The Phonological Store of Working Memory: Is It Phonological and Is It a Store?

Dylan M. Jones, William J. Macken, and Alastair P. Nicholls
Cardiff University

The phonological store is critically evaluated. Three experiments test the prediction that the effect of irrelevant sound and the effect of phonological similarity each survive the action of articulatory suppression but only when presentation of to-be-remembered lists is auditory, not visual. No evidence was found to support the interaction predicted among irrelevant speech, modality, and articulatory suppression. Although evidence for an interaction among modality, phonological similarity, and articulatory suppression was found, its presence could be diminished by a suffix, which is an acoustic, not a phonological factor. Coupled with other evidence—from the irrelevant sound effect and errors in natural speech—the action attributed to the phonological store seems better described in terms of a combination of auditory–perceptual and output planning mechanisms.
phonologically dissimilar ones (Baddeley, 1966; Conrad, 1964; Conrad & Hull, 1964; Hintzman, 1967; Wickelgren, 1965a, 1965b). A similar pattern of error emerges irrespective of whether the items are presented in the visual or auditory modality (Baddeley, 1968; Longoni, Richardson, & Aiello, 1993), hence the confusion between items is within memory at some level of abstraction beyond the physical character of the stimulus, at the level of phonemic representation. Arguably, the phonological similarity effect is one of the most significant “benchmark” phenomena in short-term memory research (Farrell & Lewandowsky, 2002; Nairne & Kelly, 1999), and most models incorporate some kind of mechanism to deal with its various empirical manifestations (see, e.g., Burgess & Hitch, 1996; Farrell & Lewandowsky, 2002; Nairne, 1990; Neath, 1999, 2000; Page & Norris, 1998). The working memory model attributes errors of phonological similarity to the phonological store, “because the phonological store relies purely on a phonological code; similar codes present fewer discriminating features between items, leading to impaired retrieval and poorer recall” (Baddeley, 1992, p. 9). The critical role of the interaction between phonological similarity and modality should not be understated; indeed, Baddeley, when speaking of this interaction, clearly considered it to be paramount: “The particular pattern of results obtained was crucial to separating the two components of the articulatory loop, the phonological store and the articulatory control process. Had the results not worked out in this way, it would have been necessary to modify the model quite seriously” (Baddeley, 1986, p. 257).

The second line of evidence about the phonological store relates to the irrelevant sound effect, that is, the tendency for to-be-ignored auditory stimuli to disrupt the serial recall of verbal sequences, even when the sequences are presented visually (see, e.g., Colle 1980; Colle & Welsh, 1976; Jones, Madden, & Miles, 1992; LeCompte, 1996; Neath, 2000; Salamè & Baddeley, 1982, 1986, 1990; see Ellermeier & Zimmer, 1997, for evidence of its effect size and stability). According to the working memory model, this phenomenon exemplifies an additional basic quality of the phonological store, that of obligatory and direct access. That is, irrelevant auditory stimuli enter the store automatically, with entry being obligatory in the sense that it is outside the control of the individual. Interference takes place within the phonological store, with the degree of disruption by irrelevant speech being the product of the similarity between the irrelevant speech and the to-be-remembered sequences. Although access by auditory stimuli to the store is direct, it is necessary for visual items to undergo a grapheme-to-phoneme conversion process that is the domain of the phonological loop. As long as the loop is free to act, then irrelevant sound in phonological form will coexist with list items also in phonological form within the store.

The action of irrelevant speech and phonological similarity can be revealed further by the interplay of articulatory suppression and the modality of presentation of the to-be-remembered material. The working memory model predicts that under conditions of articulatory suppression, irrelevant speech will have a damaging effect on serial recall when the presentation of the list is auditory but not when visual (Hanley & Broadbent, 1987). When both are auditory, the irrelevant speech and the list items enjoy direct access to the store. A very different pattern of results is expected when presentation is visual; the loop is taken up by the activities associated with suppression and therefore is not available for grapheme-to-phoneme conversion. The evidence is not controversial for the visual case: Under conditions of articulatory suppression, the irrelevant speech effect may be either completely attenuated (Divin, Coyle, & James, 2001; Miles, Jones, & Madden, 1991; Salamè & Baddeley; 1982; using vocalized suppression) or markedly though not completely attenuated (Macken & Jones, 1995; using silently mouthed suppression). For the auditory case, it is expected that direct access will carry both the irrelevant and to-be-remembered materials into the phonological store even when the loop is occupied by articulatory suppression. The irrelevant sound effect will be evident, therefore, even during articulatory suppression as long as the presentation modality of the list is auditory. Although this prediction flows logically from the working memory framework, evidence in its favor is both sparse and inconsistent (e.g., Hanley, 1996; Hanley & Broadbent, 1987). Given this background and its importance to the working memory framework, the interaction needs further examination.

The second class of interaction predicted by the working memory model is that among articulatory suppression, modality, and phonological similarity. Using broadly the same logic as for the irrelevant sound effect, the phonological similarity effect is said to arise from the action of the phonological store, but in the presence of articulatory suppression this should only be exhibited by the auditory modality, not the visual modality. In contrast to the interaction with irrelevant sound, evidence on this point is largely consistent (Baddeley, Lewis, & Vallar, 1984; Murray, 1968; Peterson & Johnson, 1971). However, details of the interaction are missing (particularly those relating to serial position), and these may be essential to a full understanding of the phonological store construct. In the experiments that follow we attempt to redress this shortcoming.

The series begins by reexamining evidence on the combined effect of articulatory suppression, irrelevant speech, and list modality. This is done on two grounds; first, there is inconsistency of the evidence in the small volume of research undertaken, and second, those few studies that have been performed have used inappropriate methodology.

**Experiment 1**

In Experiment 1 we examined the prediction that articulatory suppression will impair the expression of the effect of irrelevant sound but only when the to-be-remembered lists are presented visually. Unlike most previous studies, in the current experiment we made the comparison between auditory and visual modalities within a single study using a repeated-measures design. In addition, conditions where we present the to-be-remembered lists in the auditory modality embody a number of precautions not used in previous studies. Our aim was to minimize the possible contaminating role of factors—largely absent from its visual analogue—involving in partitioning the to-be-remembered from the to-be-ignored. Such effects can be broadly regarded as impairing encoding of the to-be-remembered lists, an effect very distinct from the usual effect on irrelevant sound with visual list presentation.

The first attempt to provide experimental evidence for the intricate interplay of irrelevant speech, suppression, and modality predicted by the model hangs by an insubstantial thread. Only one study has shown the predicted effect with auditory lists, a study by
Hanley and Broadbent (1987). However, even within that report the evidence is not consistent, in that in the first experiment in their series they found an outcome consistent with a directly contrary theoretical position, namely that the irrelevant sound effect was abolished by articulatory suppression with auditory presentation. In addition, the method of Hanley and Broadbent’s study may have given rise to a spurious irrelevant sound effect. Each irrelevant sound was presented in synchrony with a to-be-recalled stimulus, and therefore it seems possible that the irrelevant sound effect they obtained could have been due to an effect at encoding. Evidence that to-be-remembered sounds are intelligible when they are used in a task that involves listening alone does not diminish the possibility that encoding effects are at play. Arguably, the difficulty of encoding will become more evident as the task is made more demanding by the presence of a memory load. Interpretation of Hanley and Broadbent’s study is further complicated by the fact that the auditory task was studied in isolation, without comparable manipulations of a visual task. This difficulty is shared with a more recent attempt to show the same effect (Hanley & Bakopoulou, 2003) but with the further complication of a ceiling effect.

In the classical form of the irrelevant sound paradigm (see e.g., Colle, 1980; Colle & Welsh, 1976), encoding effects are avoided by having the to-be-remembered lists presented visually. If, in any auditory version of the paradigm, each irrelevant token is presented only contemporaneously with the to-be-remembered item, serial recall will be impaired; however, the effect in this case is likely to be one of encoding, not the more usual irrelevant sound effect. It seems likely that this encoding effect could survive articulatory suppression, but the effect of irrelevant sound would not be the usual one but instead be a product of masking.

The classic effect found with visually presented items is not an effect at encoding; rather, it is an effect from the rehearsal process at a postencoding stage (Salamé & Baddeley, 1982; see also Beaman & Jones, 1998; Miles et al., 1991). That this action is not an effect at encoding but is an effect on rehearsal is suggested both by the fact that disruption from irrelevant items interleaved with the to-be-remembered items is the same as when they are presented simultaneously (Salamé & Baddeley, 1982) and that the effects are broadly similar when the irrelevant speech is confined to the period of rehearsal, as when it is confined to the period of presentation (e.g., Miles et al., 1991).

Clearly, the danger of encoding effects in the auditory modality can be minimized if the to-be-ignored and to-be-remembered stimuli are made distinct from one another in pitch, timbre, or spatial position, or by changing the timing so that they do not coincide. A considerable additional advantage in isolating the to-be-remembered stimulus sequence can be obtained by capitalizing on the very powerful grouping tendencies in auditory perception, collectively known as auditory streaming (see Bregman, 1990; Warren, 1982). For example, irrelevant sequences presented at a higher rate of presentation than the items in the list will promote the fission of the to-be-remembered and to-be-ignored sequences. Likewise, starting the irrelevant stream before the list will capitalize on the tendency for stable streams to take some seconds to establish properly (see Nicholls & Jones, 2002a, 2002b). If one stream is already set up and stable when extraneous information—sharing neither timing nor timbre with the stream—is introduced, the two streams will be relatively easy to partition. Increasing the number of irrelevant items has the added attraction that it will lead to a very appreciable increase in the magnitude of the irrelevant sound effect (the “token dose effect”; see Bridges & Jones, 1996). In the case of Hanley and Broadbent’s (1987) experiments, the number of irrelevant tokens was, necessarily, the same as the number of to-be-remembered stimuli (six), which, compared with typical experiments in which visual presentation has been used, is a rather small number. Again, the fact that so few tokens produced a significant effect suggests that the irrelevant speech effect found by Hanley and Broadbent is not typical.

Additional methodological issues center on a possible masking effect of the articulatory suppression when it is spoken. Two effects of spoken articulatory suppression may be distinguished hypothetically; first, loud vocalization could mask the irrelevant sound; however, this effect can be reduced by only requiring whispered or mouthed articulation. The second effect is that vocalization itself becomes a source of irrelevant sound. Vocalized suppression is more disruptive than mouthed suppression, but the additional disruptive effect of vocalization is not the result of the speech sound generated thereby (an additional irrelevant sound effect, if you will) because the effect is just as evident when feedback to the ears is masked by noise (Macken & Jones, 1995).

It is plausible that the additional disruption is the result of increased complexity in the articulatory gestures accompanying vocalization (but see Gupta & MacWhinney, 1995, for a different view). Given both the foregoing considerations, throughout the current series the procedure we adopted was one of asking the participant to whisper the suppression and, with their permission, monitor compliance with instructions.

In Experiment 1 we contrasted auditory and visual versions of a serial short-term memory task. We were careful to make the character of the lists in the two modalities as similar as possible. A list length of seven items was adopted, in keeping with the claim that list lengths above this rely less on phonological coding (see Baddeley, 1986) while at the same time meeting the exigencies of task sensitivity, particularly in view of the very marked depression in performance levels that can be expected when articulatory suppression is introduced. The irrelevant sound was presented in a steady-state form (repeated token) and changing state (sequence of changing tokens). That repeated tokens produced less disruption than changing tokens is a highly consistent finding and may be regarded as the empirical signature of the irrelevant sound effect (see Jones, 1993, and below). The action of irrelevant sound was examined in conjunction with that of articulatory suppression on auditory and on visual lists.

Method

Participants. Forty undergraduate students from the School of Psychology at Cardiff University were given course credits for their participation. All were native English speakers reporting normal hearing and normal or corrected-to-normal vision.

Apparatus and materials. To-be-remembered items were chosen from the digits 1–9. For auditory presentation, a female voice was recorded in a monotone at a pitch corresponding to a fundamental frequency of approximately 210 Hz. For visual presentation, digits were presented in the center of a computer monitor in a 72-point Times Roman font. All spoken items were sampled with a 16-bit resolution, at a sampling rate of 48 kHz using Sonic Forge 5.0 software (Sonic Foundry, Inc., Madison, WI, 2000). Using the same software, the spoken digits were compressed digitally to 250 ms without altering acoustic features such as pitch.
Both auditory and visual sequences comprised seven to-be-remembered digits presented at a rate of one item a second. In the case of visual presentation, items were on for 250 ms and off for 750 ms. Items were ordered quasi-randomly with the precaution that no item was repeated within a trial and that consecutive digits did not follow a familiar pattern. Irrelevant items were chosen from the letters a through h and recorded by the same female speaker who recorded the to-be-remembered items. However, using Sonic Forge 5.0 software, for the irrelevant sequence, the pitch of each item was lowered by five semitones below that of the to-be-remembered items. The purpose of the pitch change was to promote fission of the auditory streams. Each irrelevant item was compressed digitally to 190 ms without further changing their pitch. Irrelevant items were presented with an onset-to-onset interval of 250 ms. In the repeated condition, the syllable a was repeated throughout the trial. In the changing condition, the alphabetic sequence a through h was presented in a random order.

For lists in both modalities, the irrelevant sound was presented at a rate of four items per second. The timing of the irrelevant stimuli in relation to the to-be-remembered stimuli was the same in both modalities. The cycle of irrelevant items in relation to each member of the to-be-remembered list was as follows: The first irrelevant token began 60 ms after the onset of the to-be-remembered item (this offset was designed to promote asynchrony and thereby contribute to fission), then the irrelevant item lasted 190 ms, and an additional three irrelevant items followed at 250-ms intervals before the next to-be-remembered item was presented. This cycle was repeated until the end of the to-be-remembered list. No irrelevant item was presented after the to-be-remembered list, to minimize the chance of a suffix effect (see below for further discussion). In the case of auditory lists, the irrelevant sequences were presented monophonically (as in Nicholls & Jones, 2002a, 2002b), with the irrelevant sequence and the to-be-remembered sequence being presented to both ears. The same relative timing between to-be-remembered and irrelevant items was preserved in the visual modality by employing PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993).

Each trial commenced with an introductory period of approximately 7 s before the onset of the to-be-remembered sequence. This introductory period was filled with either silence in the case of the quiet condition or a period of irrelevant sound (28 tokens of irrelevant sound) that continued, without a break in tempo throughout the presentation of the to-be-remembered sequence. The lead-in period was designed, along with the regularity of timing, to promote the build-up of the irrelevant stream with the aim of engendering the segregation of the to-be-remembered from the to-be-ignored sequences (see Beauvois & Meddis, 1997; Nicholls & Jones, 2002a, for a related technique).

Design. The factors of modality (visual or auditory presentation) and suppression (suppression or no suppression) were blocked in a repeated-measures design. Participants completed all trials in one modality before proceeding to the other. Within modality block, participants completed all trials in one suppression condition before commencing trials in the other suppression condition. A Latin-square design enabled the order of modality and suppression block across participants to be counterbalanced. Within each suppression block, participants received 13 trials in each of the three irrelevant sound conditions (quiet, repeated, changing). The order of irrelevant sound conditions was quasi-random such that no irrelevant sound condition was encountered more than twice in succession. Although the order of irrelevant sound conditions remained the same for each participant across blocked conditions, the to-be-remembered sequences changed across suppression block.

Procedure. All participants were tested individually in a sound-proof laboratory. Participants wore headphones throughout the experiment and the level of sound was adjusted to approximately 60 dB(A).

Participants were instructed to ignore any spoken letters they heard but to recall the digits in their order of presentation as soon as list presentation was complete. Just before each block, participants were given instructions as to whether articulatory suppression was required in addition to two practice trials from each of the three irrelevant sound conditions. In all trials, 2 s before the onset of the to-be-remembered sequence a tone was sounded. This tone was used to warn participants that the to-be-recalled sequence was about to commence and, in the case of the articulatory conditions, prompted participants to start suppression. The tone was also sounded, and was clearly audible, when the design called for an irrelevant sound sequence to be presented. Under suppression conditions, participants were asked to whisper repeatedly the sequence of syllables, x, y, and z, at a rate of approximately three items per second until they were cued to recall. During the no-suppression blocks, participants were enjoined to resist any desire to speak the to-be-remembered items aloud. Before the experimental trials began, the experimenter coached the participants in the correct rate and loudness for the articulatory suppression.

Immediately after the list was presented, participants were required to attempt written recall in a strict serial manner. Participants were monitored closely throughout the experiment to ensure that they were suppressing in accordance to instructions. With their consent, participants’ compliance with all aspects of the procedure was monitored through a video and sound link.

Participants were invited to take a 5-min break halfway through the study as they switched from one modality to the other. The experiment lasted approximately 90 min.

Results

Responses were scored by a strict serial order criterion: An item had to be in its presentation position for it to be scored correct.

The general form of the results is shown in Figure 1. In control conditions, the effect of irrelevant speech took its usual form: Changing sequences were more disruptive than repeated sequences and repeated sequences were little different to quiet. However, there was no hint of any significant difference between speech conditions when articulatory suppression was undertaken.

Both in general and in the particular the results support the view that articulatory suppression abolishes the effect of irrelevant sound and that this is equally true of auditory and visual modalities. A repeated-measures ANOVA was undertaken without the serial position factor but containing the factors suppression (present vs. absent), modality (auditory vs. visual presentation), and irrelevant sound (quiet, repeated token, changing token). For

![Figure 1](image-url)
the purposes of the hypothesis under test, the most important interaction was that of suppression and irrelevant sound, $F(2, 78) = 18.00, MSE = 0.54, p < .001$. Its form may be appreciated readily from Figure 1: The usual effect of irrelevant sound was found in conditions of no suppression, but these effects disappeared in the presence of suppression.

The Suppression $\times$ Modality $\times$ Irrelevant Sound Condition interaction was not significant, $F(2, 78) = 0.68, MSE = 0.55, p > .05$. Planned comparisons revealed that the contrast between repeated and control conditions for auditory conditions was not significant under either suppression ($F < 1$) or no suppression ($F < 1$) conditions. The same contrasts were also not significant for the visual condition (both $Fs < 1$). However, under conditions of no suppression there was a significant difference between repeated and changing conditions in both visual, $F(1, 78) = 30.99, MSE = 17.11, p < .001$, and auditory, $F(1, 78) = 18.29, MSE = 10.10, p < .001$, conditions, but this effect was not significant in either modality under suppression (both $Fs < 1$).

Figure 2 shows the interaction of suppression and modality with respect to serial position (summed over irrelevant sound conditions). These data were subject to a three-factor, repeated-measures analysis of variance (ANOVA) with the factors modality (visual vs. auditory), suppression (present vs. absent), and serial position (with seven levels). The auditory–visual modality difference took the usual form, with the main difference being the superiority of recall from auditory lists in recency, shown as a main effect of modality, $F(1, 39) = 21.42, MSE = 10.55, p < .0001$, and as a significant Modality $\times$ Serial Position interaction, $F(6, 234) = 11.65, MSE = 1.42, p < .0001$. The effect of articulatory suppression was to depress performance generally, shown as a main effect, $F(1, 39) = 250.70, MSE = 19.88, p < .0001$. For the visually presented list, this was evident throughout the sequence, with perhaps a general tendency for the disruption to increase with serial position. For the auditory lists the exception to the trend found for visual lists occurred in recency; the effect of suppression became smaller in recency (this being most evident in the terminal position). This was reflected in a Suppression $\times$ Modality interaction, $F(1, 39) = 9.92, MSE = 5.85, p < .005$, and a Suppression $\times$ Modality $\times$ Serial Position interaction, $F(6, 234) = 11.65, MSE = 0.96, p < .0001$. The locus of this effect is readily evident in Figure 2. Here, the key difference between auditory and visual presentation was in recency; auditory recency seemed to endure despite the presence of articulatory suppression. This effect is notable both for its size and likely functional significance; as becomes evident, this effect is repeated in the other experiments in this article. The effects of irrelevant speech were the same in the auditory and visual modality: The nonsignificant interaction of modality and irrelevant sound, $F = 0.14, MSE = 0.08, p > .05$, echoes that found by Nicholls and Jones (2002b) using a similar technique.

Discussion

The irrelevant sound effect is almost invariably expressed in the form of the changing state effect, namely, that a changing sequence of irrelevant items is more disruptive than a repeated sequence (e.g., Campbell, Beaman, & Berry, 2002; Jones & Macken, 1993, 1995a, 1995b, 1995c; Jones, Madden, & Miles, 1992; LeCompte, 1994, 1995). However, most of these demonstrations of the irrelevant sound effect have been with visually

![Figure 2](image-url). Serial recall percentage correct in Experiment 1 shown with respect to serial position for the Modality $\times$ Suppression interaction. The figure illustrates the strong recency with auditory presentation in the presence of articulatory suppression.
presented to-be-remembered lists (e.g., Colle & Welsh, 1976; Jones & Macken, 1993; Jones et al., 1992; LeCompte, 1995; Salamé & Baddeley, 1982). The effect of irrelevant speech on auditory sequences has been shown several times (Hamilton & Hockey, 1974; Hanley & Broadbent, 1987; Martin-Loeches, Schweinberger, & Sommer, 1997; Surprenant, LeCompte, & Neath, 2000; Surprenant, Neath, & LeCompte, 1999). Some studies used auditory stimuli interleaved with to-be-ignored stimuli but in none of the cases within this class was there a clear changing state effect (Baddeley, Papagno, & Andrade, 1993, Experiment 5; LeCompte, 1996, Experiment 3). These results appear to support the proposition derived from the working memory model that the processing of auditory lists is functionally different from that of visual lists (e.g., Baddeley, 1986; Hanley & Broadbent, 1987; LeCompte, 1996), and this in some way modulates the interaction with the irrelevant sound effect. Experiment 1 of the current series suggests that this conclusion needs to be revised. Just as Nicholls and Jones (2002b) showed, when principles of auditory streaming are used to partition the irrelevant sound from the to-be-remembered material when auditory lists are used, identical effects of irrelevant sound are found in the auditory and visual modality. Experiment 1 further reinforces this result. This isomorphism of action in the two modalities is echoed in the core factor of interest in Experiment 1, the joint action of articulatory suppression and irrelevant sound. The data are compelling in suggesting that the action of articulatory suppression in diminishing the irrelevant sound effect to vanishing point is the same in the auditory modality as it is in the visual modality. Unequivocally, the results go against those of Hanley and Broadbent (1987, Experiment 3) and the claims of the working memory model.

One additional interesting feature emerged from Experiment 1, that of the relative immunity of auditory recency from disruption by irrelevant sound. There are a variety of ways of interpreting this outcome. One can assume that automatic processes of auditory streaming are at play in recency, such that most recent items benefit from the automatic processing of order (see Nicholls & Jones, 2002a). In turn, this suggests (as Surprenant et al., 2000, demonstrated) that the suffix effect—the disruption of recency by the mere presence of a redundant to-be-ignored suffix to the list—should also be relatively immune to disruption by articulatory suppression. The immunity of the suffix from articulatory suppression stems from the same functional source as the immunity of recency to the effects of articulatory suppression. We argue in due course that this immunity reflects the process of perceptual organization of the list and that it is independent of rehearsal processes.

These reflections on the suffix effect are also germane to explaining one recent study in which the interaction predicted by the working memory model was found (Klatte, Lee, & Hellebrück, 2002). Klatte et al. found an irrelevant speech effect with auditory presentation but only in the case where the irrelevant sound continued after the list through a 10-s retention interval. Under conditions of articulatory suppression, the irrelevant sound produced an effect highly localized in recency. In Klatte et al.’s case the irrelevant sound was presented from behind the participant and the to-be-remembered speech from in front. However, far from being distinct auditory spatial locations, front–back confusions of auditory localization is the most frequent (Blauert, 1997). We argue therefore, that what Klatte et al. observed was a suffix effect, which was due to the continuation of sound beyond the list, not an irrelevant speech effect.

Auditory recency seems to survive articulatory suppression, an outcome that suggests recency is not a product of rehearsal (because suppression reduces the efficiency of rehearsal) and may be the result of some preattentive processing of the auditory material. Auditory recency is also free from disruption by irrelevant sound, which, like the failure to find an interaction with sensory modality, indicates that it is not within the province of the hypothetical phonological store. The second critical interaction predicted by the working memory model, that among phonological similarity, articulatory suppression, and presentation modality, is considered in Experiment 2.

**Experiment 2**

The working memory model predicts that the phonological similarity effect will still be evident in conditions of articulatory suppression but only for auditory, not visual, lists. This prediction rests on the assumption of direct access of speech to the phonological store even when the phonological loop is occupied. Given that this is an interaction for which there is ample prior precedent (Baddeley et al., 1984; Murray, 1968; Peterson & Johnson, 1971), why engage in a further replication? Our interest in revisiting the effect was stimulated by the survival of recency in conditions of articulatory suppression found in Experiment 1. Prior studies have not provided the necessary serial position data (with the exception of Murray, 1968, but in that case a nonstandard recall task was used), and we believe that these data are essential to understanding the role of recency and to a thorough explication of the phonological store. Specifically, we expect that effects mainly in recency will underpin the critical interaction; that is, we speculate that the survival of the phonological similarity effect is a by-product of the action of recency under suppression. Auditory recency is known to be very vulnerable to disruption, and the pattern of this disruption suggests recency is a product of acoustic organization, not phonological processing.

**Method**

**Participants.** Twenty-four undergraduate students from the School of Psychology at Cardiff University were awarded course credits for their participation. All were native English speakers reporting normal hearing and normal or corrected-to-normal vision. None of the participants had taken part in the previous experiment.

**Apparatus and materials.** All trials consisted of seven to-be-remembered letters, which were selected without replacement either from a closed set of seven phonologically dissimilar consonants (f, k, l, m, q, r, y) or from a closed set of seven phonologically similar consonants (c, b, d, g, t, v, p). For both auditory and visual presentation, to-be-remembered items were prepared and presented in the same fashion as stimuli in Experiment 1.

**Design.** As with Experiment 1, the factors of modality (visual or auditory to-be-recalled items) and suppression (present or absent) were blocked in a repeated-measures design. Participants completed all trials from one modality before proceeding to complete trials in the other modality. A Latin-square design was used to counterbalance the order of modality blocks—and the suppression blocks contained within the modality blocks—across participants. Within each suppression block, participants received 15 trials consisting of phonologically dissimilar items and 15 trials made up of phonologically similar items. The order of phonolog-
exception that suppression consisted of whispering the digits 8, 9, and 10 in canonical order and repetitively at a rate of approximately 3 items per second. Participants completed four practice trials before each block.

Results

The results were as predicted by the working memory model. Figure 3 shows the broad pattern of results, with the phonological similarity effect being produced in all pairs of conditions (auditory control conditions, visual control conditions and, critically, auditory conditions with articulatory suppression) but not with visual presentation under articulatory suppression. The effect of suppression was to depress performance markedly; in the visual presentation the effect of phonological similarity was abolished, but it remained largely undiminished with auditory presentation during articulatory suppression.

In an ANOVA that excluded serial position with three factors, suppression (suppression vs. no suppression), modality (auditory vs. visual), and phonological similarity (similar vs. dissimilar), there was a main effect of suppression, $F(1, 23) = 155.04, MSE = 4.67, p < .05$; similarity, $F(1, 23) = 78.82, MSE = 1.19, p < .05$; but no main effect of modality, $F(1, 23) = 4.42, MSE > 0.05$, nor a significant Suppression $\times$ Modality interaction, $F(1, 23) = 1.67, MSE = 1.87, p > .05$. The three-factor interaction, Suppression $\times$ Modality $\times$ Similarity, was significant, however, $F(1, 23) = 6.78, MSE = 0.77, p < .05$. The form of this interaction is readily understood with reference to Figure 3; it is completely in line with the prediction of the working memory model.

Planned comparisons revealed that differences between similar and dissimilar lists were significant for auditory and visual presentation without suppression, $F(1, 23) = 123.25, MSE = 94.32, p < .001$, and, $F(1, 23) = 86.24, MSE = 66.00, p < .001$, respectively, and for auditory lists with suppression, $F(1, 23) = 54.08, MSE = 41.39, p < .001$. Notably, the difference between similar and dissimilar lists was not significant for the visual lists under articulatory suppression ($F < 1$).

Panel A of Figure 4 shows the no-suppression conditions and Panel B of Figure 4 shows the suppression conditions. Under no-suppression conditions with auditory presentation, it is clear that similar items were very poorly recalled relative to dissimilar items and that this effect was evident throughout the serial position curve with some attenuation of the effect in primacy. The same pattern of disruption that was due to similarity is evident in the visual modality.

In a three-way, repeated-measures ANOVA covering the data in Panel A, only the effect of similarity was evident as a main effect, $F(1, 23) = 72.98, MSE = 15.26, p < .05$, and as an interaction with serial position, $F(6, 138) = 9.30, MSE = 2.75, p < .05$, reflecting the fact that the phonological similarity effect was relatively larger in serial positions beyond Serial Position 3. There was no main effect of modality, $F(1, 23) = 0.21, MSE = 25.82, p > .05$; however, and as expected, the effect of modality interacted with serial position, $F(6, 138) = 5.06, MSE = 2.56, p < .05$, reflecting the differences localized in recency. There was a trend for the similarity effect to be greater in the auditory modality, but this was not significant, $F(1, 23) = 2.81, MSE = 3.15, p > .05$. The three-factor interaction of similarity, modality, and serial position was not significant, $F(6, 138) = 1.11, MSE = 1.65, p > .05$. In the absence of suppression, the effects of phonological similarity are roughly equivalent in the auditory and visual modalities.

Turning to Figure 4B, it is evident that articulatory suppression produced a very marked effect on performance, depressing it in general. However, the most distinctive feature of the results is that of massive recency—reminiscent of that found in Experiment 1—associated with dissimilar lists when presented in auditory form. In an ANOVA for the data portrayed in Figure 4B only, the main effect terms were significant for similarity, $F(1, 23) = 13.44, MSE = 11.78, p < .05$, and for serial position, $F(1, 138) = 29.51, MSE = 14.06, p < .05$. The main effect of modality was not significant, $F(1, 23) = 0.99, MSE = 18.01, p > .05$, but was significant as an interaction with serial position, $F(6, 138) = 7.41, MSE = 4.97, p < .05$. These and other interactions are subsumed under a significant three-factor interaction—modality, serial position, and similarity—the result of marked recency with auditory presentation of dissimilar lists, $F(6, 138) = 3.01, MSE = 2.57, p < .05$.

It is worth noting that although the difference in recency was very marked, there was a small but consistent difference between similar and dissimilar lists in the primacy portion of the serial position curves for auditory lists (about the same magnitude found in the first 2 items of the control lists), but this difference was not evident in the visual lists under articulatory suppression.

Discussion

In general, we again observed the survival of auditory recency under articulatory suppression, just as we did in Experiment 1. However, the phonological similarity effect in auditory lists seems to have survived to a slight degree in primacy also (indeed the data are generally noisy in the primacy region of Figure 4A). We do not place great weight on this for several reasons: First, the effect was relatively much smaller than the effect in recency in Experiment 2;
second, the importance of recency in surviving suppression had precedent in Experiment 1; and third, and perhaps most compelling, when the similarity effect is examined in very similar conditions within Experiment 3, there were no effects of similarity in primacy.

One way of interpreting the persistence of the phonological similarity effect in recency is to assume that auditory recency reflects the action of the phonological store (although no working memory theorist has made such a claim explicitly). From another standpoint, ascribing auditory recency to the phonological store is problematic because there is evidence that auditory recency is regarded as an acoustic–auditory phenomenon, not a phonological phenomenon. This evidence comes in the form of experiments that examine the relation between the properties of the suffix and of the auditory to-be-remembered list (see Crowder & Morton, 1969; also Beaman & Morton, 2000; Page & Norris, 1998). Claims that the suffix effect is acoustic stem from a wide range of experiments in which it has been found that the physical similarity of the suffix to the list is the primary determinant of disruption. For example, the effect is attenuated when the suffix and to-be-remembered sequence differ in spatial location (e.g., Frick, 1988b; Morton, Crowder, & Prussin, 1971), voice (e.g., Greenberg & Engle, 1983; Greene, 1991), and rhythm (e.g., Crowder, 1971; Frankish & Turner, 1984). Therefore, as the phonological identity of the suffix remains fixed, changes to its acoustic identity relative to that of the list bring about substantial changes in the suffix effect. Moreover, the suffix seems to obey the rules of auditory perceptual organization and these apply with equal force to speech and nonspeech stimuli (see Bregman, 1990; Warren, 1982, for general accounts and Nicholls & Jones, 2002a, for a review in relation to suffix phenomena).

Although the greater weight of evidence is by far and away in favor of an acoustic suffix effect, opinions are sharply divided on how this acoustic effect operates. Perhaps the most pervasive view is that the suffix overwrites (or masks) similar acoustic elements in the terminal item (e.g., Crowder & Morton, 1969; Nairne, 1990; Surprenant et al., 2000; but see also Neath, Surprenant, & Crowder, 1993). An alternative view is that similarity of the suffix to the terminal item promotes grouping of the suffix with the list and that this modifies the encoding of order (e.g., Frankish, 1989; Frick, 1988b; Kahneman, 1973; Nicholls & Jones, 2002a).

The earliest masking account located the suffix effect within a precategorical acoustic store (PAS; Crowder & Morton, 1969). The PAS was described as an auxiliary but passive store prior to postcategorical short-term memory. Verbal acoustic information enters the PAS automatically, either decaying, if followed by silence, or being masked by subsequently presented items. Thus, a suffix has its impact on recency by masking the representation of the terminal item in the PAS. The PAS accounts successfully for most of the phenomena associated with the suffix effect. Moreover, PAS is only one of many theories that use the phenomenon of masking to explain the effect of a suffix on recency (e.g., Burgess & Hitch, 1999; Crowder, 1978, 1982; Greenberg & Engle, 1983; Hitch, 1975; Nairne, 1988, 1990; Neath, 2000; Page & Norris, 1998; Surprenant et al., 2000).

The grouping account suggests the recency occurs because the terminal list item occupies a distinctive position at the boundary of the to-be-remembered sequence. Items occupying boundary positions are more likely than items within the group to be recalled because they have fewer opportunities for transposition (e.g., Henson, Norris, Page, & Baddeley, 1996; Ryan, 1969; see Frick, 1988a, 1989, for possible additional factors). A suffix acoustically similar to the to-be-remembered sequence is subject to grouping processes that integrate it perceptually with the to-be-remembered sequence. The suffix now occupies the distinctive boundary position usually occupied by the terminal list item, and the result is that the encoding of order in the terminal items is disrupted. However, when the suffix is acoustically different from the list, the terminal item of the to-be-remembered sequence denotes the boundary of the to-be-remembered sequence. A suffix that is acoustically different from the to-be-remembered sequence serves the same functional purpose as silence at the end of the list, that is, to indicate perceptually the termination or boundary of the to-be-remembered sequence and to improve recall of the terminal item (e.g., Frankish, 1988b, 1990; Hitch, 1975; Nairne, 1988, 1990; Neath, 2000; Page & Norris, 1998; Surprenant et al., 2000).
A recent extensive series of experiments, coupled with a reconsideration of earlier findings, found that the evidence was very much in favor of a grouping account. Nicholls and Jones (2002a) used a technique called auditory capture to study the action of the suffix. Auditory capture is a well-established psychoacoustic phenomenon (see Bregman & Rudnicky, 1975). In the case of Nicholls and Jones (2002a), irrelevant sequences of sound accompanied the to-be-remembered auditory sequence in such a way as to be grouped with the suffix. However, even in the presence of these irrelevant items, the relation of the suffix to the last item in the list was fixed, inasmuch as no irrelevant item was interpolated—the likelihood of masking of the last item in the list by the suffix was therefore unaltered by the events with which it could be grouped—only the likelihood of its incorporation into an irrelevant (as opposed to the to-be-remembered list) was changed in experimental conditions. When conditions of capture were optimized the suffix effect was completely abolished. It is difficult for a model that is based on the overwriting of the suffix in some buffer store to explain these effects; rather, some account that is based on grouping seems more appropriate.

We interpret the literature on the action of the suffix (and, hence, by implication, the action of recency) as dependent on acoustic not phonological factors. It seems reasonable therefore to hypothesize that the effect of recency found in Experiment 2 of the current series can be impaired markedly by the action of a suffix. Of course, it may be possible that the recency observed with phonologically dissimilar lists under conditions of articulatory suppression in Experiment 2 was in some way different from the recency susceptible to disruption by a suffix that was perceptually bound with the list. Our goal in Experiment 3 was to test these possibilities. If a suffix destroys that recency that is responsible for the survival of the phonological similarity effect, it would strongly suggest, at minimum, that what was observed in the key interaction of Experiment 2 was not the action of a phonological store but a phenomenon that was due either to acoustic storage or auditory grouping.

Experiment 3

In Experiment 3 we once again examined the joint action of articulatory suppression and phonological similarity, this time with auditory lists only, but in addition we established the effect of the presence of a suffix. The inclusion of visual lists within the design would not lend it any further analytic force; an auditory suffix does not have any appreciable effect on a visual verbal list (however, it should be noted in passing that a visual–spatial suffix is effective in reducing recency in visuo–spatial recall; see Maybery, Parmetier, & Jones, 2002). The degree to which recency was manifest in the phonologically dissimilar lists in the presence of articulatory suppression was observed and, crucially, the extent to which this recency was diminished by the presence of a suffix. We predicted that the phonological similarity effect would once again appear during articulatory suppression but that the effect of phonological similarity would be abolished when a suffix was present. More specifically, we predicted that the suffix would have the effect of reducing the considerable recency evident in the dissimilar lists under articulatory suppression without a suffix and that there would be little impact of a suffix on phonologically similar lists. There are grounds for optimism in relation to this prediction; it has already been established that the suffix effect may be observed even when recall of an auditory list is undertaken under conditions of articulatory suppression (Surprent et al., 1999), so it seems reasonable to suppose that the suffix will have its action on dissimilar lists in the presence of suppression.

Method

Participants. Twenty-six undergraduate students from the School of Psychology at Cardiff University were awarded course credits for their participation. All were native English speakers reporting normal hearing and normal or corrected-to-normal vision. None of the participants had taken part in the previous experiments.

Apparatus and materials. The same digitally sampled, spoken to-be-remembered stimuli used in Experiment 2 were used again here. The word zero, spoken in the same voice and at the same pitch as the list items, was used as the suffix. All items, including the suffix, were compressed digitally to 250 ms using SoundForge 5.0 software.

As in Experiment 2, list items were presented at a rate of one item per second. When present, the suffix was recorded in tempo with the to-be-remembered list, following the offset of the last to-be-remembered item by 750 ms. Within each list, items were selected without replacement from either the phonologically dissimilar or similar set. Letters were ordered quasi-randomly in each trial with the same constraints as those imposed in the earlier experiments of this series.

Design. Conditions were contrasted in a with-in-subject design. The factor of suppression (absent or present) was blocked such that participants completed all trials in one block (suppression or no suppression) before proceeding to the other. The order of suppression block was counterbalanced across participants in an AB–BA design.

Within each suppression block, participants received 52 trials, 15 trials from each of the four list conditions. List types comprised either phonologically dissimilar or phonologically similar items and were either followed by a suffix (the irrelevant word zero in the same voice as the list) or were not. Sometimes a tone suffix is used in addition to a no-suffix condition to control for list length; however, in general, there is only a small effect over and above that of no-suffix conditions. Hence, in the interest of keeping the number of trials in the experiment to a manageable proportion, this procedure was not used. The order of the four list types was quasi-random within each block of suppression conditions such that no list type was encountered more than twice in succession.

Procedure. The procedure followed that of Experiment 2 with the additional instruction to participants that they were to ignore the word zero when present in a trial. Participants completed four practice trials in each of the four list conditions before each suppression block. With participants’ permission, compliance with all aspects of the procedure was monitored through a video surveillance system.

Results and Discussion

The findings are in line with the idea that a suffix can abolish the phonological similarity effect during conditions of articulatory suppression through its action on recency. Figure 5 shows the overall means for each of the conditions; the phonological similarity effect was strong in the control conditions, undiminished by the presence of a suffix, diminished somewhat but nevertheless evident in the presence of articulatory suppression without a suffix, and all but absent in the presence of articulatory suppression with a suffix. An ANOVA covering all conditions (but without respect to serial position) revealed a main effect of articulatory suppression, F(1, 25) = 140.34, MSE = 24.73, p < .05, and of phono-
In a three-factor, repeated-measures ANOVA on the data contributing to Figure 6A, which included phonological similarity, suffix, and serial position for the no-suppression conditions, all three main effects were significant: $F(1, 25) = 51.79, MSE = 18.71, p < .05$; $F(1, 25) = 65.58, MSE = 4.87, p < .05$; and $F(1, 25) = 43.56, MSE = 7.94, p < .05$, respectively. As we expected because of considerable precedent, the presence of a suffix interacted with serial position, $F(6, 150) = 8.87, MSE = 2.62, p < .05$. Importantly, suffix and phonological similarity did not interact significantly, $F(1, 25) = 0.53, MSE = 4.99, p > .05$, nor did suffix, phonological similarity, and serial position, $F(6, 150) = 1.23, MSE = 2.21, p > .05$.

Turning to Panel B, the results without a suffix are closely similar to those found in the auditory list condition of Experiment 2. Again, marked recency was evident in the dissimilar list. Of particular interest for the thesis developed here is the disappearance of this recency with the suffix. The results are unambiguous in showing that the suffix acted to remove the recency that allowed phonological similarity effects to survive articulatory suppression.

In a three-way, repeated measures ANOVA for data from Figure 6B there was a significant main effect of the presence of a suffix, $F(1, 25) = 47.69, MSE = 2.26, p < .05$, and a significant two-way interaction involving suffix and serial position, $F(6, 150) = 11.88, MSE = 2.27, p < .05$, with these subsumed under a significant three-way interaction of serial position, similarity, and suffix, $F(6, 150) = 6.29, MSE = 1.66, p < .05$.

The results of Experiment 3 were in line with our predictions: The recency that permitted the expression of the phonological similarity effect under articulatory suppression was abolished by the presence of a suffix. Recency was more pronounced in auditory phonologically dissimilar lists for reasons that are analogous to the action of the suffix; the end of the list formed a distinctive boundary but only with items that were themselves distinctive—the suffix effect diminished this distinctiveness. The suffix had no appreciable effect on phonologically similar lists, for which there was no recency of any noticeable magnitude in the absence of a suffix. It seems possible to construe the classic interaction of
PETERSON and JOHNSON (1971), which was replicated in Experiment 2, as a highly localized effect, indeed one that is related more to the processes of perception than to phonological storage. Notably, the results of Experiment 3 also show that there was no residual effect of phonological similarity in the primacy region under conditions of suppression. These findings also provide reassurance about the inconsistent effect of phonological similarity found in primacy with auditory lists in Experiment 2: Our inclination to place less weight on them seems justified in the light of these results (and indeed the results of Experiment 1).

It is clear that, without modification, the working memory model is not able to account for the results of Experiments 2 and 3. Generally, other models do not address this particular interaction, with the notable exception of the feature model (Nairne, 1990; Nairne & Kelly, 1999; Neath, 1999). The feature model distinguishes between modality-independent (corresponding to an abstract, possibly phonological representation) and modality-dependent features (those related to the stimulus characteristics of the list, such as the contrast between auditory and visual presentation) and assumes that they are distinct insofar as they do not interfere with each other. Recency occurs when modality-dependent features are protected from overwriting by subsequent material, such as occurs at the end of a list. Auditory traces have more modality-dependent features than visual traces and are therefore less likely to suffer interference from other modality-dependent representations in the list, with the result being enhancement of recall at the end of the list. The suffix effect can therefore be explained by overwriting of modality-dependent features (but see Bloom & Watkins, 1999; Nicholls & Jones, 2002b). Articulatory suppression impairs recall by supplying more modality-independent features, which depresses performance uniformly. On the basis of the results of Experiment 2, therefore, simulations of the feature model very accurately predict the outcome in the visual modality but are less successful for the auditory modality, because it fails to show that the survival of the phonological similarity effect is restricted to recency (in fact it survives everywhere but recency; see Figures 9 and 10 of Nairne, 1990). By the same token, the model would not have predicted the outcome of Experiment 3. However, of all the available models, the feature model has the potential to simulate our data most closely: If the model can be recast so that modality dependence is restricted to recency, then a closer approximation to the current findings might be achieved. In fact, the feature model is alone among quantitative models in addressing interaction of phonological similarity and articulatory suppression (neither the primacy model, Page & Norris, 1998, nor the start–end model, Henson, 1998, address this interaction).

General Discussion

It may be useful at this stage to summarize the key results of the experimental series. First, with respect to irrelevant sound, suppression and modality: (a) The irrelevant sound effect at presentation did not survive suppression, implying that rehearsal is a precondition for its expression; (b) the degree and form of the irrelevant sound effect was not different in auditory and visual modalities; and (c), following from (b), auditory recency seemed immune to the effect of irrelevant sound. Second, with respect to phonological similarity, articulatory suppression, and modality, (a) the phonological similarity effect survived articulatory suppression but only for auditory presentation; (b) generally this effect was highly localized—performance for the last few items (and particularly the very last) in the list was unaffected by suppression; (c) this recency was much more marked for phonologically dissimilar lists than for phonologically similar lists; and (d), furthermore, this effect was eradicated by adding a redundant suffix to the auditory list, which implies that the effect is an acoustic one related to automatic encoding of order at the end of the auditory lists. In addition to not finding any evidence for the existence of a phonological store, it seems reasonable to conclude that two processes contribute to serial recall: first, a perceptual process, made evident in the conditions involving minimal rehearsal (conditions of articulatory suppression) by recency effects and, second, a process of rehearsal, chiefly responsible for the phonological similarity effect, but one without a phonological substrate.

In the remainder of this article, experimental evidence relating to the action of two phenomena that are the defining characteristics of the phonological store construct—the phonological similarity effect and the irrelevant sound effect—is augmented by a critical review of previous work. The aim of this section is to show that the irrelevant sound effect and the phonological similarity effect do not arise from the action of a phonological store.

The Irrelevant Sound Effect

The working memory model posits necessarily that the degree of resemblance between the phonological identity of the irrelevant sound and the to-be-remembered items is the key determinant of interference. However, most studies show the contrary. That similarity between the relevant and irrelevant sequence is not important was first shown formally by Jones and Macken (1995b) and has been replicated several times (Bridges & Jones, 1996; Buchner, Irmen, & Erdfelder, 1996; Larsen, Baddeley, & Andrade, 2000; LeCompte & Shaibe, 1997; but see Hughes & Jones, in press; Salamé & Baddeley, 1982; Tolan & Tehan, 2002). As mentioned earlier, the key determinant of disruption is the degree of change within the irrelevant sequence, an observation that has been dubbed the changing state effect (Jones et al., 1992).

The changing state effect can be exploited to illustrate the point that the phonological identity of the irrelevant sequence can remain fixed, but through changes in perceptual organization of the irrelevant sequence, the magnitude of the disruption can be modulated. For example, if the repeated sequence of consonants x, y, and z is presented monaurally, it will be significantly disruptive of serial recall, just as with any other sequence of changing sound from a single source. However, if the syllable sequence is kept in that fixed order, but stereophonic presentation of sound is used to assign the x to the right ear, z to the left ear, and y to both ears, the compelling phenomenal experience is that there are three streams of repeated stimuli. If this stereophonic form of presentation is then used to present the sounds as irrelevant material, the degree of disruption is appreciably diminished (Jones, Macken, & Murray, 1993; Jones, Saint-Aubin, & Tremblay, 1999). Note that the relation between the phonological content of the list and the irrelevant sound is fixed and only the perceptual organization has changed.

Again, if the identity of a repeated token (e.g., a consonant) is fixed so that the phonological identity of the irrelevant sequence does not change, but the difference in pitch between successive consonants varies instead, there is still a strong irrelevant sound
effect (Jones, Alford, Bridges, Tremblay, & Macken, 1999). In a complementary fashion, if the pitch difference between two sounds is fixed and shows a changing state effect, speeding up the rate of presentation reduces the degree of disruption as the alternative sounds are partitioned into two streams (Macken, Tremblay, Houghton, Nicholls, & Jones, 2003). Yet again, if the identity of the token remains fixed but the regularity of the timing in between the tokens is varied, an appreciable degree of disruption occurs (Jones & Macken, 1995c). In each of these four cases the phonological identity is fixed but changes to perceptual organization augments or diminishes the disruption, at will. Similarity in phonology of the irrelevant events to the to-be-remembered events is not the important determinant of disruption, an outcome that theories embodying a phonological store construct find difficult to explain.

By the rule embodied within the phonological store construct that the store is phonological (and therefore excludes nonspeech sounds) or by the rule that irrelevant stimuli need to be similar to the to-be-remembered items—both rules amount to the same thing—nonspeech stimuli should not produce an irrelevant sound effect. Thus, the fact that nonspeech stimuli consistently produce irrelevant sound effects goes against this tenet of the phonological store. The catalogue of nonspeech stimuli that produce the effect suggests wide generality to its action, including as it does, pitch glides (Klatte, Kilcher, & Hellbrück, 1995; Jones et al., 1993), tones (Divin et al., 2001; Jones et al., 1992; Martin-Löeches & Sommer, 1998; Neath, Surprenant, & LeCompte, 1998), sine-wave speech (Tremblay, Nicholls, Alford, & Jones, 2001), and band-pass noise (Tremblay, Macken, & Jones, 2000). Some have argued for qualitative differences between speech and nonspeech stimuli (e.g., Surprenant et al., 1999) but others, using similar techniques, have found no supporting evidence (Tremblay et al., 2000). The issue of why speech sounds produce more disruption than nonspeech sounds is not addressed here (for a discussion, see Tremblay et al., 2001).

Nonspeech stimuli show the same effects of perceptual organization on serial recall and again illustrate the point that similarity of content is not an important factor: Like speech, nonspeech stimuli obey the rules of perceptual organization in relation to the irrelevant sound effect. For example, if the auditory sequence comprises two alternating tones whose difference in pitch is varied on different memory trials, a nonmonotonic relation is found between the difference in pitch between the tones and the degree of disruption. Disruption increases initially, but as the difference in pitch becomes very large, the degree of disruption diminishes significantly (Jones, Alford, et al., 1999; Jones, Saint-Aubin, & Tremblay, 1999; Macken et al., 2003). This nonmonotonicity has its counterpart in the phenomenal experience of the tones in experiments that involve listening rather than ignoring sequences of tones; at first, there is a compelling experience of a single sound but one changing in composition, but as the pitch difference increases to an appreciable degree there is increasingly the tendency to perceive two streams, each unvarying in pitch.

The irrelevant sound paradigm allows us to witness indirectly this same transformation in perceptual organization from a single changing stream to two unchanging streams by observing with such stimuli, first an increase, then a decrease in serial recall error. Note that this close mapping of phenomenal experience when the sound is being attended and the pattern of errors when the sound is being ignored implies that not only is the processing of unattended sound obligatory, it is subject to perceptual organization in much the same way as when the sound is attended (for a dialectic on this issue, see Carlyon, Cusak, Foxton, & Robertson, 2001; Macken et al., 2003).

The key conclusion from this part of the review is that one of the key determinants of disruption is the process of streaming of the irrelevant sound and that the similarity of identity (e.g., phonological description) between to-be-remembered and to-be-ignored is of little, if any, consequence (but see Hughes & Jones, in press). However, interference cannot be understood from a consideration of streaming alone, because the character of the memory task is also important in mediating the interference.

Generally speaking, there is ample evidence that irrelevant sound is much more disruptive of tasks involving rehearsal (see Banbury & Berry, 1998; Beaman & Jones, 1997; Morris & Jones, 1990) than those in which rehearsal is a negligible component (see Baddeley & Salamé, 1986; Boyle & Coltheart, 1996; Burani, Vallar, & Bottini, 1991; Richardson, 1984; Salamé & Baddeley, 1990). For example, if a test of recall for order uses a subset of 6 days of the week—Tuesday, Monday, Friday, Wednesday, Saturday, and Thursday—the usual disruption occurs. However, if instead the participant is required to report the missing day (Sunday)—a task that can seemingly be achieved by passive listening alone—the irrelevant sound effect is severely attenuated (Beaman & Jones, 1997) and critically does not exhibit its empirical signature, the changing state effect (Jones & Macken, 1993, Experiment 4). As noted when discussing Experiments 2 and 3, serial rehearsal is also a prerequisite for the phonological store account of the irrelevant sound effect but only for the visual, not for the auditory, modality. However, the results of Experiment 1 do not support this view.

The phonological store construct fails to account for a range of effects with irrelevant sound but other models also have difficulties to some degree in accounting for the overall pattern of findings. The feature model (Nairne, 1990; see, in particular Neath, 2000) shares with the phonological store notion, through its feature adoption construct, the idea that disruption is based on similarity and is thus prey to all the foregoing criticisms (see Baddeley, 2000b; Jones & Tremblay, 2000, for a fuller critique of the feature model). Other constraints, unique to the feature model, face additional difficulties. Unlike the working memory model, the feature model has the further constraint that stipulates that events in the irrelevant sound sequence be synchronous with the items in the to-be-recalled sequence (either as they are being encoded or during rehearsal), but given that subvocal rehearsal occurs as the items are being presented it is impossible to know whether this condition of synchrony is ever met. Thus it is difficult to see how this element of the feature model can ever be tested.

The changing state effect is construed within the feature model by invoking an attentional factor; one of the model’s tenets is that repeated items are easier to ignore than a stream of changing items and are subject to a habituation-like process. However, empirically, there is no evidence of habituation of the irrelevant sound effect over trials within a session (Hellbrück, Kuwano, & Namba, 1996; Jones, Macken, & Mosdell, 1997; Tremblay & Jones, 1998) nor even a diminution between sessions (Ellermeier & Zimmer, 1997; Hellbrück et al., 1996). Moreover, the mere fact of a token-dose effect (Bridges & Jones, 1996), namely that disruption in-
increases as the number of irrelevant tokens (be they speech or nonspeech) per unit time increases (for a given token set size), also goes against the idea of habituation.

The primacy model also accounts for the irrelevant sound effect in terms of the assumption that the sound draws attentional resources away from the primacy gradient representing the to-be-remembered list to avoid direct interference. This gradient, which is streamed separately from that of the to-be-yielding cues to order, leads to the setting up of a second primacy higher activation. An irrelevant sequence that changes in state, yielding cues to order, leads to the setting up of a secondary gradient, which is streamed separately from that of the to-be-remembered list to avoid direct interference. This “will draw some resources away from the primacy gradient representing the to-be-remembered items” (Page & Norris, 2003, p. 1296). This neatly accounts for the key weakness in the working memory model of the failure of between-stream similarity to produce the effect and has the additional virtue of explaining phenomena in the domains of auditory attention (specifically poor order information between streams) and short-term memory performance.

Both the feature model and the primacy model fail to predict accurately details of the effect of change, however. As the number of different tokens in the irrelevant sequence increases, both models predict a linear increase in disruption. The primacy model predicts “the more changes in state, the more likely it is that the irrelevant stimulus comes to be represented at the order-storing level” (Page & Norris, 2003, p. 1297). However, the experimental evidence shows a nonlinear effect (Tremblay & Jones, 1998): The disruption at first rises (roughly as the number of tokens increases from one to two) and then diminishes markedly as the number of tokens rises above three. Of the alternative models, the primacy model and the feature model come closest to explaining the available data and both share the additional virtue of being quantitative models of memory. However, both share the weakness of incorporating the loosely specified attention construct.

The irrelevant sound effect and the phonological store construct. Several lines of evidence converge to support the view that the phonological similarity of the irrelevant sound to the to-be-remembered sequence is not a key determinant of disruption. This is contrary to the predictions of the phonological store construct. What seems to be important is the degree of change within the irrelevant sequence; this serves both as the basis for perceptual organization into streams and, in turn, the degree of disruption. The ease with which the order of the irrelevant stream can be perceived seems to predict the degree of disruption when that information is then used as an irrelevant stimulus and the phonological status of the irrelevant sound is immaterial (but see Hughes & Jones, 2003). The other key factor is the degree to which rehearsal is used in the memory task—the more rehearsal that takes place, the greater is the disruption (see also Experiment 1 of the current series). Both elements of the setting need to share the process of order keeping, or serialization. We now discuss the other primary phenomenon associated with the phonological store, the phonological similarity effect.

The phonological similarity effect. Rather than suppose, as the working memory framework does, that the phonological store is the locus of the phonological similarity effect typically found in serial recall, we suggest here that its locus is the rehearsal process. The evidence from the current experiments is compelling in suggesting that rehearsal is a precondition for the emergence of the phonological similarity effect (see also, MacAndrew, Klatzky, Fiez, McClelland, & Becker, 2002). Furthermore, we suggest that errors of phonological similarity arise from speech planning during the rehearsal process. We argue this from the observation that errors of recall are not random, but systematic, and that the pattern of errors reveals a marked similarity to that found in natural speech (see Bailey & Hahn, 2004; Brady, Mann, & Schmidt, 1987, for developmental implications).

Errors that occur in phonologically similar lists are of a type encountered naturally as spoonerisms, and in a serial recall task they appear as errors of order. The working memory model assumes that these errors arise in the phonological store on the basis of similarity. However, the errors of similarity are not symmetrical: For example, similarity at the end of a word or syllable sequence seems more important than similarity at the beginning (see, e.g., Crowder, 1978), a fact that is not explained in the working memory model.

The idea that errors in short-term serial recall are errors of production is a relatively old one. Among the first to advance this idea was Morton (1964), but it was given perhaps most convincing realization by Ellis (1980), who put forward the idea of a response buffer “whose normal function is to allow the efficient program-ming of speech production by holding preplanned stretches of impending speech” (Ellis, 1980, p. 624). The primary function of this buffer is in the production of speech. This proposition is expressed in terms of the error equivalence hypothesis, that “the same forms of phonemic error should be detectable in both speech and short-term memory, and they should be influenced by the same variables in the same way” (Ellis, 1980, p. 625). It is not necessary to invoke the action of a buffer, however, to explain the results of Ellis (1980); such effects could occur in one of a number of ways: for example, as part of a cyclic rehearsal process not linked to any store or buffer.

The method used by Ellis (1980) was to present five syllables auditorily for serial recall and to note different classes of error as different combinations of vowels and consonants were used in the to-be-remembered list. Four classes of error were found, each of which has a corresponding error in natural language: (a) Consonant transpositions occur more frequently than vowel transpositions, which in turn occur more frequently than whole syllable transpositions—this is a classic characteristic of everyday spoonerisms (MacKay, 1970); (b) there is a feature similarity effect: The likelihood of transposition between two consonants increases as the similarity of their articulatory features increases—this is also a common characteristic of natural speech errors (MacKay, 1970; Nootboom, 1969; Shattuck-Hufnagel, 1992); (c) a contextual similarity effect is observed: Consonant transposition is more likely between pairs of syllables sharing identical vowels than between syllables with different vowels, whereas, also in naturally occurring spoonerisms, there is a greater than chance tendency for the vowels of the two syllables to be identical (Nootboom, 1969; see also Shattuck-Hufnagel & Klatt, 1979; Wilshire, 1999); (d) a syllable position effect is found: When phonemes transpose, they relocate to their corresponding positions in the respective syllables, again just as happens with spoonerisms (Boomer & Laver, 1968; MacKay, 1970; Nootboom, 1969; see also Levelt, 1989).

Broadly similar results, from a wider range of syllable types, were produced by Treiman and Danis (1988), couched in terms of the onset–rime distinction. Treiman and Danis also found errors to
be systematic: The ways in which to-be-remembered syllables break apart and recombine are not random; transpositions followed a consistent pattern, the most frequent being the onset of one to-be-remembered stimulus and the rime of another. Not all phonemes were free to transpose; for example, it was seldom the case that the phonemes constituting the rime were separated. These errors therefore appear to arise not entirely at the phonological level but also at some level above it. These data also confirm the error equivalence hypothesis of Ellis (1980). The evidence drawn from many sources—linguistic data, spontaneous errors in natural speech production, and experimental studies of word games (e.g., Treiman, 1983)—all bear remarkable similarity in their form and relative incidence to those found in serial short-term memory (see also Brady, Shankweiler, & Mann, 1983). Of course, the analysis of errors needs not be restricted to speech gestures alone—manual gestures can also show phonological similarity effects (see Leybaert & Lechat, 2001).

Against the phonological store construct must stand the failure to specify in detail the nature of the errors. Moreover, that the pattern of errors (and indeed the pattern of timing of output; see below) can be reproduced by the mere act of reading a list, in which no significant burden is placed on short-term memory (Ellis, 1980; Page & Norris, 2003; Sternberg, Monsell, Knoll, & Wright, 1978; Sternberg, Wright, Knoll, & Monsell, 1980) is also particularly telling.

That interference between item representations at the phonological level is the sole basis for the phonological similarity effect is the most widely held view (see, e.g., Burgess & Hitch, 1996, 1999; Henson, 1998; Page & Norris, 1998). For example, the primacy model (Page & Norris, 1998) has a module specifically fashioned to account for phonological similarity effects. That model embodies a “confusion” or “output” stage at which phonological forms are explicitly represented where there the likelihood of confusion between similar sounding items is greater than that between dissimilar sounding items. This second stage is effectively transparent to nonconfusable items. The great virtue of this arrangement is that it allows the model to predict a number of rather complex interactions that would otherwise be difficult to understand. For example, if similar and dissimilar items are alternated in a list, their level of performance is the same as if they came from lists composed purely either of similar or dissimilar items. The primacy model can predict this pattern of results through the action of the second stage. Perhaps the most problematic feature of this approach is the lack of parsimony of the second stage, particularly the arbitrary and unaccountable way similar items are partitioned for further analysis by that stage. The discussion of Experiments 2 and 3 described the way the feature model accounts for phonological similarity through the construct of interference. Generally, the model gives a good fit to the data but does not accurately predict the nature of the interaction among similarity, suppression, and modality.

Results from the recognition span task are often taken as evidence that phonological similarity effects are not based on rehearsal but on phonological storage (see Gathercole, Pickering, Hall, & Peaker, 2001; also Baddeley, Chincotta, Stafford, & Turk, 2002). This is a task, usually auditory, in which the role of rehearsal is minimal: A verbal sequence is repeated but on some trials one pair of adjacent items is transposed. Here, phonological similarity effects are evident in conditions in which rehearsal is minimal, a result that seems to go against the view we have been developing that phonological similarity effects are rehearsal based. However, the lesson of the current series is that auditory recency can mediate such phonological similarity effects, so the effect could just as easily be ascribed to auditory perceptual organization as to phonological storage. Serial position analysis of such data could at least clarify the role of auditory recency; regretfully, no published study on recognition span includes such an analysis. Certainly, we would predict that the phonological similarity effect in recognition span is mediated by auditory recency and that, as we demonstrated in Experiment 3, this phonological similarity effect could be removed by adding a suffix.

The phonological similarity effect. The force of the foregoing experiments and review is to suggest that the term phonological similarity effect may be a misnomer; the errors normally associated with it are perhaps better thought of as speech errors. On the basis of parsimony, it seems reasonable also to suggest that a short-term store dedicated to a phonological level of representation is superfluous. The fact that rehearsal seems to be a precondition for the phonological similarity effect (notwithstanding the vestigial role of acoustic factors shown in Experiments 2 and 3 in accounting for the survival of recency) is particularly portentous: Phonological similarity is chiefly the product of speech errors, not errors of perception, nor errors of phonological storage (but see Neath, 2000, for a different view). For a general account, therefore, it may be desirable to replace the phenomena of short-term phonological memory with processes related to speech production coupled with an account of auditory order perception. It may be that in using such mechanisms it will be possible to abandon the notion of modular memory stores altogether. Certainly, it seems reasonable in the light of the data of the current series to suggest that specific phonological representation is unwarranted.

Our consideration of the phonological similarity effect suggests that it is fruitful to construe the task of serial recall as being dominated by properties of behavior that emerge from sequence production, rather than from the recollection of a series of isolated elements. Conventionally, phonological segments are conceived as consequences of perceptual processing and as inputs to the articulatory mechanism. A computationally burdensome process of translation—unelaborated by all models of serial short-term recall incorporating phonological representation—is required to turn these discrete, static, and context-free phonological representations to the dynamic context-sensitive plan of articulation that is the substrate for vocal performance in serial recall tasks (see Fowler, Rubin, Remez, & Turvey, 1980, for a discussion). Even when the recall mode is manual, it is derived from what is primarily a motor output plan for the vocal tract; in these cases the output sequence is a product of vocal tract constraints and those of the particular manual mode, such as cursive writing (see Murray & Jones, 2002, for a discussion of the effect of coarticulation on memory span). Given this context sensitivity, a model of serial recall without a phonological representation would have a distinct advantage of parsimony, therefore, in bypassing the conundrum of this translation process. As we have argued, errors of recall can be explained adequately by recourse to speech production processes, without phonology. Furthermore, if an abrupt distinction is made between capabilities of output production and premotor output planning, we may make intelligible the capacity of congenitally speechless individuals to produce nevertheless seemingly normal short-term
memory performance (Bishop & Robson, 1989) from a plan (and not one necessarily involving speech coding), not from the production process.

Conclusions

It seems possible to begin to argue that many of the key phenomena of short-term memory do not require constructs related to functionally distinct storage entities. Rather, we may view short-term retention as parasitic on general perceptual and motor planning processes, particularly (though not solely) as they relate to language.

One key goal for future research is to establish the extent to which the distinction made here between the perceptual organization for words and the plans for their output—represented by rehearsals—can be translated into a more general framework for reproducing sequences. Instead of regarding speech planning as a special case, it is possible to speculate that it is part of the general case of reproducing configurural order. Configural sequences may take forms other than verbal–articulatory; they may be spatial or gestural (see Leybaert & Lechat, 2001), or more speculatively, they could be made up of a sequence of different attentional foci. Even with verbal material, mechanisms other than subvocal plans may be used; so, for example, if needed, fingers can be used to remember digit sequences (Reisberg, Rappaport, & O’Shaughnessy, 1984). In summary, the output plans for speech are simply a restricted case of the general constraints on the reproduction of a stimulus sequence by the production of gestures (see Corballis, 2002, for a discussion of the relation of gesture to the evolutionary development of language).

Evidence is beginning to emerge of mechanisms common to many forms of serial short-term memory. For example, there is ample evidence that memory for serial order exhibits similar characteristics whether the to-be-remembered material is verbal or spatial (e.g., Jones, Farrand, Stuart, & Morris, 1995), which in turn suggests a rehearsal mechanism common to verbal and nonverbal stimuli. The marked recency usually observed with auditory–verbal items can also be obtained with visuo–spatial items but only when the memory task mainly requires order (e.g., order reconstruction) rather than item information (Farrand & Jones, 1996; Farrand, Parmentier, & Jones, 2001; see also Avons, 1998; Smyth & Scholey, 1996). As in verbal memory, visuo–spatial serial memory has proved to be sensitive to the irrelevant sound effect (e.g., Jones et al., 1995; Tremblay, Macken, & Jones, 2001), the suffix effect (Parmentier, Tremblay, & Jones, in press), and the sandwich effect (Tremblay, Parmentier, Nicholls, & Jones, in press). Other effects such as grouping (Parmentier, Maybery, & Jones, in press) and the Hebb repetition effect (Tremblay & Couture, 2003) also appear to be key determinants of spatial serial recall. These studies constitute evidence that is accumulating against any simple notion of modular partitioning of spatial and verbal short-term memory that is embodied in the working memory model. Shared characteristics might arise from the common tendency to use configurual rehearsal, but differences in action may arise from a variety of factors such as the support of perceptual organization or from the skill base (e.g., from stored regularities) and constraints on configurational complexity of the rehearsal process and so forth.

In a broader context our standpoint harmonizes well with those who believe that short-term memory capacity is not some processing primitive but instead emerges from experience, particularly of language (see MacDonald & Christiansen, 2002; also Brady et al., 1987; Ericsson & Kintsch, 1995; but see Caplan & Waters, 2002). The working memory approach promotes partitioning of a short-term memory (and kindred cognitive architectures with dedicated storage entities), from cognition generally, and we regard this as undesirable inasmuch as it constitutes reification (see Macken & Jones, 2003). Certainly, in the foregoing discussion of what has been regarded as phonological memory within serial short-term recall, there is a strong case for a process-oriented approach rather than the deployment of constructs related either to phonology or to storage entities. We believe that evidence presented here also suggests a shift in emphasis more generally within research on short-term memory away from the delineation of stores toward describing the interplay of organization factors within perceptual systems and output planning processes.

References


suffix effect; Are the effects structural or strategic? Memory & Cognition, 11, 551–556.


Received November 4, 2003
Revision received November 17, 2003
Accepted November 20, 2003

---

**Low Publication Prices for APA Members and Affiliates**

**Keeping you up-to-date.** All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*. High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

**Essential resources.** APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association*, and on dozens of new topical books each year.

**Other benefits of membership.** Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

**More information.** Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.