Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source?

The disruption of immediate serial verbal recall by irrelevant and unattended speech in a foreign language was first demonstrated by Colle and Welsh (1976), an effect that was not readily attributable to simple distraction since no impairment was observed when speech was replaced by noise, nor was the intensity of the sound an important variable (Colle, 1980). The
effect was further investigated in a series of experiments by Salamé and Baddeley, who showed
that digit recall was as much disrupted by nonsense syllables as by meaningful words, or
indeed by the presentation of other digits, which the subject attempted to ignore (Salamé &
Baddeley, 1982). They also found no effect of sound intensity (Salamé & Baddeley, 1987), or
of noise that was modulated so as to produce the same sound intensity envelope as the speech
(Salamé & Baddeley, 1989). However, speech itself was not necessary, since disruption was
also obtained from instrumental music and from sounds whose frequency fluctuated in the
same manner as the irrelevant speech, but with intensity held constant (Salamé, 1990).

Salamé and Baddeley (1982) offered an interpretation of the effect in terms of the phono-
logical loop component of working memory (Baddeley & Hitch, 1974). It was suggested that
the irrelevant speech gained access to the phonological store. The initial suggestion was that its
effect operated through some form of “mnemonic masking”. Some evidence for this view was
offered by an incidental feature of the study suggesting the apparent importance of phonolog-
ical features—namely, the observation that disyllabic irrelevant words appeared to disrupt the
recall of monosyllabic digits slightly but significantly less than monosyllabic words. This
speculation continues to be cited as a major prediction of the model. However, this interpreta-
tion was questioned by a subsequent study that combined irrelevant speech with acoustic
similarity within the material to be recalled. It was predicted that phonologically similar items,
having fewer distinguishing cues, would be more subject to disruption by irrelevant speech
(Salamé & Baddeley, 1986). The two variables were studied across a range of sequence lengths;
additive effects occurred across moderate lengths, whereas effects tended to disappear with
long sequences, suggesting that subjects were abandoning phonological coding, as has been
found in other studies where sequence length grossly exceeds span (Hall, Wilson,
Humphreys, Tinzmann, & Bower, 1983). There was, however, no support for the simple
phonological mnemonic masking hypothesis.

The precise interpretation of the irrelevant speech effect was therefore left open, as indeed
were a number of features of the phonological loop, including, notably, the process whereby
order is stored. Irrelevant speech was, however, assumed to influence the phonological store
rather than to disrupt item perception (Baddeley & Salamé, 1986). The fact that articulatory
suppression eliminated the irrelevant speech effect with visually presented material was
consistent with the assumption that suppression prevented the material from entering the
phonological store. Hanley and Broadbent (1987) found that, with auditory presentation,
which is assumed to guarantee access to the phonological store, articulatory suppression did
not eliminate the irrelevant speech effect, provided that the material did not grossly exceed
span. Hanley and Bakopoulou (2003) replicated and extended this work, finding an irrelevant
speech effect when participants performed articulatory suppression regardless of whether the
irrelevant speech and articulatory suppression occurred while the lists were presented or only
during the retention interval. This latter finding clearly shows the irrelevant speech effect
observed under articulatory suppression was not due to a masking of the auditory stimulus.
The additive effects that Hanley and Bakopoulou found for irrelevant speech and articulatory
suppression during auditory presentation were also observed with visually presented lists
when participants were specifically instructed to encode the letters phonologically. Participants
instructed to code the letters semantically outperformed the phonological encoding
group, suggesting that semantic encoding may be a better strategy when the task becomes very
difficult.
As the Salamé and Baddeley programme of research was coming to an end, a very active series of further studies was initiated by Jones and colleagues. Their work has greatly extended knowledge of those factors that determine whether an irrelevant sound will or will not disrupt immediate recall, demonstrating that an effect can be produced by simple tones, provided they fluctuate, whereas a single tone has no effect (Jones & Macken, 1993; Jones, Macken, & Murray, 1993). An extension of this work also showed that repeatedly hearing a single token such as a letter has little or no effect on recall, whereas repeating two or more tokens clearly disrupts performance, with an increase in the number of tokens beyond two having relatively little further effect (Jones, Macken, & Mosdell, 1997; Tremblay, Macken, & Jones, 2000). Jones accounts for this and many other studies in terms of the changing state hypothesis, which proposes that only sounds that fluctuate are capable of disrupting recall, although this does not apply to noise that fluctuates in intensity, or to the fluctuation that occurs when a single item is repeated. An important feature of this approach, however, has been the way it provides a link to factors that determine auditory perception, and in particular auditory scene analysis, offering a potentially valuable bridge between short-term memory (STM) and perception (Jones, 1993; Jones & Macken, 1993; Jones & Tremblay, 2000).

A more detailed account of the process of disruption is offered by the Object-Oriented Episodic Record (O-OER) hypothesis (Jones, 1993). This proposes that the storage of the serial order of items can be conceptualized in terms of a series of pointers displayed on a multimodal surface. Irrelevant sounds that change state are themselves assumed to have a sequential structure, which then disrupts the existing cues to serial order, leading in due course to impaired recall. An important feature of this approach is its assumption that the memory representation is multimodal, based on the finding that visuo-spatial as well as verbal memory may be disrupted by irrelevant sounds (Jones, Farrand, Stuart, & Morris, 1995), although this result does not appear to be as readily replicable as most irrelevant speech effects.

A third theoretical approach to the irrelevant speech effect has been proposed by Neath (2000), who applies the feature hypothesis to the task of explaining the irrelevant speech effect. The feature hypothesis (Nairne, 1990; Neath & Nairne, 1995) assumes that memory items are stored as vectors of features, which may be modality dependent or modality independent. Memory failures are the result of these features being degraded through overwriting by new information. Neath (2000) offers an interpretation of the phenomena of irrelevant speech, together with other features attributed by Baddeley (1986) to the phonological loop. Like Jones, Neath suggests that irrelevant speech and articulatory suppression reflect a common underlying process of disruption, with the greater effect of articulatory suppression principally representing a greater attentional demand.

An even more recent interpretation of the irrelevant speech effect has been offered by Norris, Page, and Baddeley (2000) in terms of the Page and Norris (1998) computational model of the phonological loop, the primacy model. This assumes that serial order is stored in terms of trace strength, with the initial item in a sequence having the strongest encoding, and successive items being less strongly represented. Retrieval involves a competitive queuing process whereby the strongest representation is selected, emitted, and then inhibited, followed by the next strongest and so forth. The model avoids the crucial difficulties encountered by earlier chaining hypotheses of serial order (Baddeley, 1968; Henson, Norris, Page, & Baddeley, 1996) and in doing so proposes a two-component process whereby the items and information regarding their order are stored separately. Irrelevant speech is assumed to add
“noise” to the primacy-based process underlying serial recall, hence the susceptibility of order rather than item information to the influence of irrelevant speech (Beaman & Jones, 1997; Jones, 1993; Morris & Jones, 1990). The primacy model is one of a number of models that are based on the concept of a phonological loop (Burgess & Hitch, 1992, 1999; Henson, 1998), all of which maintain a distinction between a phonological storage system and a quasi-articulatory rehearsal process, a distinction that is also reflected in the neuropsychological and neuroimaging literature (Paulesu, Frith, & Frackowiak, 1993; Smith & Jonides, 1996).

Before concluding this brief review, one further recent development should be noted. Saito (1993) observed that syncopated manual tapping disrupted concurrent verbal serial recall and eliminated the acoustic similarity effect when recalling letters, although equal interval tapping had no such effect. He suggests that syncopated tapping and speech production mutually interfere because both rely on a common basic timing system (Saito, 1994).

Although the irrelevant speech effect has been studied over many years, until recently, at any given time, research tended to be confined largely to a single group, initially that of Colle (1980) and Colle and Welsh (1976), followed by Salamé and Baddeley (1982, 1986, 1987, 1989), with the paradigm then being carried forward by Jones and colleagues (for example, Jones & Macken, 1993, 1995; Jones, Macken, & Mosdell, 1997; Jones, Macken, & Murray, 1993; Jones, Madden, & Miles, 1992), and most recently by a number of groups including Neath (2000) and his associates. The different groups tend to use slightly different techniques. Salamé and Baddeley typically used immediate recall while Jones and colleagues tend to interpolate a rehearsal period between presentation and recall. Both groups used sequences of length just above span, while Neath and colleagues, coming from a background that is more influenced by long-term verbal memory, tend to use longer sequences of items, which then run the risk of encouraging the subjects to abandon the use of phonological coding (see Baddeley, 2000b, for discussion).

Given the development of interest in modelling the irrelevant speech effect, it is becoming increasingly important to establish which phenomena are robust, and which are less so. The sequence of three experiments that follow attempts to bridge some of these gaps by studying a range of variables, each of which has previously formed the focus of one particular group. The experiments do not aim to test precise predictions of any of the theories; indeed we doubt at this point whether such predictions are possible, given the relatively early stage of development of a number of the models. We hope, however, that the three studies will provide replication of basic phenomena and, perhaps equally important, evidence as to how the various effects interact. Given that the models have been created with the basic phenomena in mind, we suggest that it is these more subtle interactions that are likely to provide the best means of testing and if necessary modifying the theoretical interpretations offered.

The three manipulations we chose to study involve irrelevant speech, articulatory suppression, and manual tapping, in each case studying their effect on the serial recall of sequences of phonologically similar and dissimilar consonants. We assume that the presence of an acoustic similarity effect will provide prima facie evidence that acoustic coding is involved, while its absence would suggest a change in coding to a nonacoustic code.

These three types of disruption are applied across three separate studies. In Experiment 1, the irrelevant material comprised a single repeated token. In the case of irrelevant speech, the participants heard the digit 2, in the articulatory suppression condition, participants repeated the digit 2, while in the manual tapping condition, a single key was repeatedly
pressed. Experiment 2 was identical, with the exception that the rhythm of the presentation or operation of the irrelevant material was changed from equal interval to syncopated. Finally, Experiment 3 returned to equal-interval presentation, but increased the number of tokens from one to six. We attempt tentative predictions as to the effects to be expected from these variables on the basis of the phonological loop model, the O-OER model of Jones (1993), and the Nairne and Neath feature model (Nairne, 1990; Neath, 2000; Neath & Nairne, 1995), while accepting that those proposing the latter models may disagree with our interpretation.

EXPERIMENT 1

This study essentially replicates a number of phenomena that have already been observed, although typically not within the same experiment, as each has characteristically been used as a control for more complex conditions. Jones and colleagues have investigated the effect of a number of tokens within the irrelevant speech on its capacity to disrupt performance, concluding that “a sound sequence made by repeating a token produces small effects that are usually nonsignificant . . . whereas a sequence consisting of different tokens produces markedly greater disruption” (Tremblay & Jones, 1998, p. 660). This statement is based on numerous observations in which the effect of a single repeated token is not significantly different from a quiet control condition. It has occasionally been found that a single item may disrupt performance (for example, Tremblay & Jones, 1998), so the O-OER theory makes no clear prediction in this case, although any effects would be expected to be small. Neither the phonological loop nor, to the best of our knowledge, the feature hypothesis makes any clear prediction on the number of tokens required to create an irrelevant speech effect.

Articulatory suppression is a manipulation that has been used most extensively by proponents of the phonological loop hypothesis, where it is assumed to block both subvocal rehearsal and the phonological encoding of visually presented material. This hypothesis therefore predicts both impairment in memory performance and the removal of the phonological similarity effect with articulatory suppression. The O-OER hypothesis would predict that articulatory suppression would have a somewhat greater impact than irrelevant speech, because of the act of vocalization (Macken & Jones, 1995), but would expect this effect to be small or nonsignificant because suppression comprised a single item. Nothing in the O-OER model seems to predict that articulatory suppression would remove the phonological similarity effect. In the case of feature theory, one would assume that both a heard and a repeated single item would be expected to disrupt retention, perhaps producing a similar but more marked effect to that found with irrelevant speech. Whether either or both would interact with phonological similarity is hard to determine without performing simulations.

In the case of equal-interval tapping, the phonological loop hypothesis would predict little impairment, given that the task does not appear to place heavy attentional demands on executive processes. It would also predict little influence of equal interval tapping on the phonological similarity effect. The O-OER hypothesis would likewise predict little effect, other than that due to the minimal attentional load, since single repeated items should not have a great effect on serial recall. In the case of the feature hypothesis, there would appear to be a prima facie argument for disruption on the grounds that the items to be remembered will have their relevant features disrupted by the features of the concurrent motor task. Should this not occur, however, it could presumably be attributed to the absence of cue overlap, a plausible
suggestion that does, however, open up the question of how one independently determines degree of feature overlap. Once again, none of the theories would appear to predict that simple tapping would eliminate the phonological similarity effect.

Method

Participants
A total of 24 undergraduates at John Carroll University participated to partially fulfil an introductory psychology course requirement or to earn extra credit in an upper division psychology class.

Memory materials
On each trial participants saw either six letters that were phonologically similar, CDGPTV, or six letters with no vowel sounds in common, BFHJQR. The six letters were presented in a different random order on each trial. The same random orders were used for all participants. Each letter was shown in the centre of the computer screen for 250 ms, followed by a blank screen for 250 ms, resulting in a rate of two letters per second.

Concurrent conditions
In the control condition, participants simply watched the letters appear on the screen and waited six seconds for the cue to write them down. The irrelevant speech condition consisted of the single word TWO in a female voice presented on head phones and repeated at the rate of twice a second. The irrelevant speech began when the first letter was presented, and it continued to the end of a 6-s delay period, when participants were cued to recall the letters. In the articulatory suppression condition, participants repeated the word TWO approximately twice a second, and in the tapping condition they tapped the 0 key on the number key-pad twice a second. They performed these concurrent tasks while seeing the letters and during the 6-s delay period before the recall cue. On each trial in the tapping condition there were four computer beeps at the target rate, to set the pace, before the letters began to appear on the screen. There were no beeps during the presentation of the letters or during the delay period as this would have provided an additional irrelevant sound. The computer monitored the participants’ tapping rate and, at the end of each trial, informed the experimenter of the number of 500-ms intervals in which there was no tap and the number in which there were two or more taps. If participants made more than five errors in matching the tapping rate, the experimenter cautioned them to be more careful to match the rate.

Procedure
Participants watched the six letters on a computer monitor and waited for the word RECALL, to appear on the screen. Then they wrote their answers on an answer sheet containing six lines for each trial. They filled in the lines from left to right and wrote something on each line before going on to the next line. After each trial participants covered their responses with a piece of cardboard so that they were not distracted by their responses on previous trials. Participants did four practice trials before each condition. In each condition there were 24 trials. Letter type was random with the constraint that 12 lists from each letter set occurred, with each type of letter presented twice within each block of four trials. All participants did the control condition first. The order of the other three concurrent tasks was counterbalanced across participants, with four people experiencing each of the possible orders.
Results

The cell means and standard deviations are shown in Table 1.

A $2 \times 4$ (letter set by concurrent task) analysis of variance (ANOVA) of the proportion of letters correct out of six showed that there was a main effect for letter set, $F(1, 23) = 40.34, p < .001$, a main effect for concurrent task, $F(3, 69) = 42.74, p < .001$, and a significant interaction between letter set and concurrent task, $F(3, 69) = 15.48, p < .001$.

The alpha level was set at .05 for the tests of simple effects, and all differences reported as significant meet or exceed this level. The Bonferoni approach to control for familywise error rates was used. Tests of simple effects of letters within each concurrent task condition showed that participants recalled a higher proportion of the phonologically dissimilar letters than of the phonologically similar letters in the control, irrelevant speech, and tapping conditions. There was no difference in the proportions of the two kinds of letters recalled in the articulatory suppression condition. Thus, the phonological similarity effect was found in the control condition and in the presence of irrelevant speech and equal-interval tapping. As expected, articulatory suppression abolished the phonological similarity effect.

Tests of simple effects of concurrent tasks within each of the letter sets showed that, with the phonologically dissimilar letters, articulatory suppression resulted in poorer performance than that found with the other three conditions, which did not differ from each other. With the similar letters, there was also no irrelevant speech effect—that is, there was no difference between the control and irrelevant speech conditions. There was no difference between tapping and articulatory suppression, but people performed better in both the control and irrelevant speech conditions than in the tapping and articulatory suppression conditions. Thus, with phonologically similar letters, the effect of equal-interval tapping was similar to the effect of saying the word TWO repeatedly.

Discussion

Consistent with the earlier work of Jones, Madden, and Miles (1992), we found no irrelevant speech effect for a single repeated item. As predicted by the phonological loop hypothesis, articulatory suppression by a single item markedly impaired performance and removed the influence of phonological similarity. Finally, with equal interval tapping the phonological

<table>
<thead>
<tr>
<th>Letter set</th>
<th>Phonologically similar</th>
<th>Phonologically dissimilar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concurrent task</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Control</td>
<td>.61</td>
<td>.17</td>
</tr>
<tr>
<td>Irrelevant speech</td>
<td>62</td>
<td>.17</td>
</tr>
<tr>
<td>Tapping</td>
<td>53</td>
<td>.19</td>
</tr>
<tr>
<td>Articulatory suppression</td>
<td>46</td>
<td>.17</td>
</tr>
</tbody>
</table>
similarity effect remained, and with equal-interval tapping there was a moderate effect on overall performance. The theoretical implications of these results will be discussed in conjunction with those of Experiments 2 and 3.

**EXPERIMENT 2**

This experiment is identical to Experiment 1 with the exception that the irrelevant material involved either hearing or generating at a syncopated rather than an equal-interval rhythm. Its principal function was to replicate Saito’s (1993, 1994) finding that syncopated tapping not only impairs performance, but also removes the phonological similarity effect. What might be the effects of syncopated irrelevant speech? Jones and Macken (1995) studied the effect of the same token presented at unpredictable irregular intervals and found that this disrupted serial recall more than the same token repeated at regular intervals. Would the predictable, syncopated sequence be similar to the irregular tokens in their study, or would it form what they refer to as a higher order object? The O-OER hypothesis might, then, predict either an irrelevant speech effect based on the irregularity of an incoming stimulus or its absence, on the grounds that the predictability of the syncopation would allow the stimulus to be organized into a higher order perceptual object. Once again, it is unclear what the feature hypothesis would predict.

In the case of syncopated articulatory suppression, all the theories might suggest that generating a syncopated rhythm is more attentionally demanding and hence would produce an overall impairment in performance. If this is the case, then any interpretation of the articulatory suppression effect, in terms of its attentional demand, would presumably predict an even greater level of impairment when suppression is syncopated. The phonological loop hypothesis would predict the elimination of the phonological similarity effect under conditions of suppression and, in Saito’s modification, under syncopated tapping. It would not predict a reduction of the phonological similarity effect as a result of syncopated irrelevant speech. Once again, since the interpretation of the phonological similarity effect is less focal to the O-OER and feature hypotheses, it is less clear what they would predict.

**Method**

**Participants**

A total of 24 undergraduates at John Carroll University, who had not taken part in the first experiment, participated to partially fulfil an introductory psychology course requirement or to earn extra credit in an upper division psychology class.

**Materials and procedure**

The to-be-remembered materials and general procedures were identical with those used in Experiment 1. The irrelevant speech was the word TWO recorded in a female voice and presented over headphones. The word TWO was repeated in the syncopated rhythm of long (L) intervals, 500 ms, and short (s) intervals, 250 ms, used by Saito (1993, 1994). Two repetitions of this syncopated pattern, L.s.L.s.L.s.L.s.L, lasted 4 s, and the entire pattern was repeated throughout the letter presentation and delay before recall. Note that this is a syncopated and not a simple repeating pattern, because each replication begins and ends with a long interval. In the articulatory suppression condition, people
repeated the word TWO in this syncopated rhythm. In the tapping condition, people tapped the 0 key on the number key pad in this syncopated rhythm. Because of the difficulty of the tapping task, participants did several practice trials before beginning this condition. The computer monitored the participant’s tapping rate and listed the number of intervals in which there was no tap and the number in which there were two or more taps. Throughout the study, if participants made more than five tapping errors on a trial, they were cautioned and encouraged to be more accurate.

Results

The cell means and standard deviations are shown in Table 2.

A 2 × 4 (letter set by concurrent task) ANOVA of the proportion of letters correct out of six showed that there were main effects of letter set, \( F(1, 23) = 41.91, p < .001 \), and concurrent task, \( F(3, 69) = 53.23, p < .001 \), and a significant interaction between letter set and concurrent task, \( F(3, 69) = 23.45, p < .001 \).

Tests of simple effects of letters within each concurrent task condition showed that people recalled more phonologically dissimilar than phonologically similar letters in the control and the irrelevant speech conditions, but there was no difference in the recall of the two types of letters in the tapping and articulatory suppression conditions. Thus both the syncopated tapping task and syncopated articulatory suppression abolished the phonological similarity effect, but syncopated irrelevant speech did not.

Tests of simple effects of concurrent tasks within each of the letter sets showed that, with the phonologically dissimilar letters, there was significantly better performance in the control condition than in the other three tasks. Performance in the irrelevant speech condition was significantly poorer than in the control condition, but significantly better than in the tapping and articulatory suppression conditions. The tapping and articulatory suppression were not different from each other. Thus, for unrelated letters there was an irrelevant speech effect with syncopated repetition of the same single word. The effects of syncopated tapping and syncopated articulatory suppression were equivalent. With the phonologically similar letters, the pattern was identical to that found in the first experiment. There was no difference between the control condition and irrelevant speech, showing no irrelevant speech effect for phonologically similar letters when the irrelevant speech was a single word repeated in a syncopated rhythm. There was no difference between tapping and articulatory suppression. However, the

<table>
<thead>
<tr>
<th>Letter set</th>
<th>Similar</th>
<th>Phonologically dissimilar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concurrent task</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Control</td>
<td>.53</td>
<td>.19</td>
</tr>
<tr>
<td>Irrelevant speech</td>
<td>.53</td>
<td>.18</td>
</tr>
<tr>
<td>Tapping</td>
<td>.39</td>
<td>.19</td>
</tr>
<tr>
<td>Articulatory suppression</td>
<td>.38</td>
<td>.11</td>
</tr>
</tbody>
</table>
control and irrelevant speech conditions were both different from tapping and articulatory suppression. For both types of letters, syncopated tapping had an effect similar to articulatory suppression in the same syncopated rhythm.

Discussion

The first point to note about Experiment 2 is that it successfully replicates Saito (1993, 1994) in observing a clear decrement as a result of syncopated manual tapping, together with the removal of the phonological similarity effect. A very similar picture occurs when syncopated articulatory suppression is required. There is no significant difference between these two conditions. In the case of irrelevant speech, introduction of syncopation does lead to a small decrement, presumably because it makes the sound more speech-like and/or introduces a more complex serial component.

This pattern of results clearly fits the phonological loop hypothesis, as modified by Saito (1993, 1994). The irrelevant speech data are consistent with the observation of an effect of a single unpredictable spoken item found by Jones and Macken (1995). It is unclear, however, how the O-OER hypothesis would account for the dramatically greater influence of syncopated tapping and of the removal of the phonological similarity effect by this and by articulatory suppression. One possible way out of this problem for both the O-OER and feature hypotheses might be to suggest that syncopated tapping is highly attentionally demanding, and that this interferes with phonological encoding. We return to this issue in the final discussion. Before doing so, Experiment 3 is described, in which the number of tokens is increased from 1 to 6. Extensive research by Jones and colleagues has shown that increasing the number of auditory tokens beyond one markedly increases the effect of irrelevant speech.

EXPERIMENT 3

This is identical to Experiment 1, with the exception that instead of a single token, a sequence of six items occurs. In the case of irrelevant speech, the digits 1, 2, 3, 4, 5, and 6 are spoken repeatedly in the canonical order. Under articulatory suppression, the subject utters these, while in the tapping condition, the subject is required to successively tap six keys. There is abundant evidence from Jones and his group that increasing the number of irrelevant auditory items beyond one greatly increases the irrelevant speech effect (Jones et al., 1992; Jones & Macken, 1993; Tremblay, Macken, & Jones, 2000). In the case of articulatory suppression, there is some evidence that more complex suppression impairs memory to a somewhat greater extent (Jones & Macken, 1995; Saito, 1997), whereas to the best of our knowledge, the effect on immediate verbal memory of size of token set in a tapping task with equal interval tapping of a series of keys has not been studied. In terms of theoretical predictions, the phonological loop hypothesis would predict an irrelevant speech effect, together with a more dramatic effect of suppression that might be slightly greater than that involved in repeating a single item. Tapping a series of keys may have a slightly greater attentional demand than tapping a single key, together with some potential spatial disruption, although the impact of this on immediate verbal recall is likely to be minimal. Of these manipulations, only suppression would be expected to eliminate the phonological similarity effect.
The O-OER hypothesis would clearly predict impairment from this changing irrelevant speech and a greater degree of impairment from articulatory suppression because of the combined effects of changing items and the attentional demand of vocalization (Macken & Jones, 1995). It might also be expected to predict a substantial effect of the serial key tapping, since this is clearly an activity involving order information, which should presumably interfere with ordered recall of the letter sequences, given the assumption of a single common multimodal surface on which serial order is represented. It is again difficult to know what to predict from the feature hypothesis. All three disrupting conditions involve the presentation of irrelevant material simultaneously with the letters to be recalled. In principle, one might expect all to be capable of disrupting storage. Again, however, it would seem to be possible to accommodate any of a range of results by making appropriate assumptions about the degree of overlap between the features in the remembered and irrelevant items.

Method

Participants

A total of 24 undergraduates at John Carroll University participated to partially fulfill an introductory psychology course requirement or to earn extra credit in an upper division psychology class.

Materials and procedure

The to-be-remembered materials were identical with those used in Experiments 1 and 2. The irrelevant speech comprised the words ONE, TWO, THREE, FOUR, FIVE, SIX, repeated in that order at the rate of two per second. In the irrelevant speech condition, people repeated the words ONE, TWO, THREE, FOUR, FIVE, SIX, in that order at the rate of about two words per second. For the tapping condition, the keys on the number key-pad were relabelled so that the number 1 was on the 4 key, 2 on the 5 key, and so on, so that tapping the numbers 1, 2, 3, 4, 5, 6, was in a clockwise loop, starting at the upper left of the relabeled keys. Participants tapped the keys, in order, at the rate of two keys per second. If they deviated from this pattern and rate, the experimenter cautioned them and told them what they were doing wrong.

Results

The cell means and standard deviations are shown in Table 3. A 2 × 4 (letter set by concurrent task) ANOVA of the proportion of letters correct out of six showed that there was a main effect for letter set, \( F(1, 23) = 35.02, p < .001 \), a main effect for concurrent task, \( F(3, 69) = 37.36, p < .001 \), and an interaction between letter set and concurrent task, \( F(3, 69) = 21.35, p < .001 \).

Tests of simple effects of letters within each concurrent task condition showed that people remembered more phonologically dissimilar letters than phonologically similar letters in the control, irrelevant speech, and tapping conditions, but in the articulatory suppression condition there was no difference in the proportions of the two types of letters recalled. Only articulatory suppression abolished the phonological similarity effect.

Tests of simple effects of concurrent tasks within each of the letter sets showed that, with phonologically dissimilar letters, people remembered more letters in the control condition than when hearing irrelevant speech. People also remembered fewer letters in the tapping and
articulatory suppression conditions than in the control condition. There was no difference in the effects of irrelevant speech and tapping, but performance with articulatory suppression was poorer than that with any of the other tasks. With the phonologically similar letters, there was no irrelevant speech effect, but both tapping and articulatory suppression interfered with recall compared to the control condition. Irrelevant speech interfered significantly less than articulatory suppression, but the effect of irrelevant speech was not different from the effect of the tapping condition. Tapping and articulatory suppression were not different from each other.

Discussion

As expected on the basis of existing literature, irrelevant speech involving multiple tokens had a clear impact on recall performance, a result consistent with all three explanatory hypotheses. The phonological similarity effect remained, however, consistent with the phonological loop hypothesis, with the O-OER hypothesis, and with considerable earlier research (see Larsen, Baddeley, & Andrade, 2000, for a review), but inconsistent with feature theory as presented by Neath (2000). Articulatory suppression again has a substantial impact on performance, reducing recall and eliminating the phonological similarity effect. This fits neatly into the phonological loop hypothesis while presenting a challenge, though not necessarily an insuperable one, for the two alternative hypotheses. Finally, sequential tapping has a small effect on performance but does not remove the effect of phonological similarity. Each of the theories could probably deal with this result, without necessarily predicting it. The O-OER hypothesis might suggest that its serial component has an effect analogous to that of irrelevant speech. This, however, still leaves the problem of why both syncopated tapping and articulatory suppression have effects that are substantially greater, and which obliterate the phonological similarity effect. A similar problem confronts the feature hypothesis.

GENERAL DISCUSSION

The data from the three experiments are summarized in Table 4, which presents the magnitude of the effects of the various manipulations on performance in the dissimilar condition.
For each condition, the difference in percentage correct between it and the single task baseline is represented. Hence, the reported 2% effect of an equal-interval repeated single item is based on the difference between performance in the single task condition (81%) and the performance of those participants in the control condition (83%). In addition, Table 4 gives the effect size, eta squared, for that comparison. For each condition, the magnitude of the phonological similarity effect is also given, this time by subtracting the percentage correct in the phonologically similar condition from that with dissimilar sequences. Again, effect size is given.

The data for the dissimilar letters fall into four separate groups, depending on the magnitude of the effect shown:

1. Irrelevant speech consisting of a single item repeated at equal intervals clearly does not impair performance (2% decrement, effect size .062).
2. There are small effects of equal interval tapping, with either one or six tokens (10%, .226; 11%, .388, respectively).
3. Somewhat more robust effects are produced by the need to ignore a syncopated single item (11%, .467), and by the standard irrelevant speech effect whereby multiple items are repeated at equal intervals (15%, .539).
4. Finally, there are four conditions that show a substantial degree of impairment: syncopated tapping (37%, .844), and articulatory suppression, whether with a single item at equal intervals (37%, .844), with a single syncopated item (41%, .8981), or with multiple items (42%, .878).

The pattern of results for the effect of phonological similarity shows two broad groupings. The three articulatory suppression conditions and the syncopated tapping all completely remove the effect of phonological similarity, whereas the phonological similarity effect survives in all other conditions.

### Table 4
Summary of the magnitude of the effect of each variable manipulated across Experiments 1–3

<table>
<thead>
<tr>
<th>Condition</th>
<th>Effect</th>
<th>(\eta^2)</th>
<th>Effect</th>
<th>(\eta^2)</th>
<th>Effect</th>
<th>(\eta^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Single token at equal intervals</td>
<td>Synopated single token</td>
<td>Multiple tokens</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irrelevant speech</td>
<td>Decrement</td>
<td>2</td>
<td>.062</td>
<td>11</td>
<td>.467</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Similarity effect</td>
<td>19</td>
<td>.532</td>
<td>12</td>
<td>.462</td>
<td>11</td>
</tr>
<tr>
<td>Manual tapping</td>
<td>Decrement</td>
<td>10</td>
<td>.226</td>
<td>37</td>
<td>.844</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Similarity effect</td>
<td>20</td>
<td>.587</td>
<td>0</td>
<td>.004</td>
<td>24</td>
</tr>
<tr>
<td>Articulatory suppression</td>
<td>Decrement</td>
<td>37</td>
<td>.827</td>
<td>41</td>
<td>.881</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Similarity effect</td>
<td>0</td>
<td>.000</td>
<td>-3</td>
<td>.053</td>
<td>-3</td>
</tr>
</tbody>
</table>

*Note:* Magnitude of decrement from the control condition for the dissimilar letters is given, in percentages, together with the effect size (eta squared). The effect of acoustic similarity (the difference between similar and dissimilar letters) and effect size (eta squared) is also shown for each condition.
Finally it should be noted that there was a reduced effect of phonological similarity in the two conditions in which irrelevant speech impairs performance, with a syncopated single token and with multiple tokens (both 12%, compared to effects of 19%–24% for the remaining conditions). It is clear from Tables 2 and 3 that this reduction reflects the absence of any effect of irrelevant speech in the presence of phonological similarity.

In calculating the percentage decrement resulting from irrelevant speech, one should note that the silent baseline is, of course, different for similar and dissimilar sequences. Could this be responsible for the lack of an irrelevant speech effect for similar letters? One possibility might be a floor effect. This seems unlikely, since performance on similar sequences under irrelevant speech conditions is not only well above zero (53–64%), but is also consistently above performance on such sequences when they are accompanied by articulatory suppression or tapping. Could one argue that, rather than taking a difference between two percentages, the difference should be expressed as a proportion of the silent baseline condition? This would have little effect, since the difference is very small (3% in Experiment 3, absent in Experiment 2, and minimally reversed, –1%, in Experiment 1).

We attempt to give as simple an account as possible of this overall pattern of data based on a number of plausible assumptions, applied within the theoretical framework of a phonological loop. For present purposes, we will adopt the changing state hypothesis (Jones, 1993) as being the best, indeed virtually the only, detailed account of what characterizes irrelevant auditory items that do and do not interfere with memory. We deviate from the position of Jones, however, in not adopting the O-OER hypothesis.

Our assumptions are as follows:

1. The phonological similarity effect represents the storage component of the phonological loop, with the traces of phonologically similar items having less discriminable cues than do dissimilar items, hence leading to poorer recall.
2. Irrelevant speech impacts on the same system as phonological similarity. The nature of the impact is discussed below.
3. Articulatory suppression and, following Saito (1993), syncopated tapping interfere with performance in two ways. First, they prevent visual stimuli being represented within the phonological store, and second, they prevent maintenance rehearsal. In the present study, both effects are present, whereas if material were presented auditorily, only the prevention of rehearsal would occur.
4. The requirement to perform a second active task concurrently with the span task will result in a modest deficit, presumably due to attentional demands on the central executive.
5. Finally, reliance on a phonological code is an optional strategy, and that it tends to be abandoned when performance drops below a critical point (Baddeley, 2000b).

None of these involves assumptions over and above those already made by the phonological loop model (Baddeley, 1986; Baddeley & Logie, 1999). We now consider the previously described pattern of results. We assume that the presence of a phonological similarity effect is an indication that a phonological code is present and is actively contributing to performance in the dissimilar condition. If this were not the case, there is no reason to assume that phonological similarity would have any impact on performance.
In the case of the repetition at equal intervals of a single item, our results coincide with those reported most frequently by Jones, although it should be noted that this condition does occasionally give rise to a small irrelevant speech effect (Jones & Tremblay, 2000). As such, our result broadly coincides with the predictions of the changing state hypothesis.

The fact that the requirement to tap at equal intervals causes a modest decrement in performance, for both single and multiple tokens, is consistent with the assumption that the demand to combine two active tasks places a load on executive processes, as indeed is assumed by many theories of attention (Pashler, Johnston, & Ruthruff, 2000). It is for that reason that equal-interval tapping is often used as a control in studies involving suppression and the phonological loop (Baddeley, 1986). The fact that tapping does not reduce the phonological similarity effect is consistent with the assumption that its impact is on the central executive, not on the phonological loop. Furthermore, the equivalence of repetition of a single response and repetition of multiple responses distinguishes it from the effect of irrelevant speech. This is consistent with the assumption of a different point of origin for the two effects, as the phonological loop model proposes.

We obtained a moderately large effect of irrelevant speech, both from a single item presented in a syncopated rhythm and from multiple items presented at equal intervals. The single item effect clearly needs to be taken into account by the changing state hypothesis, but does not refute it. Of particular interest here is the clear effect of irrelevant speech on memory for dissimilar sequences, together with its equally clear absence for phonologically similar ones. We interpret this difference in the effect of irrelevant speech on the recall of the two types of letters by assuming that the effect of phonological similarity is to reduce substantially the usefulness of the phonological code (a difference in the percentage of phonologically similar and dissimilar letters recalled of 22%–23% in the baseline condition). We suggest that phonological similarity leaves the acoustic code severely impoverished and that it is hence abandoned (Baddeley, 2000b). There is virtually no further disruption in recall of the phonologically similar letters by irrelevant speech (–1% to 3%).

This leaves us with the four conditions in which impairment of recall of the dissimilar letters by the concurrent task is substantial (37%–42%). These tasks are syncopated manual tapping and articulatory suppression with all three of the concurrent tasks. If we accept Saito’s (1993, 1994) assumption that tapping in a syncopated rhythm interferes with the capacity to generate speech, then the absence of a phonological similarity effect in all four conditions is to be expected, given that subjects are likely to have difficulty in converting the visual stimulus into an acoustic code. Performance is likely to be further impaired by the executive demand of concurrently generating a sequence of taps or utterances. Finally, these four conditions are all likely to prevent the use of maintenance rehearsal during input, storage, and retrieval. Even so, it is important to note that performance does not reduce to zero, with subjects still being able to reproduce two or three letters. This indicates that the phonological loop is not the only means of maintaining information, giving further support for the assumption of an additional visual or, perhaps more likely, multimodal store, such as the episodic buffer proposed by Baddeley (2000a).

Our results therefore seem to fit readily into the phonological loop model, with relatively minimal further assumptions, one being an acceptance of Saito’s (1993, 1994) evidence for the link between rhythm and speech, and the other being the acknowledgment of the generally
accepted assumption that performing two simultaneous tasks is more demanding than performing a single task.

We do not see how the results can readily be fitted into alternative hypotheses. In particular, whereas the effect of phonological similarity and its interaction with the various conditions is exactly as expected from the phonological loop hypothesis, it is unclear how this variable would be handled by either the O-OER or the feature hypotheses. Clearly, both will surely be able to present some account of the effect of phonological similarity since it is a major variable, but we are unclear as to how the complex pattern of interactions that we observe would be explained.

A possible solution to this problem is offered by the observation that the absence of a phonological similarity effect occurs only in conditions where performance, even on the dissimilar letter set, is low. We are, however, clearly not observing a simple floor effect, since performance in these conditions ranges between .35 and .52, two to three letters. Our own interpretation assumes a strategy switch away from reliance on phonological coding. Evidence for this is discussed in Baddeley (2000b), where increasing sequence length substantially beyond span appears to lead to the abandonment of phonological coding, whether measured by the effect of acoustic similarity within letter sequences (Hanley & Broadbent, 1987; Salamé & Baddeley, 1986) or word length (Neath & Nairne, 1995).

The fact that patients with a specific short-term phonological memory deficit can increase their span markedly when presentation is visual rather than auditory (Vallar & Shallice, 1990), argues for the need to assume an additional nonphonological storage system. Whether this involves purely visual coding (see Logie, Della Sala, Wynn, & Baddeley, 2000, for a recent discussion) or reflects a more multidimensional store such as the episodic buffer system proposed by Baddeley (2000a) remains an open question.

The data cited so far have inferred the nature of coding from the effects of similarity or stimulus material. A more direct approach is to give subjects explicit instructions to use one or other strategy. Wickelgren (1964, 1967) used this method to provide further evidence that grouping digits in threes led to enhanced performance. A recent study by Hanley and Bakopoulou (2003) instructed subjects to remember sequences of phonologically similar or dissimilar consonants presented auditorily and followed by either silence or irrelevant speech. One group was asked to encode phonologically, a second to seek semantic links, while the third group was given no strategy instruction. The clearest effects occurred in the group instructed to use phonological coding, with marked effects of both phonological similarity and irrelevant speech. A similar though less pronounced pattern occurred in the uninstructed control group, while neither phonological similarity nor irrelevant speech effect influenced performance when subjects were told to encode semantically. Interestingly, this latter group performed at a higher level than the other two, suggesting that, under these circumstances at least, semantic coding was an effective strategy. It is clear, however, that, as proposed above, performance is strategy dependent, with both irrelevant speech and phonological similarity having a pronounced effect when subjects are instructed to encode phonologically, and furthermore that some, if not necessarily all, subjects spontaneously appear to use this strategy in the control group.

A second problem for both theories would appear to arise from their assumption that articulatory suppression and irrelevant speech involve essentially the same mechanism. If so, why does the hearing of a single repeated item have no effect in contrast to multiple items, whereas saying one item is just as disruptive as saying a sequence of six?
The question of stimulus regularity would also appear to present problems for both approaches. Jones could clearly handle the irrelevant speech effect by assuming that syncopated sounds involve a changing state, whereas equal interval sounds do not. This would, however, leave the problem of why regularity should be the most important variable with tapping, and number of tokens is of little importance, while the opposite pattern occurs for irrelevant speech. Furthermore, why should syncopated tapping but not equal-interval tapping eliminate the phonological similarity effect?

Similar problems would appear to occur in the case of the feature hypothesis; why should tapping a syncopated rhythm produce a much more dramatic impairment than tapping at equal intervals, on one or on six keys, whereas articulating a syncopated rhythm is no more disruptive than articulating a regular sequence? We do not wish to claim that there is no possible way in which either model could explain our results, but would like to encourage our colleagues to say how they would do so.

We should conclude by returning to the question of how the irrelevant speech effect can be explained within the phonological loop hypothesis. We begin with an issue that has dogged the phonological loop hypothesis over recent years, namely that of the effect of phonological similarity between the irrelevant sound and the material being remembered. Based on a single result, Salamé and Baddeley (1982, Exp. 5) speculated that the effect of irrelevant speech might occur as a result of a form of mnemonic masking, with speech masking memory, because both involved a phonological code. This effect did not replicate, and it is now clear that there is no effect of the similarity between the remembered and ignored material on subsequent recall (Buchner, Irmen, & Erdfelder, 1996; Jones & Macken, 1995; Larsen et al., 2000; LeCompte & Shaibe, 1997).

As specified elsewhere (Baddeley, 2000b; Larsen et al., 2000), there are at least two ways of accounting for this result within the phonological loop model.

The first of these is to modify the suggestion of Neath (2000) that interference occurs at the level of features rather than items, but to propose that such features operate at the phoneme or syllable level. Consider, then, a sequence of letters with a common vowel sound (e.g., b, c, g, d) with a set of rhyming ignored words (he, key, flea, etc.). If the interference were phoneme specific, then interference with the remembered letters sequence would principally operate upon the vowel sound, which, being redundant, would have little effect on performance. Although suggested by the feature hypothesis, this is not the interpretation offered by Neath, and it would clearly need further development and testing before it could be accepted as an adequate explanation.

An alternative type of explanation might be based on the family of more detailed models of the phonological loop that have emerged in recent years, concerned with providing an adequate account of how serial order is stored and retrieved within the loop (Burgess & Hitch, 1992, 1999; Henson, 1998; Page & Norris, 1998). Such models typically assume that item and order information are stored and retrieved by different processes. For example, the Page and Norris (1998) model assumes that each item is linked to a starter node by an association, with subsequent items having successively weaker links. Retrieval involves a competitive queuing process whereby the strongest link to the starting point is selected and emitted, then inhibited before selecting the next strongest. Such a model could account for the lack of an effect of phonological similarity between the items recalled and of those ignored by assuming that irrelevant speech adds noise to the associative links, while phonological similarity is...
represented at the level of the item. It would be of considerable interest to see if this or other models could provide a precise account of the current pattern of results.

REFERENCES


*Original manuscript received 14 February 2002
Accepted revision received 23 July 2002*