Word Dose in the Disruption of Serial Recall by Irrelevant Speech: Phonological Confusions or Changing State?

Andrew M. Bridges and Dylan M. Jones

University of Wales College of Cardiff, Cardiff, U.K.

Irrelevant background speech disrupts serial recall of visually presented lists of verbal material. Three experiments tested the hypothesis that the degree of disruption is dependent on the number of words heard (i.e. word dose) whilst the task was undertaken. Experiments 1 and 2 showed that more disruption is produced if the word dose is increased, thereby providing evidence to support the experimental hypothesis. It was concluded from the first two experiments that the word-dose effect might be the result of increasing the amount of changing-state information in the speech. The results of Experiment 3 supported this conclusion by showing an interaction between word dose and changing-state information. It was noted however that the results might be explained within the working memory account of the disruptive action of irrelevant speech. A further two experiments cast doubt on this possibility by failing to replicate the finding that the phonological similarity between heard and seen material affects the degree of interference (Salamé & Baddeley, 1982). The findings are discussed in relation to the changing state hypothesis of the irrelevant speech effect (e.g. Jones, Madden, & Miles, 1992).

It is a well-documented and robust finding that “unattended” background speech reduces performance on serial recall tasks by approximately 30% (Baddeley & Salamé, 1986; Colle, 1980; Colle & Welsh, 1976; Hanley & Broadbent, 1987; Jones, 1994; Jones & Macken, 1993; Jones, Macken, & Murray, 1993; Jones et al. 1992; Jones, Miles & Page, 1990; Miles, Jones & Madden, 1991; Morris & Jones, 1990a, 1990b; Salamé & Baddeley, 1982, 1986, 1987, 1989, 1990). The term “irrelevant speech effect” (Jones et al., 1990) was suggested to be a better description of the phenomenon in preference to referring to the speech as unattended, thereby avoiding pre-judging the nature of its processing. It has

Andy Bridges is now at the Psychology division, Aston University. Requests for reprints should be sent to Andy Bridges, Psychology Group, Aston University, Aston Triangle, Birmingham, B4 7ET, UK. Email: a.m.bridges@aston.ac.uk

The work was carried out at the School of Psychology, Cardiff, University of Wales during the period of an ESRC studentship awarded to the first author. Thanks are due to Paul Farrand, Bill Macken, and Alison Murray for help in the development of this paper and to Graham Hitch, Rick Hanley, and one anonymous referee for comments on earlier drafts.

© 1996 The Experimental Psychology Society
been shown that gross characteristics of the speech source such as meaning and intensity do not determine the disruption (Colle, 1980; Colle & Welsh, 1976; Jones et al., 1990) and, by varying the section of the task during which the speech is played, that the effect occurs in memory and not at encoding (Jones, 1993; Macken & Jones, 1995; Miles et al., 1991).

The identification of aspects of processing that are pre-attentive was noted by Massaro and Cowan (1993) to be important in an information-processing approach to the study of cognition. This allows the study of processes that arise without active attention and can provide considerable insight into the mechanisms of both peripheral and higher processing. Investigation of the irrelevant speech effect has facilitated this approach in the study of auditory cognition, contributing considerably to the progression of theories concerning memory and attention (see Jones & Morris, 1992 for a review). To date, however, a range of factors concerning the characteristics of auditory material that produce this disruption has yet to be fully explored. Two specific issues are addressed in this study: the degree to which the number of words (word dose) in the irrelevant speech determines the disruption; and whether the phonological similarity between heard and seen material is an important factor in the effect.

The Working Memory Account of the Irrelevant Speech Effect. A number of influential studies have investigated whether irrelevant acoustic material other than speech produces similar disruption of serial recall. It was found that continuous white noise fails to produce disruption (e.g. Baddeley, 1968), bursts of white noise produce some disruption (Salamé & Baddeley, 1982, Experiment 2; Salamé & Wittersheim, 1978), and the most robust interference is produced by speech. The conclusion was drawn, therefore, that non-speech material does not have the capacity to disrupt serial short-term memory to the same degree as speech.

This position harmonizes well with Baddeley and colleagues' influential theory of working memory (Baddeley & Hitch, 1974; see Baddeley, 1992a, 1992b for a review of the model), according to which items are represented by way of phonological codes in short-term memory, automatically in the case of speech, and by deliberate articulation if they are of visual origin. The irrelevant speech effect is therefore the result of an "alphabet soup" in which codes representing rehearsed visual items become confused with the speech codes that have automatic access to the memory store. Non-speech material does not have this automatic access to the phonological store, and therefore does not produce disruption of the representations of the visual material in the same way as speech. Recent studies have, however, shown that it is not only speech that disrupts serial recall. Specifically, sounds such as tones are equally likely to produce an irrelevant "speech" effect (e.g. Jones & Macken, 1993). These findings, by casting doubt on the working memory explanation, have motivated an alternative account of the effect—that is, the changing-state hypothesis (Jones et al., 1992).

The Changing-state Hypothesis. The finding that irrelevant speech consisting of a single repeated syllable did not produce appreciable disruption of serial recall (Jones et al., 1992) demonstrated that speech was not a sufficient constituent of an irrelevant sound for it to bring about an irrelevant speech effect. It was argued, instead, that the constituent
items in the speech must change from one to the next ("changing state") in order for the effect to be shown. This view stands in contrast to the working memory account, which proposes that speech has mandatory access to the phonological store, and will, therefore, automatically command some of the resources required for the maintenance of deliberately rehearsed visual material. Hence, any irrelevant speech sound would be expected to interfere with serial short-term memory.

The early studies of irrelevant sound were largely motivated by requirements to produce optimum noise level guidelines for industrial work places (for a review see Jones & Broadbent, 1991). Non-speech sounds used in these studies corresponded to the type of background noise that might be heard in such an environment, such as continuous white noise (Baddeley, 1968). Thus, the conclusion that the disruption of serial recall by irrelevant sound was virtually unique to speech was based on the investigation of the effects of "steady-state" sound. The possibility therefore remained that an "irrelevant speech" effect might be produced by changing-state non-speech sounds. Experimental investigation of this possibility demonstrated that disruption of serial recall, of a magnitude equivalent to that produced by speech, was produced by changing state tones (Jones & Macken, 1993), segmented glissandi (Jones, Macken, & Murray, 1993), and band-pass noise (Jones & Macken, 1995a). The conclusions drawn from these studies, formalized as the changing-state hypothesis, were that speech is neither a necessary nor sufficient constituent of an irrelevant sound for it to produce disruption of serial recall. Rather, the critical factor in determining the disruption is whether or not the irrelevant sound contains changing-state information, that is, each physical unit within the sound must be different to the one that preceded it.

Manipulating Changing State Using "Word Dose". According to the changing-state hypothesis, increasing the amount of changing-state information should increase the degree of disruption. To date this has only been investigated in one way—by manipulating the phonological similarity between the items in the speech (Jones & Macken, 1995b). If the words in the speech are more phonologically similar to each other—as, for example, in rhyming words—there will be less change from word to word, and hence less changing-state information. This type of speech should therefore produce less disruption of serial recall. Experimental support for this position was provided by a series of experiments that contrasted rhyming words (e.g. knee, key, fee) with non-rhyming words (e.g. deaf, pay, bell). It was found that, as predicted, the changing-state (non-rhyming) speech produced more disruption than the steady-state (rhyming) words (Jones & Macken, 1995b). The purpose of the present study was to investigate an alternative technique for manipulating the amount of changing-state information in the speech, by varying the total number of irrelevant spoken words heard during the course of a serial recall task, that is, the "word dose".

Changing-state information, it is suggested, may be derived from the difference between subsequent units in the irrelevant sound. Increasing the number of units would, therefore, increase the number of between-unit changes, and hence the amount of changing-state information. It would therefore be expected that increasing the word dose in an irrelevant speech sequence would increase the degree of disruption to serial recall. The first experiment of this study sought to examine this possibility, by contrasting
the effects of three types of irrelevant speech, each containing a different level of word dose, on a serial recall task.

**EXPERIMENT 1**

The first experiment in the series was intended to investigate the possibility that word dose is an important factor in the disruptive effect of irrelevant speech on serial recall. Word dose was manipulated by contrasting three types of speech: one in which there was *no gap* between words (the treatment with most words), one in which there was a *small gap* between words (fewer words), and one in which there was a relatively *big gap* between words (fewest words). It was predicted that increasing the word dose would increase the amount of changing-state information. Therefore, the prediction according to the changing-state hypothesis is that a continuum of increased disruption is expected, with most interference occurring in the condition with the most words, and least interference shown when there is the smallest number of words in the speech stream.

**Method**

**Subjects**

Twenty-four undergraduate student volunteers were paid for participating in the experiment. All subjects had English as their first language and reported normal hearing.

**Apparatus and Materials**

Items to be recalled were presented serially on the screen of an Apple Macintosh Quadra microcomputer. Lists were constructed from the random arrangement of the letters F, K, L, M, Q, R, S, T, and Y, with the constraint that a letter could not appear in the same serial position in two consecutive lists and that recognizable words, letter strings, or acronyms were not included. The lists were stored and presented within a *Hypercard* environment.

All three speech conditions consisted of the same five letter names—*B* (bee), *I* (eye), *J* (jay), *N* (enn), and *Z* (zed)—in random order (these will be referred to as *utterances*, as they do not all strictly constitute words). All utterances were edited using digital signal processing techniques\(^1\) to last exactly 350 msec, but the length of the silence placed between the utterances was varied between the conditions. In the no-gap condition, no silence was placed between the utterances. The small gap condition contained 350-msec periods of silence placed between the utterances, and 700-msec gaps were used in the long gap condition. This resulted in the subjects hearing the most words in the no-gap condition, the fewest in the long-gap condition, and an intermediate number in the small-gap condition. The three speech conditions were contrasted with a quiet control. All words were spoken in a male voice, recorded in a random order using SoundEdit software and stored as “snd” resources within Hypercard. In this manner a recording was constructed for each condition, which lasted approximately 20 sec and was repeatedly looped. Sound was delivered via Sony CD 250 headphones at 65 dB(A) as measured by an artificial ear.

---

\(^1\) All sound used in the experiments in this paper were recorded and digitally edited using digital signal processing techniques to 8-bit resolution at a sampling rate of 22 kHz.
Experimental Design

A repeated measures design was used in which the four treatments were blocked, with 16 trials in each condition. The order of presentation of the blocks was randomized between subjects.

Procedure

Subjects were tested individually, seated in a sound-proofed laboratory approximately 0.5 m from the computer’s screen. Before the start of the trial, standard instructions were read by the subject. These informed them of the nature of the recall task and instructed them to ignore any sounds they might hear. In each trial, the nine letters were displayed in random order, as described earlier. Each letter was displayed for 500 msec with an inter-letter interval also of 500 msec. Each list was preceded and followed by a tone. After presentation of the second tone, the word “wait” was flashed on the screen for 10 sec, during which time the subject was expected to rehearse the list. As it has been shown that the disruption occurs in memory and not encoding, it was expected that exposure to the irrelevant speech during a retention interval would result in more marked effects than immediate recall. After this retention interval another tone was played, at which signal the subject was instructed to recall the list in strict serial order. Speech was played throughout the presentation, retention, and recall of the lists. Responses were written on a blank grid comprising rows of nine boxes. The subjects were given 15 sec to recall each list. The experiment was preceded by a five-trial practice session, and the experiment lasted about 40 min in all.

Results

A two-factor ANOVA with auditory condition (four levels) and serial position (nine levels) as factors was used to analyse errors. The mean errors at each serial position are shown in Figure 1. There was a significant effect of condition, $F(3, 69) = 14.36; p < .0001$, and position, $F(8, 184) = 34.96; p < .0001$, with a significant interaction between the two, $F(24, 552) = 1.58; p < .05$. For each condition, the total mean errors out of 16 and standard errors were: 6.51 ± 0.30 for the quiet condition, 7.38 ± 0.31 for the big-gap condition, 8.17 ± 0.30 for the small-gap condition, and 9.20 ± 0.31 for the no-gap condition.

Planned comparisons between the auditory conditions showed there to be significant differences between quiet versus big gap, $F(1, 17) = 4.215; p < .05$, small gap versus no gap, $F(1, 17) = 5.86; p < .05$, and big gap versus no gap, $F(1, 17) = 18.03; p < .0001$, and no significant difference between small gap versus big gap, $p = .07$. The continuum of increased interference from big gap through to no gap was examined using an orthogonal polynomial analysis, which showed there to be a highly significant linear trend, $F(1, 23) = 19.35; p < .0001$.

Inspection of Figure 1 shows a possible reason for the interaction between condition and position—namely, the reduced difference between conditions at Serial Positions 1 to 3. This may also explain the finding that the difference between big gap and small gap did not reach significance, despite the statistical evidence from the polynomial analysis of a continuum of interference as dose increased. A series of post-hoc comparisons were therefore used to examine the difference between the conditions in just the last six serial positions. Significant differences were found between all conditions: quiet versus big
gap, $F(1, 17) = 35.40; p < .0001$; big gap versus small gap, $F(1, 17) = 23.90; p < .0001$; small gap versus no gap, $F(1, 17) = 36.30; p < .0001$.

**Discussion**

The results of the first experiment form a coherent pattern; as the amount of silence between the utterances in an irrelevant speech stream decreases—that is, as the number of words increases—the amount of disruption to serial recall caused by that stream increases: a "word dose" effect. This finding appears to support the hypothesis made earlier that one factor affecting the amount of disruption, possibly by altering the amount of changing state information, is the number of words in the speech stream. It could be argued, however, that the number of utterances has been confounded with the total speech exposure. That is, in the no-gap condition there were many more utterances than in the big-gap condition, because it contained no silence. This design, however, resulted in greater exposure to speech in the former condition than the latter. A further experiment was undertaken, to manipulate both dose and speech exposure thereby avoiding the possibility of attributing the results to word dose in error.
EXPERIMENT 2

It has been argued that word dose, the number of distinct and different units in an irrelevant speech stream, is an important factor in determining the extent of its interference with serial recall. We would not, therefore, attribute the results of Experiment 1 to differential amounts of speech exposure across conditions. Experiment 2 sought to test this proposition by contrasting speech streams with the same number of utterances, but different amounts of total speech exposure: short utterances with long gaps, long utterances with short gaps, and short utterances with short gaps. To avoid the possible confound that long utterances would be phonologically different from short utterances, sound-editing software, using digital signal processing techniques that lengthen sounds but retain essential characteristics such as pitch, was used to construct the long utterances from the short utterances. It was predicted that the stream with short utterances and long gaps would show an equivalent degree of disruption to that with long utterances and short gaps (i.e. the same number of utterances in each), despite the latter providing greater speech exposure. The stream containing the greatest number of utterances (short utterances and short gaps) was expected to produce the most disruption.

Method

Subjects

Twenty undergraduate student volunteers were paid for participating in the experiment. All subjects had English as their first language and reported normal hearing.

Apparatus and Procedure

Items to be recalled were presented serially on the screen of an Apple Macintosh Quadra micro-computer. Lists were the same as for Experiment 1, as was the procedure.

Materials

All three speech conditions consisted of the same five utterances of letter names—B (bee), I (eye), J (jay), N (enn) and Z (zed)—in random order. Two sets of recordings of the letter names were made. In the first, the utterances were timed to last exactly 350 msec. These recordings were then transformed using Sound Designer software to produce a second set of recordings, each of which lasted exactly 700 msec. Qualities of the original recording, such as pitch, were kept exactly the same for the two sets of recordings. It is important to note that these techniques of transformation do not decrease intelligibility, and that the resultant utterance is perceived as natural speech. The speech materials were assembled from either the short (350 msec) or long (700 msec) utterances with either short (350 msec) or long (700 msec) periods of silence between the utterances to produce recordings of speech for three auditory conditions; short word with short gap, long word with short gap, and short word with long gap. In this manner a recording was constructed for each condition which lasted approximately 20 sec and could be repeatedly looped. All words were spoken in a male voice and stored as resources within Hypercard. The three speech conditions were complemented with a quiet control. The same number of utterances was contained in the recordings for the short-word with long-gap and long-word with short-gap conditions, although the total amount of speech, in terms of
time, was greater in the latter condition. The amount of speech exposure was also greater in the short-word with short-gap condition than in the short-word with long-gap condition; however, the former condition contained many more utterances. Sound presentation was as for Experiment 1.

Results

A two-factor ANOVA was carried out on serial recall errors with respect to auditory condition (four levels) and serial position (nine levels). Mean errors at each serial position are displayed in Figure 2. There was a significant effect of condition, $F(3, 57) = 5.89; p < .001$, and serial position, $F(8, 152) = 23.87; p < .0001$ but no significant interaction, $p > 0.5$. The results of planned comparisons between conditions showed there to be significant differences for quiet versus short word with long gap, $F(1, 17) = 3.99; p < .05$, quiet versus long word with short gap, $F(1, 17) = 4.60; p < .05$, short word with long gap versus short word with short gap, $F(1, 17) = 4.85; p < .05$, and long word with short gap versus short word with short gap, $F(1, 17) = 4.22; p < .05$. No significant difference was shown for long word with short gap versus short word with long gap, $p > 0.5$.

![Figure 2: Results of Experiment 2, showing mean errors of serial recall with respect to serial position, contrasting performance in quiet with conditions of irrelevant speech in which the constituent words were 350 msec long, with between-word intervals of either 350 msec (short word with short gap) or 700 msec (short word with long gap), or the words were 350 msec long with 350-msec intervals (short word with short gap).](image-url)
Discussion

The results of Experiment 2 are as predicted by the word dose hypothesis. The two conditions in which there were the same number of utterances in the speech stream (short word with long gap and long word with short gap) showed an equivalent amount of disruption of serial recall despite there being a greater amount of total speech exposure in the latter condition. The condition in which the speech stream contained many more utterances (short word with short gap) produced significantly more disruption than either of the other two speech conditions. Taking these results in conjunction with those of Experiment 1, we can now state more confidently that the total duration of speech is not the important factor in the dose effect. Rather, as hypothesized, a major determinant of the degree of disruption appears to be the number of distinct utterances in the speech stream. In summation, we have observed a dose effect—that is, the greater the “dose” of units in a speech sequence, the greater the disruption of serial order information that the sequence will produce.

A further prediction that may be derived from the changing-state hypothesis is that there should be an interaction between changing state and word dose. It is suggested that disruption of serial recall is dependent on item-to-item change, hence steady-state material produces little disruption. Therefore, increasing the word dose in a steady-state sound should have little effect on the amount of disruption produced. Certainly, any increase in disruption should not be as marked as that shown in the preceding experiments by changing-state sound. A third experiment was undertaken to examine this prediction.

EXPERIMENT 3

The third experiment was designed to investigate the manipulation of word dose in irrelevant speech consisting of the same repeated utterance (steady state) and speech that contained differing utterances (changing state). It is proposed, from the changing-state hypothesis, first, that changing-state speech would produce significantly more interference than steady state. Second, increasing the dose in steady-state speech should not increase the amount of changing-state information and therefore would not increase the disruption of serial recall. Hence there should be an interaction between changing state and word dose.

Method

Subjects

Twenty-seven undergraduate student volunteers were paid for participating in the experiment. All subjects had English as their first language and reported normal hearing.

Apparatus and Procedure

Items to be recalled were presented serially on the screen of an Apple Macintosh Quadra microcomputer. Lists were the same as for Experiment 1, as was the procedure.
Two types of speech sequences were constructed. The first consisted of a repeated (steady-state) utterance: the syllable B (bee); the second consisted of the same letter names used in Experiment 1 in random order (changing state): B (bee), I (eye), J (jay), N (enn), and Z (zed). All utterances were spoken in a male voice and edited using digital signal-processing techniques to last exactly 350 msec. Word dose was manipulated in both of the speech types by inserting either a 0-msec (high word dose) or 700-msec (low word dose) gap of silence between the utterances. This resulted in the production of four speech conditions: steady/low, steady/high, changing/low and changing/high. The recording for each condition lasted approximately 20 sec and could be repeatedly looped. These recordings were stored as resources within Hypercard. The four speech conditions were contrasted with a quiet control. Sound presentation and delivery were as for Experiment 1.

Results

The errors were analysed first using an ANOVA with two factors: condition and serial position. Mean errors at each serial position are displayed in Figure 3. There was a significant effect of condition, \( F(4, 104) = 9.20; \ p < .0001 \), and position, \( F(8, 208) = 63.05; \ p < .0001 \), and no interaction, \( p > .1 \). For each condition, the total mean errors out of 16 and standard errors were: 6.48 ± 0.29 in the quiet condition, 6.78 ± 0.27 in the steady/low condition, 6.95 ± 0.29 in the steady/high condition, 7.51 ± 0.28 in the changing/low condition, and 9.17 ± 0.27 in the changing/high condition.

Planned comparisons showed there to be significant differences between quiet and the two changing state conditions: quiet versus changing/low, \( F(1, 26) = 4.21; \ p < .05 \), and quiet versus changing/high, \( F(1, 26) = 29.04; \ p < .0001 \). No significant differences were found between quiet and the two steady-state speech conditions: quiet versus steady/low, \( p > .1 \), and quiet versus steady/high, \( p > .1 \).

The data for the quiet condition were then removed, and a 2 (speech type-changing vs. steady state) × 2 (dose) ANOVA was used to analyse the data from the speech conditions. There was a significant main effect of speech type, \( F(1, 26) = 20.99; \ p < .0001 \), and dose, \( F(1, 26) = 6.62; \ p < .05 \). Importantly, there was also a significant interaction between dose and speech type, \( F(1, 26) = 4.95; \ p < .05 \). Planned comparisons showed that a significant dose effect was produced only with changing-state speech, changing/low versus changing/high, \( F(1, 26) = 12.27; \ p < .01 \); no significant difference was found between the two steady-state conditions, steady/low and steady/high, \( p > .1 \). Analysis of the effect of speech type showed that significantly more errors were produced in the changing/high condition than in either of the conditions of steady-state speech: changing/high versus steady/low, \( F(1, 26) = 25.33; \ p < .0001 \), changing/high versus steady/high, \( F(1, 26) = 21.88; \ p < .0001 \), whereas changing/low speech did not produce more errors than either of the steady state conditions: changing/low versus steady/low, \( p > .1 \); changing/low versus steady/high, \( p > .1 \).

Discussion

The predictions drawn from the changing-state hypothesis concerning the relationship between word dose and changing-state information were fulfilled in the results of Experiment 3. First, there was a significant interaction between dose and speech type, with only
the changing-state speech producing a significant dose effect. Second, significantly more errors were produced under conditions of changing-state speech when presented at a high dose. It would appear that the reduction in changing-state information caused by presenting the changing-state speech at a low dose resulted in this condition producing no more disruption than steady-state speech. Changing-state speech presented at a low dose did, however, produce more errors than when the task was performed in quiet, unlike either of the steady-state conditions.

The first two experiments demonstrated that as the number of discrete words in an irrelevant speech source increases, the greater the disruption to serial recall caused by that speech. This was suggested to be due to an increase in changing-state information in the sound. The third experiment supports this conclusion with the demonstration of an interaction between word dose and whether the sound consisted of changing- or steady-state items.

A weakness of the preceding conclusions, however, is that the word-dose effect might also be explained within the working memory model. An important extrapolation of the working memory account is that, if the irrelevant speech effect is due to phonological confusions, speech that is phonologically more similar to the items to-be-recalled should produce more interference. Evidence for this position was provided by Salamé and Baddeley (1982, Experiment 5), who contrasted the effects of three irrelevant speech sources on the
serial recall of visually presented digits: (1) digits identical to the visual presentation (semantically similar condition), (2) words that contained the same phonemes as the digits in different order—for example, tun, gnu, tree, sore (phonologically similar condition), or (3) phonologically dissimilar words—for example tennis, tipple, wicket (phonologically dissimilar condition). It was found that the semantically similar and phonologically similar conditions produced an equivalent degree of disruption, which was significantly greater than that produced by the phonologically dissimilar condition (all speech conditions produced significantly more errors of recall than when the task was performed in quiet).

One of the conclusions drawn from these findings was that “(the results) seem to offer strong support for the view that the amount of disruption by irrelevant speech is determined by the phonological similarity between the material being rehearsed and the irrelevant distracting material” (Salamé & Baddeley, 1982, p. 160). A second conclusion was that the ordering of the phonologically similar phonemes was inconsequential. This argument was drawn from the finding that digits, which were identical to the visual material, and the phonologically similar speech, which contained the same phonemes as the digits in a different order, produced similar degrees of disruption.

The conclusions drawn by Salamé and Baddeley (1982) could lead to the suggestion that it was not changing-state information per se that produced the increased interference shown by high-dose sound. Rather, it could be argued that increasing the word dose simply increases the number of confusions within the phonological store. Equally, the interaction demonstrated in Experiment 3 could be the result of the changing-state speech containing more phonemes that were also present in the to-be-remembered list. That is, the changing-state sound was phonologically more similar to the to-be-remembered items. The changing-state state speech contained the syllable $B$ (bee) (containing the vowel sound $ee$ present in $T$ in the visual list), $J$ (jay) ($ay$ present in $K$ in the visual list), $I$ (eye) ($eye$ present in $Y$ in the visual list), and $N$ ($enn$) & $Z$ ($zed$) ($ed$ present in $M$ in the visual list). In comparison, the steady-state speech only contained the letter name $B$ (bee). It may be, therefore, that the changing-state speech produced greater disruption because it created more phonemic confusions in the phonological store. Furthermore, increasing the dose of the changing-state speech would produce proportionally greater potential for phonemic confusions than the higher dose in the steady-state speech.

It should be noted that indirect evidence, from studies of the irrelevant speech effect, is not very supportive of the hypothesis that the effect is related to the phonological similarity between heard and seen material. A number of studies using very different types of speech have produced broadly similar degrees of disruption—such as narrative (Colle, 1980; Jones et al., 1990), speech in a language that is not understood by the subject (Jones et al., 1990), sung passages (Morris, Jones, & Quayle, 1989), and reversed speech (Jones et al., 1990). Furthermore, the findings discussed earlier that certain types of non-speech produce similar disruption to speech (e.g. Jones & Macken, 1993) cast even more doubt on the proposal that the phonological similarity between the speech and the to-be-remembered items is an important factor in the interference. Direct evidence contrary to the results of Salamé and Baddeley (1982, Experiment 5) comes from another study in which, although stimuli different to the original experiment were used, an effect of the phonological similarity between heard and seen items was not found (Jones & Macken, 1995b, Experiment 2).
Despite these findings, it may be feasible to object to the conclusions of the present series on the basis of the phonological similarity hypothesis detailed above. It seemed desirable therefore to attempt a direct replication of the experiment that first demonstrated the effect (Salamé & Baddeley, 1982, Experiment 5) to examine whether the phonological similarity effect is as robust as the dose effect demonstrated in the preceding three experiments.

EXPERIMENT 4

The effects of irrelevant speech have been explained within the working memory model of Baddeley and colleagues as a result of phonological confusions between heard and seen material (e.g. Salamé & Baddeley, 1982). It has been noted that the word-dose effect demonstrated in the previous three experiments could also be accommodated within this model. Increasing the word dose may simply increase the number of phonological confusions. Furthermore, it is possible that the speech containing a single repeated utterance (steady state) used in Experiment 3 was phonologically less similar to the visual stimuli than the speech containing several words (changing state). This difference in phonological similarity could therefore be argued to account for the finding that steady-state speech did not produce a dose effect. In order to be certain that phonological similarity was not an important factor in the results of Experiment 3, a replication was attempted of the original experiment, which demonstrated the effect (Salamé & Baddeley, 1982, Experiment 5).

Method

Subjects

Eighteen undergraduate student volunteers were paid for participating in the experiment. All subjects had English as their first language and reported normal hearing.

Apparatus and Materials

Lists of items to be recalled were presented serially on the screen of an Apple Macintosh II microcomputer. Lists were constructed from the random arrangement of the digits one to nine, as in Salamé and Baddeley (1982), with the constraint that a digit could not appear in the same serial position in two consecutive lists. The lists were stored and presented within a Hypercard environment.

The three types of irrelevant speech used were similar to those used by Salamé and Baddeley (1982), but with the safeguard that all words were timed to exactly 750 msec using editing software. Material in the semantically similar condition was identical to that used in the visual lists (the digits one to nine) and phonologically similar material contained the same phonemes arranged into different pairs: tun, gnu, wee, sore, thrive, fix, heaven, and sign. As in Salamé and Baddeley (1982), the digit seven, and its rhyme (heaven) were pronounced with the intention of approximating a mono-syllabic word (sev’n and ‘eav’n), to conform with the rest of the utterances in their respective speech streams. The phonologically dissimilar words comprised tennis, jelly, tipple, double, tunnel, hackle, valley, pickle, and wicket. The three speech conditions were contrasted with a quiet control. All words were spoken.
in a male voice, recorded using SoundEdit software and stored as “snd” resources within Hypercard. The words were recorded in a random order with a gap of 250 msec between each one; this, combined with the fact that all words were of the same length, ensured that the same word dose was contained in each speech condition. In this manner a recording was constructed for each condition, which lasted approximately 20 sec and was repeatedly looped. Sound delivery was as for Experiment 1.

**Experimental Design**

A repeated-measures design was used in the experiment. The four treatments were blocked, with 20 trials in each condition, and the order of presentation of the blocks was randomized between subjects.

**Procedure**

The procedure was as for Experiment 1, except that in order to replicate the methodology used by Salamé and Baddeley (1982, Experiment 5), no retention interval was used. Subjects were required to recall the items immediately following presentation.

**Results**

Errors were analysed using a two-factor ANOVA with acoustic condition (four levels) and serial position (nine levels) as factors. There were highly significant effects of condition, \( F(3, 51) = 15.44; p < .0001 \), and serial position, \( F(8, 136) = 33.56; p < .0001 \). Planned comparisons between the main effect of acoustic conditions showed there to be no significant differences between any of the speech conditions (\( p > 0.05 \) in all cases), but significant differences between each of the speech conditions and quiet: quiet versus semantically similar, \( F(1, 17) = 20.99; p < .0001 \), quiet versus phonologically similar, \( F(1, 17) = 39.93; p < .0001 \), quiet versus phonologically dissimilar, \( F(1, 17) = 27.07; p < .0001 \). There was also a significant interaction between the two factors, \( F(24, 408) = 2.15; p < .01 \). There are two possible accounts for the interaction that is displayed in Figure 4. First, phonologically similar words produce more disruption than the other speech types at Positions 3, 4, and 5; however, of the three contrasts only that at Position 5 just reaches significance, and then only at the 5% level. Testing of an isolated point among many is not strictly legitimate, and it therefore seems likely that the significant difference is a chance result. A second possible reason for the interaction is the convergence of all speech conditions with the quiet condition at Serial Positions 1, 2, and 8. The latter explanation seems more likely, given that an analysis of the amount of variance accounted for by differences across conditions is extremely small, \( \eta^2 = 0.07 \), in comparison to that accounted for by change across serial position, \( \eta^2 = 0.29 \).

**Discussion**

In showing no increased disruption of serial recall by irrelevant speech containing the same phonemes as the to-be-remembered material, Experiment 4 failed to replicate the findings of Experiment 5 of Salamé and Baddeley (1982). The phonologically similar
material did, however, show a (largely non-significant) tendency to produce more disruption at three serial positions. It would seem unwise to attach too much significance to this, considering the analysis of the amount of variance accounted for by change across condition and serial position. Second, if the increase in disruption was due to phonological similarity, we would have expected the semantically similar speech to have shown a similar trend.

The failure to find a phonological similarity effect, however, supports the evidence, discussed earlier, from other irrelevant speech studies, which concluded that phonological similarity between heard and seen items was not an important factor in the disruption of serial recall (e.g. Jones & Macken, 1995b, Experiment 2). We would suggest, from this evidence, that the sole experimental demonstration of a phonological similarity effect available (Salamé & Baddeley, 1982, Experiment 5) was, in fact, the result of a Type II sampling error. The result allows us to be more confident that the word-dose effect demonstrated in previous experiments was not the product of increasing the phonological similarity between heard and seen material.

Despite supporting predictions, a possible weakness of Experiment 4 is that only the words in the phonologically dissimilar condition were disyllabic. It could be argued that
dose may be increased, not only by increasing the number of words in the irrelevant speech, but also by increasing the number of syllables in those words. We could, therefore, have observed, in the results of Experiment 4, an effect of both dose and phonological similarity. It could be suggested that as the phonologically dissimilar words were disyllabic, this condition provided a greater dose than the other speech conditions, hence producing an increased degree of disruption, equivalent to that produced by phonological similarity. The results of Experiment 4 could therefore be revealing two mechanisms of interference rather than just the one. It should, however, be noted that if this position is taken, a dose effect should have also been shown by the disyllabic words in Salamé and Baddeley’s original investigation. Furthermore, in another experiment in the same series (Salamé & Baddeley, 1982, Experiment 4), it was shown that multisyllabic words did not produce more interference than did speech containing monosyllabic words.

Despite that evidence, a final experiment was designed to check that disyllabic words do not produce more disruption than monosyllabic words. The aim of Experiment 5, therefore, was to support our contention that the results of Experiment 4 were due not to a manifestation of both phonological similarity and dose effects, but solely to a failure to find a phonological similarity effect.

EXPERIMENT 5

This experiment contrasted the effect of speech containing words that were phonologically identical to the visual stimuli with speech containing either monosyllabic or disyllabic phonologically dissimilar words. All words were edited to exactly the same duration, and therefore the same word dose was maintained in all conditions. If dose increases as the number of syllables in a word increases, disyllabic words should produce more disruption of serial recall than monosyllabic words.

Method

Subjects

Eighteen undergraduate student volunteers were paid for participating in the experiment. All subjects had English as their first language and reported normal hearing.

Apparatus and Procedure

Items to be recalled were presented serially on the screen of an Apple Macintosh II micro-computer. Lists were the same as for Experiment 4, as was the procedure.

Materials

Two of the types of irrelevant speech used—semantically similar (the digits one to nine) and disyllabic phonologically dissimilar (tennis, jelly, tipple, etc.)—were identical to those used in Experiment 4. The third type comprised monosyllabic phonologically dissimilar words: bed, sap, pick, stop, neck, tip, nut, cat, and duck. Word length and rate of presentation were as for Experiment 4, again ensuring the same word dose in each condition. The three speech conditions were contrasted with a
quiet control. All words were spoken in a male voice, recorded using SoundEdit software, and stored as “snd” resources within Hypercard. The words were recorded in a random order, with a gap of 250 msec between each one. In this manner a recording was constructed for each condition, which lasted approximately 20 sec and could be repeatedly looped. The method of presenting sound was as for Experiment 4.

Results

The mean errors at each serial position are illustrated in Figure 5. There were significant effects of auditory condition, $F(3, 51) = 7.25; p < .001$, and serial positions, $F(8, 136) = 52.62; p < .0001$, with no interaction between the two. Planned comparisons showed there to be no differences among the speech conditions, $p > 0.1$ in all cases, and significant differences between each of the speech conditions and quiet: quiet versus semantically similar, $F(1, 17) = 13.95; p < .001$, quiet versus monosyllabic dissimilar, $F(1, 17) = 12.45; p < .001$, quiet versus disyllabic dissimilar, $F(1, 17) = 16.62; p < .001$.

FIG. 5. Results of Experiment 5, showing mean errors of serial recall with respect to serial position, performance in quiet was contrasted with performance under conditions of irrelevant speech that contained words that were either identical to the visual stimuli (digits), or were phonologically dissimilar and either monosyllabic (e.g. pet, cat, duck), or disyllabic (e.g. tennis, valley, wicket).
Discussion

The results of Experiment 5 show that disyllabic words do not produce more disruption than do the monosyllabic phonologically dissimilar words. This suggests that dose is not affected by the number of syllables in the words comprising the irrelevant speech. As the purpose of this experiment was to investigate an alternative explanation for the results of Experiment 4, it is useful to consider the two sets of results together. Experiment 4 showed no effect of the phonological similarity between an irrelevant speech source and the visually presented items of a serial recall task. This finding is contrary to the results of Salamé and Baddeley (1982, Experiment 5) but harmonizes with the results of several other irrelevant speech studies (e.g. Jones & Macken, 1995b, Experiment 2). It was noted, however, that the results may have been due to the action of two mechanisms. If it is assumed that dose increases as the number of syllables in the constituent words of the irrelevant speech increases, disyllabic words would provide a greater dose of speech and hence would be expected to create more interference. Experiment 5, however, shows that monosyllabic and disyllabic phonologically dissimilar words produce equivalent amounts of disruption, and, furthermore, they produce interference at a level comparable to that produced by words that are phonologically similar to the visually presented digits. This result does not concur with the hypothesis that the findings from Experiment 4 were the result of both phonological similarity and dose. It does, however, provide further support for the conclusion that the results of Experiment 4 were due solely to a failure to find an effect of the phonological similarity between heard and seen items.

GENERAL DISCUSSION

The results of the present series form a clear and coherent pattern. Experiment 1 showed there to be a monotonic relationship between the number of words in an irrelevant speech source, the word dose, and its propensity to disrupt serial recall. Experiment 2 showed that this effect was independent of the overall speech exposure provided by the speech. The results of Experiment 3 suggested that the effect of word dose was due to an increase in the amount of changing state information contained in the speech, as steady-state speech did not produce a significant dose effect. It was noted, however, that this result may be due to the changing-state speech containing more phonemes in common with the to-be-remembered material, resulting in greater phonemic confusion. This possibility was examined in Experiment 4, which attempted to replicate the finding that speech that is more phonologically similar to the visual stimuli produces more disruption (Salamé & Baddeley, 1982, Experiment 5). No effect of phonological similarity was found, but it was noted that the result may have been confounded by the fact that only the phonologically dissimilar words were disyllabic. Experiment 5 examined the possibility that dose may be increased by increasing the number of syllables in the words contained in an irrelevant speech source. Monosyllabic and disyllabic words produced equivalent degrees of disruption, implying that dose (in discrete presentation at least) is dependent on the number of words and not the number of syllables contained in those words. Experiment 5 increases our confidence that the results of Experiment 4 were due solely to a failure to find a phonological similarity effect. This, in turn, supports the conclusions drawn
from the first three experiments that increased disruption produced by high-dose speech is the result of an increase in changing-state information within the speech, not an increase in phonological confusions between the speech and the visual material.

**Word Dose in Continuous Speech.** Whilst it has been argued that the results point to a word-dose effect of irrelevant speech, we have been careful to point out throughout the paper that the experiments have only investigated the effect of discrete presentations of words. We would also expect to see a dose effect of continuous speech; however, we would not expect that dose would correspond to the number of written dictionary (WD) words. A number of studies have suggested that the parsing of continuous speech is facilitated by the location of stressed syllables (e.g. Cutler, 1989; Cutler, Mehler, Norris, & Segui, 1986; Grosjean & Gee, 1987). It is beyond the scope of the current discussion to consider the merits of these arguments, but it may be that dose in continuous speech would be found to correspond to the number of parsed units or phonological words (Grosjean & Gee, 1987) created by such a segmentation process. This possibility has relevance to the finding that serial recall is disrupted by irrelevant speech in a language that was not understood by the subject (e.g. Jones et al., 1990). Although it has been argued that speech segmentation processes are different in different languages (Cutler, 1993), native English speakers will parse an unfamiliar language in the same way as they would parse English. Therefore, we would still expect to see a dose effect, as the unfamiliar speech would be segmented into discrete units, albeit different units from those that would be produced by a native speaker of the language.

**Implications for Auditory Short-term Memory.** The irrelevant speech effect was suggested by Baddeley and Andrade (1994) to be one of the four primary sources of evidence for the existence of the phonological store. It was noted, however, that a deeper understanding of the operations of these effects was necessary for a more precise model of phonological short-term memory. We believe that the findings of the present study, in conjunction with other reported findings, go some way towards providing that understanding of the irrelevant speech effect, and may therefore begin to answer this call for a more precise model.

The major novel finding of the experimental series is that the number of discrete segmentable units in an irrelevant sound source is an important factor in its propensity to disrupt serial recall. This was interpreted as adding to the considerable body of literature that suggests that changing-state information is largely responsible for the irrelevant “speech” effect. The conclusion to be drawn from these arguments is that changing-state information has an important role in the cognitive processing of both auditory and visual material.

A secondary finding was that, contrary to the results of Salamé and Baddeley (1982, Experiment 5), the degree of phonological similarity between heard and seen material did not affect the disruption. This provided at least some evidence to suggest that the effect is unlikely to be the result of phonemic confusions. This secondary result may provide some insight into the finding that instrumental music produces some disruption of serial recall (Salamé & Baddeley, 1989). The conclusion drawn from the result was that music is in some way “speech-like” enough to gain access to the phonological store. However, it is
difficult to understand how instrumental music might create the phonemic confusions suggested to be integral to the irrelevant speech effect. An alternative proposal is that even continuous music is perceptually segmented into units that differ from one to the next. Hence, in contrast to a steady state continuous sound such as white noise, it has the propensity to disrupt serial recall.

We would not argue that the results discussed above are sufficient to form the basis for a more precise model of short-term phonological memory. A number of questions have yet to be addressed concerning the exact mechanisms that produce a changing-state effect of irrelevant speech. For instance, why should changing-state speech be afforded a greater degree of processing than steady-state, and exactly what parameters define the degree of changing-state information in a stimulus source? It is apparent, however, that major revisions of Baddeley and colleagues’ theory of the phonological store would have to be made to accommodate these findings from the study of irrelevant speech. At the very least, the assumptions that speech is the only material to gain automatic access to the store and that the irrelevant speech effect is the result of phonological confusions between heard and seen items would have to be dropped. Furthermore, considerable additions would have to be made to the current model to account for the importance of changing-state information.

REFERENCES


Original manuscript received 5 May 1994
Accepted revision received 1 November 1995