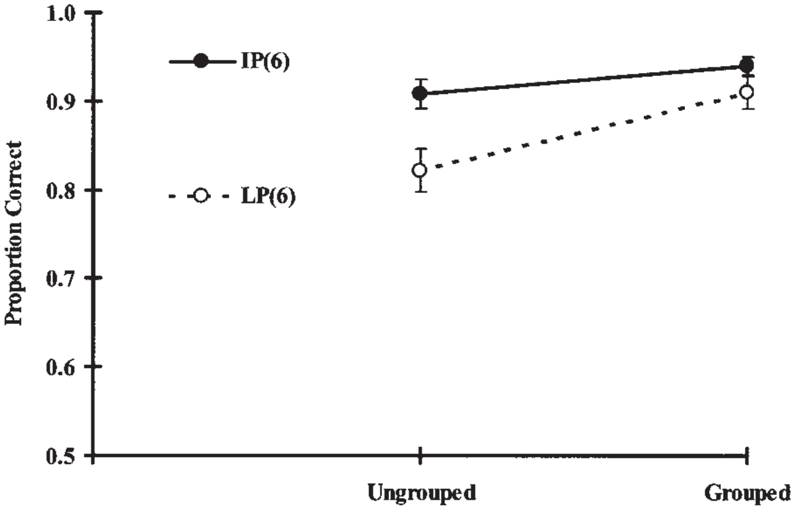

Figure 9. Overall performance with and without temporal grouping for six-item lists (Experiment 2). See Figure 2 for more details.

interaction between grouping and probe task approached significance, $F(1, 17) = 3.23$, $p = .09$, and was reliable when considered as a one-tailed test of the predicted interaction. Post hoc *t*-tests showed a significant effect of grouping on the LP task, $t(18) = 4.04$, $p < .001$, and not the IP task, $t(18) = 1.61$, $p = .13$. However, interpretation of the differences between the probe tasks is compromised by a possible ceiling effect in the IP task.

Serial position effects

Accuracy as a function of probe position for each condition is shown in the upper panel of Figure 10. Most noticeable is the greater accuracy for Position 3 than Positions 2 or 4 in both ungrouped and grouped conditions of the LP task. This pattern suggests that participants were spontaneously subjectively grouping the ungrouped lists. A 2 (grouping) × 2 (probe task) × 5 (probe position) ANOVA showed a significant effect of grouping, $F(1, 17) = 5.21$, $MSE = 0.25$, $p < .05$, a significant interaction between probe task and probe position, $F(3.13, 53.22) = 3.01$, $MSE = 0.54$, $p < .05$, and an interaction between grouping and probe task that again approached significance, $F(1, 17) = 2.22$, $MSE = 0.57$, $p = .06$.

As before, RTs in the LP task generally increased with probe position (lower panel of Figure 10), but with a marked deviation from monotonicity on Position 3 (like the accuracy data). Grouping tended to reduce the overall gradient of RTs against probe position. Grouping had less obvious effects on RTs in the IP task. This greater effect of grouping on LP than on IP probe position curves was confirmed by a significant three-way interaction between grouping, probe task, and probe position, $F(3.10, 52.78) = 3.11$, $MSE = 0.15$, $p < .05$. Two-way ANOVAs on the LP and IP tasks separately confirmed an interaction between grouping and probe position on the LP task, $F(2.93, 49.89) = 2.74$, $MSE = 0.26$, $p = .05$, but not in the IP task, $F(2.52, 42.77) = 1.14$, $MSE = 0.05$, $p = .34$.

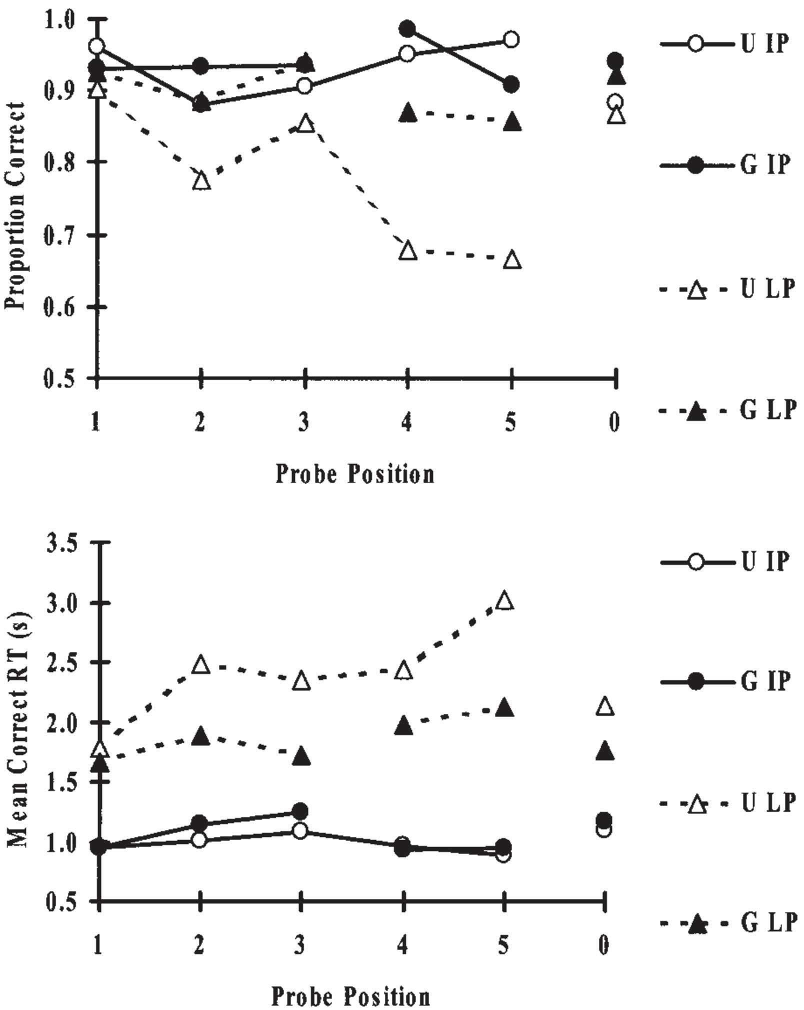

Figure 10. Accuracy and reaction times (RT) as a function of probe position with and without temporal grouping (Experiment 3). G = grouped; U = ungrouped. See Figure 5 for more details.

Discussion

The present experiment provides support for the hypothesis that temporal grouping primarily improves short-term memory for serial order. Although the interaction between grouping and probe task on overall performance may have been confounded by ceiling effects, accuracy and RTs as a function of probe position confirm that grouping exerted a greater influence on performance of the LP than on that of the IP task. Grouping increased performance and lowered RTs on most if not all positions in the LP task, particularly positions in the second group. However, grouping had no reliable effects on probe position curves in the IP task.

Both grouped and ungrouped probe position curves in the LP task showed improved performance on Position 3 relative to Position 2. This "mini-recency" effect at the end of the groups is indicative of grouping strategies (Ryan, 1969). The fact that such evidence was apparent even in the ungrouped condition suggests that participants were spontaneously grouping in threes (the modal group size, Henson, 1996), despite our precautions to vary list length unpredictably in the ungrouped condition and to avoid carry-over effects by testing the ungrouped condition first. The presence of subjective grouping in our ungrouped condition would have weakened the interaction between grouping and task. This is a recurring problem with (visual) grouping manipulations (Henson, 1996).

The mini-recency effect on Position 3 in the LP task represented faster and more accurate responding to negative list probes in which the paired transposition straddled a group boundary. Such deviations of the grouped structure may be particularly easy to detect, improving accuracy. However, it is difficult to explain why such responses are faster than those on the preceding position according to a strict, forward rehearsal strategy from the start of the list. One possibility is that participants have direct access to different groups and sometimes initiated their rehearsal from the start of the second group. In this case, they could immediately notice an erroneous item at the start of the second group when the transposition straddled the group boundary, thus producing shorter RTs on Position 3 than on Position 2. This strategy would result in a shallower gradient across probe position for grouped than for ungrouped lists, consistent with the interaction pattern between grouping and probe position in the LP task.

EXPERIMENT 4
Finger tapping

Experiment 4 examined the relative interference on IP and LP tasks of a concurrent tapping task, following the suggestion that such rhythmic production tasks might impair the encoding of serial order in short-term memory by competing for a common timing signal (Burgess & Hitch, 1999). This suggestion is consistent with the imaging study of Henson et al. (2000), in which a dorsal premotor brain region was differentially activated as a function of serial rehearsal and temporal grouping in STM, given that the same brain region has been implicated in rhythmic motor finger movements by imaging and neuropsychological studies (Catalan et al., 1998; Halsband et al., 1993). If the same timing signal is responsible for serial rehearsal and rhythmic tapping, then the LP task should show a greater detrimental effect of concurrent tapping than the IP task. Furthermore, by increasing the temporal complexity of the tapping task, from simple repetitive tapping to a more complex "changing-state" rhythm (Jones & Macken, 1995), we expected to see an increase in the amount of interference.

Method

Participants

The 18 volunteers replying to adverts in the UCL Psychology Department consisted of 9 men and 9 women (mean age of 28.1 years), and they were paid for participating.

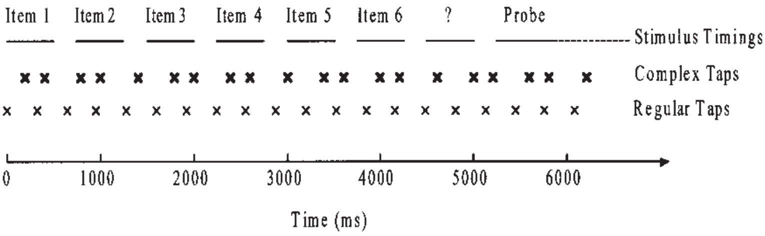

Figure 11. Schematic of relative timing of probe tasks and regular and complex tapping in Experiment 4.

Materials and procedure

Six sets of 30 lists were created in the same manner as that in Experiment 1. The IP and LP tasks were combined with one of three tapping conditions—no tapping, regular tapping, and complex tapping—performed by each participant in six separate blocks. The three blocks for each probe task were presented contiguously, with the order of tapping conditions counterbalanced across participants. Half the participants performed the LP blocks first, and half performed the IP blocks first.

In the regular and complex tapping tasks, participants pressed the spacebar with their nondominant hand in synchrony with a computer-generated tone (responses to the probe tasks being made with their dominant hand). The tones had a duration of 200 ms and pitch of 2 kHz. In the regular tapping task, tones were produced at 320-ms intervals. In the complex tapping task, tones were produced in a syncopated rhythm like the one used by Saito (1994). The relative timing of tones and items in the probe tasks varied due to their different frequencies, as shown in Figure 11. Tones began a few seconds before the first stimulus appeared and continued uninterrupted throughout the block.

Baseline measures of performance in both tapping tasks were obtained for approximately 90 s at the beginning and end of the experiment. Participants were given 10 practice trials at each probe task in the absence of concurrent tapping before receiving the experimental blocks involving that task. Three covert practice trials were added to the beginning of each block to ensure that participants had become accustomed to combining the tapping and probe tasks. The only other procedural difference from Experiment 1 was that trials were not self-paced: After recording a response to the probe task, the next trial was delayed until the beginning of the next two bar cycle of the tone sequence. The experiment lasted approximately 1 hr 15 min.

Results

Overall performance

Concurrent tapping impaired performance of both IP and LP tasks (upper panel, Figure 12), more so when tapping a complex than simple rhythm (impairment by regular and complex tapping on IP task, $M = 0.03$, $SD = 0.10$ and $M = 0.11$, $SD = 0.09$, respectively, and on LP task, $M = 0.07$, $SD = 0.07$ and $M = 0.16$, $SD = 0.09$, respectively). This was confirmed by a 2 (probe task) × 3 (tapping task) ANOVA showing significant effects of probe task, $F(1, 17) = 20.89$, $MSE = 0.01$, $p < .001$, and tapping condition, $F(1.97, 33.35) = 26.28$, $MSE = 0.01$, $p < .001$. The interaction between probe task and tapping approached significance, $F(1.83, 31.07) = 3.07$, $MSE = 0.005$, $p < .06$. A planned comparison between performance with tapping, (complex + regular)/2, versus no tapping revealed a significant interaction with probe task, $F(1, 17) = 4.61$, $p < .05$. However, the difference between complex and regular tapping did not interact with the probe task

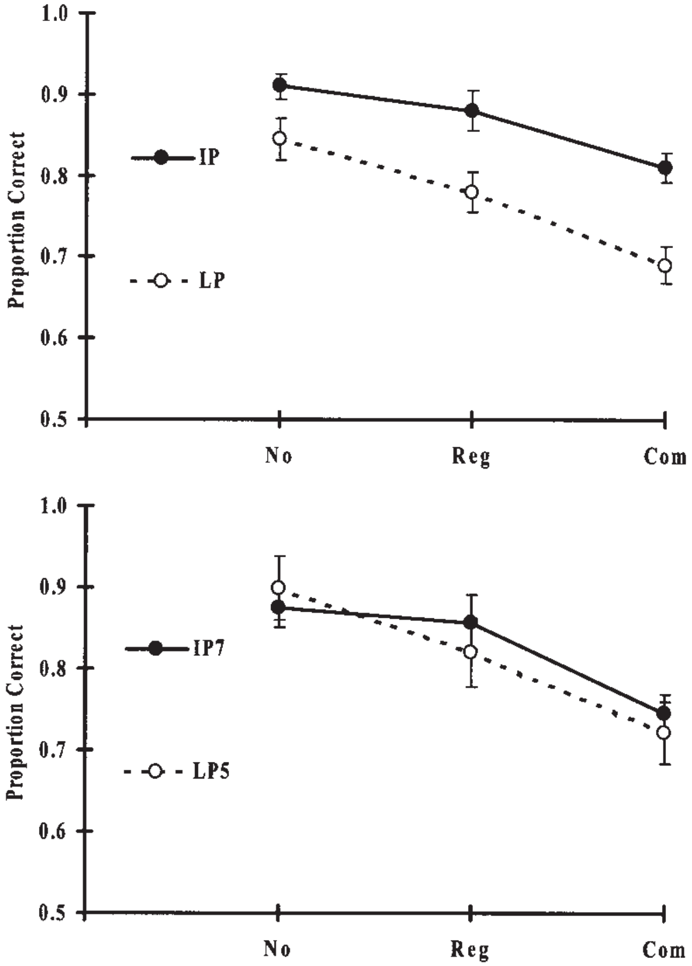

Figure 12. Overall performance (upper panel) with no tapping, regular tapping, and complex tapping (Experiment 4), and when control performance in IP and LP tasks equated (lower panel). No = no tapping; Reg = regular tapping; Com = complex tapping. See Figure 2 for more details.

($F < 1$). Thus tapping caused a significantly greater impairment on the LP than on the IP task, but there was no evidence that this impairment differed for regular versus complex tapping.

When performance of the LP and IP tasks in the no tapping condition was equated by comparing seven-item IP lists with five-item lists (lower panel, Figure 12), the interaction between probe and tapping tasks did not reach significance, $F(1.77, 30.17) < 1$, and nor did either of the planned comparisons, $F(1, 17) < 2.14$, $p > .16$ (though this might reflect reduced power, given the smaller sample sizes). Thus it remains possible that the above interaction between probe task and the presence of tapping reflects a range effect, given the overall differences in performances of the two probe tasks.

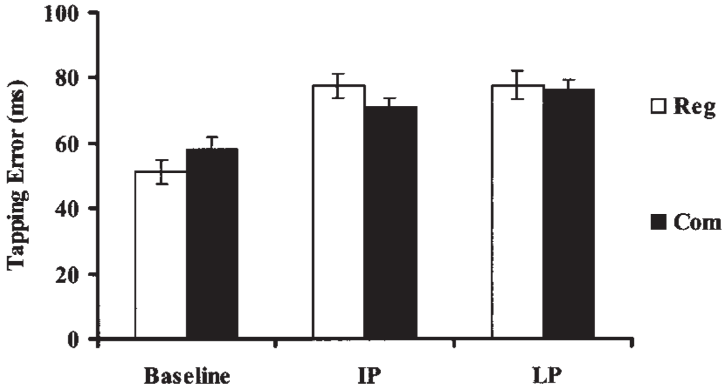

Figure 13. Tapping error (as variability of tone–tap offset; see text) during baseline and during probe tasks in Experiment 4. Reg = regular tapping; Com = complex tapping.

Tapping performance was analysed to see whether interference between tapping and memory affected the secondary rather than the primary task. Performance was indexed by matching the timing of each tap to the nearest tone and calculating the mean and the standard deviation of the offsets. These two measures were highly correlated ($r > .6$, $p < .01$), indicating that when performance was impaired, both accuracy and variability were affected (participants did not, for example, tap the correct rhythm with a constant lag). The following analyses are confined to the variability measure (Figure 13; note that this measure is not directly comparable across the regular and complex tapping tasks). A 3 (IP vs. LP vs. baseline) × 2 (regular vs. complex tapping) ANOVA showed a significant interaction between probe task and tapping task, $F(1.51, 25.75) = 7.91$, $MSE = 168.82$, $p < .01$. However, this interaction resulted from the difference between baseline performance and performance during the two probe tasks: namely, a greater increase in tapping variability from baseline to dual-task conditions for regular than for complex tapping. A further 2 × 2 ANOVA confined to tapping performance during the two probe tasks confirmed there was no interaction, $F(1, 17) = 2.37$, $MSE = 55.88$, $p = .14$. Thus, there was no evidence of an interaction between tapping and type of probe task in the tapping performance.

Phonological similarity and probe type effects

Performance on the IP task was relatively unaffected by the phonological confusability of the probe (Figure 14). A 3 (tapping task) × 2 (probe type) × 2 (probe confusability) ANOVA showed a significant effect of tapping task, $F(1.93, 32.87) = 14.62$, $MSE = 0.10$, $p < .001$, but no other effects or interactions ($F < 2.2$, $p > .12$). A phonological similarity effect was observed in the LP task, however, for which the corresponding analysis showed a significant effect of tapping task, $F(2.0, 33.9) = 12.07$, $MSE = 0.10$, $p < .001$, and a significant interaction between probe type and probe confusability, $F(1, 17) = 11.46$, $MSE = 0.15$, $p < .01$. As in Experiments 1 and 2, the effect of probe confusability in the LP task was associated with negative list probes.

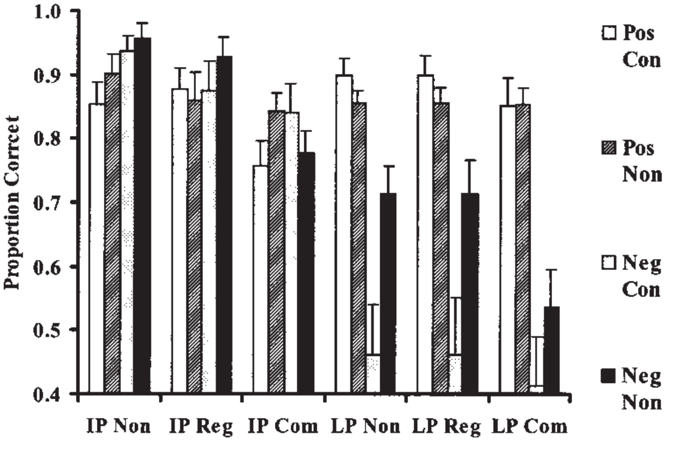

Figure 14. Accuracy as a function of probe type and probe confusability for no, regular, and complex tapping (Experiment 4). Non = no tapping; Reg = regular tapping; Com = complex tapping. See Figure 4 for more details.

Serial position effects

Serial position functions for accuracy and RTs were similar to those reported in the previous experiments. As there were no significant interactions involving either probe task or tapping conditions, they are not reported.

Discussion

The present experiment provided limited support for a disruptive effect of concurrent tapping on short-term memory for serial order. Thus, although tapping produced a greater impairment on the LP than on the IP task, we could not rule out range effects in this interaction. Moreover, though tapping a complex rhythm produced a greater overall impairment than tapping a regular, isochronous rhythm, we found no differential effect of the type of tapping on performance in the IP and LP tasks, or on the variability of tapping error during the IP and LP tasks. The lack of interaction with type of tapping is contrary to what might be expected if the LP and tapping tasks competed for a common timing signal, on the assumption that complex tapping makes greater demands on such a timing signal than does regular tapping.

Regular tapping has not generally been thought to interfere with short-term memory, often being used as a nonphonological control task (Baddeley, 1986). In these situations, however, the tapping has usually been self-paced, and it is possible that participants can adjust their rate of tapping to (a harmonic of) the presentation or rehearsal rate. The present data show that regular tapping can interfere with short-term memory when externally paced at a different (nonharmonic) rate. However, part of this impairment may reflect the general demands of dual-versus single-tasking. The even greater impairment produced by complex tapping on both probe tasks confirmed that participants found the complex rhythm more challenging than the regular tapping. Indeed, one possible reason for the lack of interaction between probe task and type of tapping is that complex tapping was simply too difficult to combine successfully with either task.

Unlike Saito (1994), we did not find that complex tapping interacted with the phonological similarity effect, at least in the LP task. One reason may be a lack of dynamic range in performance of the LP task (given that performance for confusable probes was close to chance); another reason may be that participants were able to switch temporarily between the memory and tapping tasks (akin to the explanation given in Experiment 2 for the residual phonological similarity effects in the LP task under concurrent articulatory suppression).

A recent study by Burle and Bonnet (2000) found an interfering effect of repetitive, auditory clicks on performance of an IP task, which was selective to certain click rates. These authors agued that the critical rate was that corresponding to the gamma rhythm in the human brain (as proposed by Treisman, Faulkner, Naish, & Brogan, 1990), and that click rates close to this frequency interact with the temporal oscillators assumed by the model of Lisman and Idiart (1995) to underlie serial scanning in an IP task. However, these frequencies (around 20 Hz) are an order of magnitude higher than the present rates of tone presentation (approximately 2 Hz), and the rate of assumed serial scanning (which is still a matter of debate; see Introduction) is much faster than the rate of serial rehearsal assumed in the models of short-term memory considered here. Thus it may be unwise to compare data and models across these different timescales.

GENERAL DISCUSSION

The present study introduced a novel means to test factors that selectively interfere with short-term maintenance of serial order, by comparing two probe tasks—the item probe (IP) and the list probe (LP)—that differed principally in their serial order requirements. The assumption that only the LP task requires serial rehearsal was confirmed by analyses of RTs, which showed a monotonic increase with probe position for the LP task but not the IP task. The two tasks also showed different patterns of accuracy as a function of probe position, with the LP task being dominated by a large primacy effect, and the IP task showing mainly a recency effect. Both tasks showed effects of phonological similarity, confirming that they engage phonological short-term memory (Baddeley, 1986). The IP task produced slightly better overall performance than the LP task, but performance could be reasonably well equated by titrating the different list lengths used for each task. Given that both tasks involve the same list presentation and yes/no response requirement, we propose that using them in tandem will prove valuable for investigating the mechanisms hypothesized to underlie short-term memory for serial order. However, it is important to emphasize that neither task is necessarily "pure". Thus, as already noted, participants may be able to make some use of item information to answer questions about serial order and vice versa.

Nonetheless, we found that several experimental manipulations produced greater interference with the LP than with the IP task, providing evidence that the maintenance of serial order is indeed a separable process. This evidence is consistent with a range of different assumptions about the way serial order is separately coded (Brown et al., 2000; Burgess & Hitch, 1992, 1999; Henson, 1998; Lee & Estes, 1981). We concentrated here on the hypothesis that memory for serial order is subserved by a timing signal derived from temporal oscillators, as assumed by the computational models of Burgess and Hitch (1999) and Brown et al. (2000).

Thus the factors examined—irrelevant sound, articulatory suppression, temporal grouping, and rhythmic tapping—were selected on the basis that they have a temporal component that might engage the same timing signal and hence cause a greater effect on the LP than on the IP task.

Experiments 1 and 2 confirmed that irrelevant sound and articulatory suppression do have a greater detrimental effect on the LP than on the IP task, suggesting that these manipulations are particularly disruptive of the mechanisms underlying short-term memory for serial order. Experiment 3 showed that temporal grouping produces a greater improvement in performance on the LP than on the IP task, while Experiment 4 indicated that concurrent tapping to unsynchronized tones produces a greater impairment in the LP than in the IP task, although the possibility of a range effect limits the impact of this latter observation. Thus, despite considerable disparities among the factors examined, a common pattern has emerged that is broadly consistent with predictions from the hypothesis that order information is mediated by a timing signal. However, more detailed considerations suggest it would be premature to take the present results as unequivocal evidence for such a signal.

For example, the greater relative interference of irrelevant sound and articulatory suppression on the LP task is consistent with other explanations, such as the disruption of inter-object links (Jones, 1993), the prevention of phonological recoding by articulatory suppression (Baddeley, 1986), or forcing the use of visuospatial representations (see Experiment 2, Discussion). Experiments 1 and 2 did not attempt to distinguish such possibilities. Indeed, one suggestion is that both factors might simply impair subvocal articulation. However, the present results make this particular possibility seem unlikely. Whereas RT functions indicate that suppression disrupted speech-based serial processing in the LP task, no corresponding effect was observed with irrelevant sound. To explain this difference, it seems necessary to assume that irrelevant sound disrupted order information without disrupting subvocal articulation. Nevertheless, it should be clear that neither the effect of irrelevant sound nor that of suppression requires that the disruption of order information is necessarily mediated by a timing signal.

We would argue that the selective improvement from temporal grouping in the LP task is more specific evidence for the timing-signal hypothesis. It was notable that grouping exerted a clear effect on the probe position functions of the LP task in both RTs and accuracy, but had little effect on the corresponding functions in the IP task. Temporal grouping effects have a natural explanation in terms of recruitment of oscillators with different frequencies (Brown et al., 2000; Burgess & Hitch, 1999; Henson & Burgess, 1997; Hitch et al., 1996). Furthermore, the only obvious change between ungrouped and grouped versions of the LP tasks in Experiment 3 was the timing of presentation and response, with little apparent change in the amount of subvocal articulation required.

Selective interference from finger tapping on the LP task would also comprise more specific evidence for the timing-signal hypothesis. However, the results of Experiment 4 provide only limited support for this. Thus, although tapping interfered with the LP task more than the IP task, we could not rule out range effects in this case. Moreover, there was no support for the prediction that complex ("changing-state") tapping produces a greater decrement on the LP task than on the IP task, as would be expected if tapping a more complex rhythm made greater demands on the same timing signal. One explanation of our results is that the motor task relies on an entirely separate timing mechanism (although this would not be consistent with neuroimaging data suggesting that left dorsolateral premotor activation is common to grouped probe tasks and rhythmic finger tapping, see Henson et al., 2000). Alternatively, speech and nonspeech tasks may share a common timing signal composed of multiple oscillators (as in the OSCAR model for example, Brown et al., 2000). If this were the case, the degree to which the secondary task disrupts serial memory depends not on its complexity, but on the degree to which some oscillators are required for both tasks. The selective interference caused by irrelevant sound and articulatory suppression might then be the result of a range of different oscillators being recruited by these stimuli/tasks, including the critical oscillators used for the encoding/rehearsal of serial order. The two tapping tasks, although differing in complexity, are matched for mean rate and would thus recruit rather similar sets of oscillators. They might therefore be roughly equivalent in their capacity to interfere with the oscillators used for remembering the sequence. In support of this, we note that there are independent theoretical grounds for believing that multiple oscillators are required for the timing signal: (1) to account for temporal grouping effects (see above); (2) to encode sequences where the rate of presentation may vary (e.g., in continuous speech, Hartley, 2002; Vousden, Brown, & Harley, 2000); and (3) to account for cases where position appears to be coded relative to both the start and the end of a sequence (Henson, 1999b; Henson & Burgess, 1997).

CONCLUSION

The main purpose of the present study was to introduce two probe tasks that allow one to test for factors that selectively affect short-term memory for serial order. We have shown how the combination of both tasks can be applied to test one particular theory of serial order—namely, that serial order is represented by a timing signal derived from a set of temporal oscillators. We found limited support for this theory, in that (part of) the effects of irrelevant sound, articulatory suppression, and grouping may reflect modulation of this timing signal. Further factors need to be examined, however, in order to distinguish the timing-signal account from other theories, and in particular to examine the interfering effects of nonspeech tasks with a temporal/rhythmic component.

REFERENCES

Allport, D. A. (1984). Auditory–verbal short-term memory and conduction aphasia. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 313–326). Hove, UK: Lawrence Erlbaum Associates Ltd.

Ashby, F. G., Tein, J. Y., & Balakrishnan, J. D. (1993). Response time distributions in memory scanning. *Journal of Mathematical Psychology*, *37*, 526–555.

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

Baddeley, A. D., & Ecob, J. R. (1973). Reaction time and short-term memory: Implications of repetition effects for the highspeed exhaustive scanning hypothesis. *Quarterly Journal of Experimental Psychology*, *25*, 229–240.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *Recent Advances in the Psychology of Learning and Motivation*, *1*, 47–90.

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.

Beaman, C. P., & Jones, D. (1997). The role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *23*, 459–471.

Bisanz, J., & Resnick, L. B. (1978). Changes with age in two components of visual search speed. *Journal of Experimental Child Psychology*, *25*, 129–142.

Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior*, *13*, 80–97.

Brown, G. D., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*, 127–181.

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language*, *31*, 429–460.

Burgess, N., & Hitch, G. J. (1996). A connectionist model of STM for serial order. In S. Gathercole (Ed.), *Models of short-term memory* (pp. 51–71). Hove, UK: Psychology Press.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*, 551–581.

Burle, B., & Bonnet, M. (2000). High-speed memory scanning: A behavioural argument for a serial oscillatory model. *Cognitive Brain Research*, *9*, 327–337.

Catalan, M. J., Honda, M., Weeks, R. A., Cohen, L. G., & Hallett, M. (1998). The functional neuroanatomy of simple and complex sequential finger movements: A PET study. *Brain*, *121*, 253–264.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology*, *13*, 65–77.

Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnrude, I., Frackowiak, R., & Kleinschmidt, A. (2000). Representation of the temporal envelope of sounds in the human brain. *Journal of Neurophysiology*, *84*, 1588–1598.

Hacker, M. J. (1980). Speed and accuracy of recency judgements for events in short-term memory. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 651–675.

Halsband, U., Ito, N., Tanji, J., & Freund, H. J. (1993). The role of premotor cortex and the supplementary motor area in the temporal control of movement in man. *Brain*, *116*, 243–266.

Hartley, T. (1995). *The role of syllable structure in verbal short-term memory*. PhD Thesis, University of London, UK.

Hartley, T. (2002). Syllabic phase: A bottom-up representation of the temporal structure of speech. In J. A. Bullinaria & W. Lowe (Eds.), *7th Neural Computation and Psychology Workshop*. Singapore: World Scientific.

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*, *31*, 1–31.

Healy, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 644–655.

Henson, R. N. A. (1996). *Short-term memory for serial order*. Unpublished PhD thesis, University of Cambridge, UK.

Henson, R. N. A. (1998). Short-term memory for serial order: The Start–End Model. *Cognitive Psychology*, *36*, 73–137.

Henson, R. N. A. (1999a). Coding position in short-term memory. *International Journal of Psychology*, *34*, 403–409.

Henson, R. N. A. (1999b). Positional information in short-term memory: Relative or absolute? *Memory & Cognition*, *27*, 915–27.

Henson, R. N. A., & Burgess, N. (1997). Representations of serial order. In J. A. Bullinaria, D. W. Glasspool, & G. Houghton (Eds.), *4th Neural Computation and Psychology Workshop* (pp. 283–300). London: Springer.

Henson, R. N. A., Burgess, N., & Frith, C. D. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: An fMRI study. *Neuropsychologia*, *38*, 426–440.

Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *Quarterly Journal of Experimental Psychology*, *49A*, 116–139.

Jones, D. (1993). Objects, streams and threads of auditory attention. In A. Baddeley & L. Weiskrantz (Eds.), *Attention, awareness and control* (pp. 87–104). Oxford: Oxford University Press.

Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect—implication for phonological coding. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*, 369–381.

Jones, D. M., & Macken, W. J. (1995). Organizational factors in the effect of irrelevant speech: The role of spatial location and timing. *Memory & Cognition*, *23*, 192–200.

Lee, C. L., & Estes, W. F. (1981). Item and order information in short-term memory: Evidence for a multi-level perturbation process. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 149–169.

LeCompte, D. (1996). Irrelevant speech, serial rehearsal, and temporal distinctiveness: A new approach to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 1154–1165.

Lisman, J. E., & Idiart, M. A. P. (1995). Storage of 7±2 short-term memories in oscillatory subcycles. *Science*, *267*, 1512–1515.

Logie, R. H., Della Sala, S., Wynn, V., & Baddeley, A. D. (2000). Visual similarity effects in immediate verbal serial recall. *Quarterly Journal of Experimental Psychology*, *54A*, 626–646.

Macken, W. J., & Jones, D. M. (1995). Functional characteristics of the inner voice and the inner ear: Single or double agency? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 436–448.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General*, *118*, 346–373.

McElree, B., & Dosher, B. A. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General*, *112*, 291–315.

McNicol, D. (1971). The origins of transpositions in short-term memory. *Australian Journal of Psychology*, *23*, 9–17.

Monsell, S. (1978). Recency, immediate recognition memory and reaction time. *Cognitive Psychology*, *10*, 465–501.

Murdock, B. B. (1976). Item and order information in short-term serial memory. *Journal of Experimental Psychology: General*, *105*, 191–216.

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology*, *21*, 263–276.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, *7*, 403–423.

Page, M. P. A., & Norris, D. G. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*, 761–781.

Pashler, H., & Badgio, P. C. (1985). Visual attention and stimulus identification. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 105–121.

Ryan, J. (1969). Grouping and short-term memory: Different means and patterns of grouping. *The Quarterly Journal of Experimental Psychology*, *21*, 137–147.

Saito, S. (1994). What effect can rhythmic finger tapping have on the phonological similarity effect? *Memory & Cognition*, *22*, 181–187.

Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for structure of working memory. *Journal of Verbal Learning and Verbal Behaviour*, *21*, 150–184.

Salamé, P., & Baddeley, A. D. (1990). The effects of irrelevant speech on immediate free recall. *Bulletin of the Psychonomic Society*, *28*, 540–542.

Sternberg, S. (1967). Retrieval of contextual information from memory. *Psychonomic Science*, *8*, 55–56.

Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction time experiments. *American Scientist*, *57*, 421–457.

Treisman, M., Faulkner, A., Naish, P. L. N., & Brogan, D. (1990). The internal clock: Evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. *Perception*, *19*, 705–743.

Tremblay, S., & Jones, D. M. (1999). Change in intensity fails to produce an irrelevant speech effect: Implications for the representation of unattended sound. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1005–1015.

Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology*, *41*, 101–175.

Wickelgren, W. A., & Norman, D. A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, *3*, 316–347.

Original manuscript received 7 June 2001
Accepted revision received 5 August 2002