The Functional Determinants of Short-Term Memory:

Evidence from Perceptual-Motor Interference in Verbal Serial Recall

Robert W. Hughes and John E. Marsh

RUNNING HEAD: The Functional Determinants of Short-Term Memory

KEYWORDS: Serial Recall; Perceptual Organization; Motor Planning; Auditory Distraction; Order Incongruence

Correspondence:

Robert W. Hughes, Department of Psychology,

Royal Holloway, University of London, Egham, Surrey, United Kingdom,

TW20 0EX; Phone: (+44) 01784 443280

email: Rob.Hughes@rhul.ac.uk

Abstract

A functional, perceptual-motor, account of serial short-term memory is examined by investigating the way in which an irrelevant spoken sequence interferes with verbal serial recall. Even with visual list-presentation, verbal serial recall is particularly susceptible to disruption by irrelevant spoken stimuli that have the same identity as—but which are order-incongruent with—the to-be-remembered items. We test the view that such interference is due to the obligatory perceptual organization of the spoken stimuli yielding a sequence that competes with a subvocal motor-plan assembled to support the reproduction of the to-be-remembered list. In support of this view, the interference can be eliminated without changing either the identities or objective serial order of the spoken stimuli but merely by promoting a subjective perceptual organization that strips them of their order-incongruent relation to the to-be-remembered list (Experiment 1). The interference is also eliminated if subvocal motor sequence-planning is impeded via articulatory suppression (Experiment 2). The results are in line with the view that performance-limits in verbal serial short-term memory are due to having to exploit perceptual and motor processes for purposes for which they did not evolve, not the inherently limited capacity of structures or mechanisms dedicated to storage.

KEYWORDS: Short-Term Memory; Motor Planning; Perceptual Organization; Serial Recall; Irrelevant Sound; Auditory Distraction

The capacity to retain and reproduce verbal input in serial order over the short term has long been recognized as critical for many higher-order cognitive functions including key aspects of language comprehension and learning, problem-solving, and reasoning (e.g., Baddeley, 1986). Given the importance of verbal serial short-term memory, it seems surprising that it is so acutely vulnerable to disruption by the mere presence of task-irrelevant stimuli, particularly from irrelevant spoken material as well as other kinds of sound (e.g., Beaman & Jones, 1997; Colle & Welsh, 1976; Ellermeier, Kattner, Ueda, Doumoto, & Nakajima, 2015; Elliott, 2002; Hanley & Hayes, 2012; Salamé & Baddeley, 1982). Such vulnerability has often been cited as support for theoretical frameworks in which short-term memory performance is supported by a distinct structure or memory space that is intrinsically fragile, highly prone to decay or/and interference (e.g., Baddeley, 2007; Larsen & Baddeley, 2003; Neath, 2000; Salamé & Baddeley, 1982, 1989, 1990). The present research is embedded within an alternative theoretical framework that avoids the paradox of positing dedicated yet fragile short-term memory structures. On the perceptual-motor account, performance in verbal serial short-term memory tasks is parasitic on general-purpose perceptual organization and motor-planning processes that are co-opted on the fly in an attempt to meet task demands (e.g., Hughes, Marsh, & Jones, 2009, 2011; Hughes, Chamberland, Tremblay, & Jones, 2016; Jones, Hughes, & Macken, 2006, 2007; Macken, Taylor, & Jones, 2014, 2015; see also Buchsbaum & D'Esposito, 2008; MacDonald, 2016; Melby-Lervåg & Hulme, 2010; Postle, 2006; Wilson & Fox, 2007). From this standpoint, the vulnerability of short-term performance is not so surprising because the processes supporting that performance were not specifically designed for that purpose. In particular, in this view, it is the inherent permeability of a motor-plan assembled in the face of a highly novel sequence that leaves verbal serial short-term memory performance vulnerable to task-irrelevant sequences (e.g., Hughes & Jones, 2005; Macken et al., 2015). In the present study, we test this perceptual-motor account by studying verbal serial recall performance in the presence of a spoken distractor sequence that would be expected to offer particularly strong competition for the motor-planning process. In support of the account, we show that a to-be-ignored spoken sequence containing the same items as in a (visually-presented) to-be-remembered list—but in an incongruent order—is particularly disruptive of serial recall but that this disruption is eliminated if either an obligatory perceptual organization of the spoken sequence is promoted that strips it of its competitiveness (Experiment 1) or if articulatory motor-planning is precluded (Experiment 2).

Verbal Serial Recall in the Face of Task-Irrelevant Sound

The classic test of serial short-term memory is verbal serial recall in which, typically, around five to eight verbal items (e.g., digits or letters) are presented at the rate of one or two items per second. The participant is required to recall the items in serial order immediately following the last item or following a short retention interval (e.g., Baddeley, 1966, 1986; Conrad, 1964). It is well established that serial recall, even when the to-be-remembered items are presented visually, is impaired appreciably by irrelevant spoken stimuli even though participants are explicitly told that the sound is irrelevant to their task, that they will not be tested on its content, and that they are therefore to ignore it the best they can (e.g., for reviews, see Beaman, 2005; Hughes & Jones, 2001). It is important to recognize that the distractors need not be speech (or verbal) to produce disruption however: a sequence of pure tones (Divin, Coyle, & James, 2001; Elliott, 2002; Jones & Macken, 1993; Sörqvist, 2010), pitch-glides (Jones, Macken, & Murray, 1993; Klatte, Kilcher, & Hellbruck, 1995), noise-bursts (Tremblay, Macken, & Jones, 2001), and nonvocal music (Klatte, Kilcher, & Hellbrück, 1995; Perham & Vizard, 2012; Schlittmeier, Hellbrück, & Klatte, 2008; Salamé & Baddeley, 1989) also impair verbal serial recall. Rather, the necessary and sufficient condition for reliable disruption is that the sound comprises a sequence of segmentable, acoustically changing, elements: Thus, changing-state sound (e.g., “b, f, q, r, t…”; or a sequence of tones changing in frequency from one to the next) produces appreciable disruption whereas a steady-state sound (e.g., “b, b, b, b, b…”; or the same tone repeated) produces little if any disruption compared to quiet (i.e., the changing-state effect; e.g., Campbell, Beaman, & Berry, 2002; Hughes, Tremblay, & Jones, 2005; Jones, Madden, & Miles, 1992).

Of particular interest in the present article is the finding that whereas similarity between the irrelevant and relevant material is certainly not necessary for sound to be disruptive of serial recall, the disruption is greater when the spoken distractors are postcategorically identical to the to-be-remembered items. That is, when the spoken distractors are, for example, “8, 5, 3, 6, 1, 4, 7, 2”, the serial recall of the (visually-presented) list *57812643* is impaired to a greater degree than when the spoken distractors are relatively dissimilar to the to-be-remembered items (e.g., letter-names; Hughes & Jones, 2005; Salamé & Baddeley, 1982; see also Jones & Macken, 1995b). On the face of it, this finding seems both intuitively obvious and in line with the classical concept of similarity-based interference embodied in several accounts of the irrelevant sound effect (Gathercole & Baddeley, 1993; Salamé & Baddeley, 1982; Neath, 2000). However, the starting point for the present study is that this apparent item-similarity based effect is in fact an *order incongruence effect*: the interference is uniquely located at the sequence-level, not at the level of the individual items. Specifically, when the particular order in which the identical set of distractors is presented is incongruent with the to-be-remembered sequence (e.g., “8, 5, 3, 6, 1, 4, 7, 2” when the to-be-remembered list is *57812643),* then indeed serial recall is poorer than when the distractors are dissimilar. However, if the order of those same distractors is congruent (but out of temporal phase) with the to-be-remembered list (e.g., “4, 3, 5, 7, 8, 1, 2, 6”), those distractors—despite still being postcategorically identical to the to-be-remembered items—no longer impair serial recall compared to dissimilar distractors. Thus, item- (or sub-item-) level interference of the sort often postulated in short-term/working memory models (e.g., Neath, 2000; Nairne, 1990; Oberauer, Farrell, Jarrold, & Lewandowsky, 2016) cannot account for this phenomenon (Hughes & Jones, 2005). In the present study, we use the order incongruence effect to reveal the contribution to verbal serial short-term memory performance of general-purpose mechanisms of sequential perceptual organization and motor-sequence planning.

A Perceptual-Motor View

Several theories have emerged in recent years that conceive of verbal short-term memory performance as parasitic on processes and systems that are not specifically memorial. Some of these appeal to the systems involved in language processing and suppose that performance in verbal short-term memory tasks reflects nothing more than language comprehension and production skills (Acheson and MacDonald, 2009a,b; Buchsbaum & D'Esposito, 2008; MacDonald & Christiansen, 2002; Melby-Lervåg & Hulme, 2010). Another parasitic-type account that we will use here as our main theoretical framework appeals to even more general-purpose processes: On the *perceptual-motor account* (e.g., Hughes et al., 2009, 2011, 2016; Jones, Hughes, & Macken, 2006), verbal short-term retention is the byproduct of motor-sequence planning processes involved in producing any coherent sequential action (including, but not confined to, vocal action; e.g., Rosenbaum, 2009) and, particularly when auditory stimuli are involved, preattentive and involuntary processes of sequential perceptual organization (which again apply to verbal stimuli but not uniquely so; Bregman, 1990; Sussman, Bregman, & Lee, 2014). An increasing number of key serial recall phenomena that putatively reflect the operation of a dedicated storage space are being successfully recast purely in terms of perceptual organization and motor-planning; these now include the 'phonological' similarity effect and its interaction with articulatory suppression and modality (Jones et al., 2006, 2004, 2007; Maidment & Macken, 2012; Sjöblom & Hughes, 2016; see also General Discussion), modality and suffix effects (Macken, Taylor, Kozlov, Hughes, & Jones, 2016; Maidment, Macken, & Jones, 2013), the influence of long-term linguistic knowledge (Macken, Taylor, & Jones, 2014; Woodward, Macken, & Jones, 2008), and perceptual variability effects (Hughes et al., 2009, 2011, 2016).

Within a perceptual-motor view, verbal serial recall may be construed as a setting in which there is an extreme under-specification of action-parameters problem (cf. Hommel, 2010; Neumann, 1987, 1996). By design, the action required in a serial recall task is highly under-specified: A list of items is presented in which the constituent items are sequentially unrelated, that is, they will not (or are very unlikely to) match any extant long-term unitized representation (e.g., “5, 1, 6, 3, 7…” might be presented but not “1, 2, 3, 4, 5…” or “Mary had a little lamb…”; but see Jones & Macken, 2015). In the face of such sequential novelty, the skill of speaking (or, more accurately, ‘inner-speaking’) is co-opted, not to refresh decaying items in a dedicated memory space as in the classical view (see, e.g., Baddeley, 2007), but in order to try to bind items that bear little or no pre-existing sequential relation to one another. That is, the sequentiality and continuity of speech provides a common carrier upon which to place each to-be-remembered item such that they are no longer unrelated but instead become embodied within a single, temporally-extended, motor-object. The prosodic and co-articulatory characteristics of natural speech (e.g., Sternberg, Wright, Knoll, & Monsell, 1980) further imbues the motor-plan with cues that support and constrain the serial order of items (Hughes et al., 2009; Macken et al., 2014; Maybery, Parmentier, & Jones, 2002; Neisser, 1967; Woodward et al., 2008). However, any skill (including speaking) is, by definition, an abstract entity: whereas it specifies the general set of action-parameters required to produce a certain type of behavior (e.g., the set of parameters that govern the way the various components of the vocal tract must move to produce coherent speech), it remains to be populated with specific content (i.e., the words, phrases, sentences, and so on, that are to be produced; e.g., Hommel, 2010; Neumann, 1987, 1996). We suggest that it is this inherent openness—or in-need-of-populating characteristic—of a motor sequence-plan that leaves serial recall susceptible to interference by any input that could plausibly be a candidate for populating the plan but which may not specify task-appropriate action-parameters. In more general terms, in this view, interference in verbal serial recall does not reveal the existence of mnemonic capacity-limits but instead reflects the use of a perfectly functional process (motor-planning) being applied in the face of highly impoverished input (i.e., a highly, if not entirely, novel sequence).

It has been argued that one major source of irrelevant input that can threaten the integrity of the motor-planning process is that derived from the obligatory (i.e., non-volitional) process of perceptually organizing sound into streams (e.g., Jones & Macken, 1993; Hughes & Jones, 2005). Auditory streaming refers to the Gestalt processes whereby the initially undifferentiated mixture of inputs received by the ears is partitioned into coherent perceptual groups or objects (e.g., Koffka, 1935). The most important aspect of auditory streaming for present purposes is the computation of whether or not successive sounds have been produced by the same environmental event (e.g., Bregman, 1990; Moore & Gockel, 2002; Warren, 1999). Sounds that follow one another tend to be assigned to the same stream to the extent that they are acoustically similar (e.g., in terms of frequency, timbre, or inter-aural level or time difference) or/and show "good continuation", just as is the case for static visual stimuli on the spatial dimension (Bregman, 1990). Important for present purposes is that when sounds differ from one to the next but are nevertheless still similar enough to be assigned to the same stream—such as a series of different words but all spoken in the same voice—the serial order of those sounds is readily perceived. In contrast, it is notoriously difficult to discern the order of successive sounds that differ to such an extent that they are likely to be partitioned into distinct streams (e.g., different words spoken in different voices or emanating from different spatial locations; e.g., Bregman & Campbell, 1971; Lackner & Goldstein, 1974). Thus, a sequence of changing sounds that nevertheless share a common ground form a strong, ordered, sequence. Indeed, when auditory stimuli are presented as the to-be-remembered material in a serial recall task, performance is much better when they form a single, coherent, stream than if they are partitioned into different streams (Hughes et al., 2009, 2011, 2016). However, when that same coherent changing-state sequence is presented as a *task-*

irrelevant auditory sequence, it competes as a candidate for inclusion in the motor-plan and impairs serial recall (i.e., the changing-state effect; e.g., Jones & Macken 1993).

The hypothesis tested here is that the order incongruence effect (Hughes & Jones, 2005) reveals the action of a passive auditory sequencing process competing with an active subvocal motor-planning process but which, compared to the general effect of changing-state sound, occurs at the much finer level of the specific transitional information between particular verbal events. When presented with an auditory sequence of digits (spoken in the same voice), the serial transitions between the items (e.g, “three-five…”) are processed non-volitionally as a by-product of sequential auditory streaming. When these transitions are incongruent with those in the to-be-remembered list (e.g., three-seven), they compete with the deliberate process of specifying the articulatory transitions to be embodied in a subvocal sequence motor-plan. When the order of the irrelevant items is congruent with that of the presented items, there is no such sequence-level conflict. The current experiments test two straightforward predictions of the perceptual-motor interference account using manipulations designed to selectively alter the perceptual organization of the distractors and the motor-planning of the to-be-remembered list. In Experiment 1, we address the perceptual-input aspect of the account and test the prediction that the order incongruence effect should be attenuated if an objectively order-incongruent distractor-sequence is presented in such a way as to promote a subjective perceptual organization that renders that sequence no longer order-incongruent with the to-be-remembered list. In Experiment 2, we go on to test the prediction that impeding the capacity for subvocal motor-planning via the method of articulatory suppression—hence reducing the possibility of perceptual-motor interference—should also attenuate or eliminate the order incongruence effect.

Experiment 1

In this experiment, we test the idea that it should be possible to modulate the order incongruence merely by influencing the way the irrelevant distractor-sequence is likely to be perceptually organized. As noted, there is good evidence that the way in which sounds are perceptually organized has dramatic consequences for order processing. For example, Bregman and Campbell (1971) found that if a set of low-frequency tones (1, 2, 3) is alternated with a set of high-frequency tones (A, B, C) in a looping sequence (i.e., “1A2B3C1A…”) such that two interleaved streams are likely to be generated based on frequency-range, participants tend to inadvertently report the order of the tones by stream (e.g., 123ABC) instead of their temporal order. Similarly, we have shown that spoken to-be-remembered lists in which successive items alternate in terms of voice (Hughes et al., 2009, 2011) or ear-of-presentation (Hughes et al., 2016; see also Treisman, 1971) are particularly difficult to recall in serial order (compared to single voice or ear lists). Whereas the foregoing studies relate to order processing for attended sound-sequences, other studies have shown the role of streaming of to-be-ignored sound in the context of the greater disruptive effect on serial recall of irrelevant changing-state compared to steady-state sound. For example, if two tones presented in an alternating fashion are similar enough in frequency to cohere into one single changing-state stream, the usual appreciable disruption of serial recall is produced. However, when the two tones are separated further in frequency to the extent that a two stream percept is now more likely—with each stream comprising one steady-state tone—the disruption is markedly reduced (Jones, Alford, Bridges, Tremblay, & Macken, 1999; for further evidence of the preattentive nature of auditory streaming, see, e.g., Jones & Macken, 1995a; Sussman, Horváth, Winkler, & Orr, 2007; Sussman et al., 2014; Winkler, Denham, & Nelken, 2012).

Here we sought to demonstrate that auditory perceptual organization is a critical determinant of the impact of order incongruence on verbal serial recall. Our rationale begins with the assumption that with an auditory sequence comprising a series of different digits, the broader physical similarity between the digits—such as their shared voice and the fact that each is presented from the same spatial location—promotes their integration into a single coherent stream (cf. Hughes et al., 2009, 2016). We posit that it is this obligatory process of integrating the successive changing digits that yields information pertaining to the transitions between them and hence, in the case of an order-incongruent sequence, perceptual-motor interference. This leads to the prediction that if the successive items were to be presented such as to demote their integration into a single stream, such transitional information would be impoverished. In turn, it should no longer matter under such circumstances whether the transitions are order-congruent or order-incongruent with the to-be-remembered list. That is, the order incongruence effect should be reduced or eliminated when the spoken items no longer form a single stream.

As illustrated in Figure 1, we tested the perceptual streaming component of our account by including ‘with-alternation’ as well as ‘no-alternation’ versions of both an order-incongruent sequence and order-congruent sequence. In the no-alternation condition, all the items in a given sequence were presented in either a female voice or a male voice and presented to either the left ear or to the right ear. In this condition, therefore, the order incongruence effect should be replicated because the successive stimuli should cohere into a single stream: Serial recall should be poorer in the presence of the order-incongruent sequence than in the presence of the order-congruent sequence. In the with-alternation condition, the very same sequences were presented but now successive items alternated between the male and female voice and between the left and right ears. The alternation should promote the perceptual partitioning of successive distractors to different streams or at least reduce the likelihood of successive items cohering as strongly into a single-stream percept. If so, an objectively order-incongruent sequence would now, in perceptual terms, cease to be one: successive distractors would be highly acoustically dissimilar and hence the sequential relation between them would be rendered perceptually ambiguous (Bregman, 1990). Thus, the difference in recall between order-incongruent and order-congruent conditions should be attenuated or eliminated despite the fact the items themselves and their (objective) temporal order is identical to that in the no-alternation conditions.

Another feature of the design allowed us to test a strong as well as weak version of the streaming hypothesis. Regardless of alternation condition, the distractor sequences were generated such that in the order-incongruent condition, the nonsuccessive items (i.e., items in positions 1, 3, 5, etc., and likewise 2, 4, 6, etc.) were order-congruent with the to-be-remembered list and, conversely, in the order-congruent condition the nonsuccessive items were order-incongruent with the to-be-remembered list. Thus, on a strong version of the streaming hypothesis, in the with-alternation condition, the effect of an order-incongruent compared to an order-congruent sequence should be reversed (and not merely attenuated) because the alternation may lead *nonsuccessive* items to cohere into two interleaved streams on the grounds that they would now share voice and spatial location. That is, in the with-alternation condition, the order-congruent sequence should lead to poorer performance than the order-incongruent sequence. The weaker version of the hypothesis allows for the possibility that given the relatively short series of just four nonsuccessive sounds in each voice/ear, and the fact stream-integration takes some time to 'build up' (Anstis & Saida, 1985; Bregman, 1990), nonsuccessive items may be unlikely to cohere strongly together. According to this weaker hypothesis, the main action of the alternation manipulation will be to reduce the perceptual integration of successive items (rather than promote the integration of nonsuccessive items) and hence, as noted, it predicts a reduction or elimination rather than a reversal of the order incongruence effect.

Method

Participants. Forty-six undergraduate students at Royal Holloway, University of London, took part in exchange for course credits. All reported normal hearing and normal or corrected-to-normal vision.

Apparatus and Materials. Each trial of the focal serial recall task involved the visual presentation of the eight digits *1-8* presented without replacement, one at a time at the center of a computer screen in a 72-point Times font. The order of the digits was determined pseudorandomly for each list with the constraint that there were no ascending or descending runs of more than two digits. Each digit lasted 350 ms, and the interstimulus interval (ISI; offset to onset) was 400 ms. For the auditory stimuli, a set of spoken digits (“one”– “eight”) were recorded in both a female voice and a male voice (within each voice, the digits were spoken at an approximately even pitch) with a Sennheiser ME 65 microphone to 16-bit resolution at 22 kHz sampling rate using *Sony Sound Forge Pro 10* software (Sony Creative Software). Using the same software, each item was then digitally edited to last 250 ms. The male-spoken digits were then pitch-shifted down by two semi-tones and the female-spoken digits were pitch-shifted up by two semi-tones (without altering their duration) to further accentuate the acoustic difference between the two voices. Care was taken to ensure that the editing and pitch-shifting did not lead to any loss of intelligibility. The ISI (offset to onset) in the auditory sequences was 500 ms. The onset of each of the eight auditory items preceded each of the eight visual digits by 75 ms. All stimuli were presented using a PC running *Eprime 2.0 Professional Software* (Psychology Software Tools).

Design. A repeated-measures design was used with three factors: congruence (order-congruent and order-incongruent), alternation (irrelevant items alternating between voices/ears or not) and serial position. Regardless of alternation condition, an order-congruent sequence involved having the same sequence of digits as to-be-remembered and irrelevant material but the irrelevant sequence lagged behind the to-be-remembered sequence by 4 items. Thus, a concurrently presented to-be-remembered and irrelevant digit never matched but all but one of the transitions between temporally successive items in the irrelevant sequence also occurred in the to-be-remembered list. Note that the one transition within a digits-congruent order sequence but absent from the to-be-remembered list (“two”–“five” in the example in Figure 1) still does not conflict with any transition within the to-be-remembered list, because these items appear at each end of the to-be-remembered list. In the order-incongruent condition, again regardless of alternation condition, all the pairwise transitions differed from those in the concurrent to-be-remembered list.

It was also ensured in this condition that a concurrently presented to-be-remembered and irrelevant digit was never the same digit.

Within the no-alternation condition, all auditory items were presented to the same ear (either right or left) and in the same voice (either the female-spoken stimuli or the male-spoken stimuli). In the alternation condition, the auditory sequences were presented in an alternating female-male voice and alternating left-right ear fashion (see Figure 1). For any given alternating sequence, ear and voice of presentation were always perfectly correlated (e.g., male-item to left ear followed by female-item to right ear followed by male-item to left ear, and so on). Note also that for the alternation condition we retained the labels ‘order-congruent’ and ‘order-incongruent’ to refer to the formal status of the irrelevant sequences only; in practice, the voice/ear alternation was predicted to strip the two conditions of congruence/incongruence or to reverse their congruent/incongruent relation to the to-be-remembered list.

There were 16 trials in each of the four [2(Congruence) × 2(Alternation)] conditions presented in a pseudo-random fashion in a single block of trials with the constraint that a trial from each condition was presented once every four trials. In each of the no-alternation conditions, four of the irrelevant sequences were female-voice/left-ear, four were female-voice/right-ear, four were male-voice/left-ear, and four were male-voice/right-ear. In the with-alternation condition, four of the irrelevant sequences started with a female-voice/left-ear item and then alternated thereafter with a male-voice/right-ear item (as shown in the example in Figure 1), four started with a female-voice/right ear item alternating thereafter with a male-voice/left ear item, four started with a male-voice/left ear item alternating thereafter with a female-voice/right-ear item, and four started with a male-voice/right-ear item alternating thereafter with a female-voice/left-ear item.

Procedure. Participants were tested individually in a quiet room. Each participant was provided with instructions on the screen explaining what the serial recall task involved and were told that any speech heard over the headphones was irrelevant to their task and hence was to be ignored. Participants were also informed that the trials would be presented at a preset pace: 50 ms following the offset of the last visual to-be-remembered item, the screen flashed from white to black for 150 ms, which signaled the start of a 16.5 s written response period. Participants were required to write out the list of digits in the same order as they saw them. A 500 ms tone was presented over the headphones 13 s into the 16.5 s of writing time to signal that the first item of the next to-be-remembered list was imminent. Four practice trials, one from each condition, were given before the experiment proper.

Results

The data from both experiments reported in the present article were scored according to the strict serial recall criterion as standard: An item was only recorded as correct if its output position corresponded to its absolute temporal position in the presented list (*correct-in-absolute-position scoring*). Figure 2 shows the proportion of correctly recalled items at each of the eight serial positions (as well as the mean and standard error collapsed across serial position) from the order-incongruent and order-congruent conditions with and without alternation.

It is clear that alternation attenuated appreciably the order incongruence effect. Confirming this impression, a 2 (Congruence) × 2 (Alternation) × 8 (Serial position) repeated-measures ANOVA revealed a main effect of Congruence, $F(1, 45) = 6.12$, $MSE = .057$. $p < .02$, $\eta_P^2 = 0.12$, and, whilst there was no main effect of Alternation, $F(1, 45) = 2.04$, $MSE = .033$, $p > .05$, the critical interaction between Congruence and Alternation was significant, $F(1, 45) = 4.54$, $MSE = .020$, $p < .05$. Simple effects analyses showed that in the no-alternation condition the difference between the incongruent and congruent conditions was significant, $F(1, 45) = 10.11$, $MSE = .005$, $p < .005$, $\eta_P^2 = 0.18$ (a 'large' effect according to Cohen, 1988), whereas in the with-alternation condition, it was not, $F(1, 45) = 1.14$, $MSE = .005$, $p > .05$, $\eta_P^2 = 0.025$. As expected, the main effect of serial position was also significant, $F(7, 315) = 34.20$, $MSE = .076$, $p < .01$, as was its interaction with congruence, $F(7, 315) = 2.64$, $MSE = .010$, $p < .05$. We do not attempt to attach any functional significance to this latter interaction however; it may simply reflect decreased sensitivity to the effects of congruence at the first one or two serial positions where performance is near ceiling. Finally, whilst there is some evidence in Figure 2 for a tendency for the attenuation of the order incongruence effect in the alternation condition to be less marked at some serial positions than others (e.g., positions 3 and 4), the three-way interaction was not significant, $F(7, 315) = 1.92$, $MSE = .008$, $p > .05$.

Discussion

Experiment 1 showed that the order incongruence effect is eliminated when the distractors are presented such as to demote their perceptual organization into a single coherent stream. This is in line with our supposition that in a standard spoken sequence (i.e., same voice, same location), it is the integration of the items into a single stream that gives rise to information pertaining to the order of successive items. When such coherence is broken by presenting successive items in different voices and to different ears[1], such transitional information is impoverished or lost. Thus, in effect, the two types of sequence lose their differential status as order-congruent *vs.* order-incongruent.

The fact that the order incongruence effect was eliminated rather than reversed in the with-alternation condition means that the strong version of the streaming hypothesis was not upheld. That is, the shared voice and spatial location of the nonsuccessive distractors in the alternation conditions does not appear to have promoted the sequential integration of those nonsuccessive distractors. Had this been the case, we would have expected the order-congruent and order-incongruent conditions to have switched roles in the alternation condition, leading to a detriment in the order-congruent (with-alternation) compared to the order-incongruent (with-alternation) condition. There are a number of possible non-mutually exclusive reasons why such integration of

[1] The present design did not allow us to determine the extent to which each cue to stream segregation used here—ear/spatial location and voice—was effective. However, this is not relevant to the purpose of the current experiment; two covarying cues were used rather than one simply to increase the chances of inducing such segregation.

nonsuccessive distractors does not seem to have occurred. For example, as noted earlier, there were only four sounds within each voice/ear in the alternation conditions; this may be too few to allow for the build-up of two coherent streams (Bregman, 1990). Alternatively, or in addition to there being too few nonsuccessive distractors, the relatively long interstimulus interval between them (1250 ms), compared to successive distractors (500 ms), may also have demoted the formation of two coherent streams (cf. van Noorden, 1975). Thus, the action of the alternation manipulation here seems, in line with the weaker version of the streaming hypothesis, to have been to reduce the coherence of successive distractors without necessarily producing coherence between nonsuccessive distractors. Regardless, by confirming even this weaker hypothesis, we have shown that streaming processes are critical to the order incongruence effect.

Experiment 2

The perceptual-motor interference account predicts that the impact of order incongruence should also be attenuated or eliminated if subvocal motor-planning of the to-be-remembered items is precluded: In the absence of motor-planning, there can be no perceptual-motor interference. We therefore implemented a commonly used technique for blocking (or at least impeding) the use of subvocal motor-planning, namely, articulatory suppression, in which participants are instructed to repeat a task-irrelevant utterance (subvocally, vocally, or as used implemented here, in a whispered manner) during the memory task (e.g., Baddeley, Lewis, & Vallar, 1984; Murray, 1968; Jones et al., 2004). In this experiment, therefore, we contrasted performance under order-congruent and order-incongruent conditions while participants engaged, or did not engage, in articulatory suppression. We also included an auditory condition comprising a sequence of irrelevant letters as well as a quiet condition in this experiment (which were also undertaken with or without articulatory suppression). The letters condition was included to ensure that the order incongruence effect observed in the no-alternation condition in Experiment 1 was a true replication of the original effect reported in Hughes and Jones (2005), that is, a disruptive effect of an order-incongruent compared to order-congruent sequence, not a facilitative effect of an order-congruent compared to an order-incongruent sequence (cf. Bell, Mund, & Buchner, 2011). For a true replication, performance in the order-incongruent condition should, in the no-suppression condition, be poorer than in the letters condition while performance in the order-congruent sequence should not differ from the letters condition. We also included a quiet condition because it is already established that articulatory suppression removes the general effect of changing-state irrelevant sound compared to quiet (Hanley, 1997; Jones et al., 2004). Thus, if we were to observe that the order incongruence effect survives articulatory suppression, contrary to our predictions, we would have an additional independent indication of whether that was because the articulatory suppression manipulation had been ineffective; that is, the general effect of irrelevant sound should also still be evident in such a case.

Method

Participants. Thirty undergraduate students at the University of Central Lancashire took part in exchange for course credit. All reported normal hearing and normal or corrected-to-normal vision.

Apparatus and Materials. These aspects of the method were the same as Experiment 1 except for the following details. We used the original, non-pitch-shifted versions of the female-spoken stimuli throughout for all auditory sequences and all auditory stimuli were presented to both ears. In addition, a set of spoken letter-names ("b," "h," "j," "k," "l," "m," "q," and "s") was recorded in the same female voice as the digits, again spoken at an approximately even pitch and edited to last 250 ms each while ensuring no loss of intelligibility.

Design. The experiment had a repeated-measures design with three factors: auditory condition (four levels; quiet, letters, digits-congruent order, and digits-incongruent order), articulatory suppression (no-suppression vs. under-suppression), and serial position (eight levels). The letters condition involved presentation of the eight letters in a random order for each trial. On this occasion, in the digits order-congruent condition, the irrelevant sequence lagged behind the to-be-remembered list by five items rather than four (cf. Experiment 1) or two (cf. Hughes & Jones, 2005). We did this in light of a study by Bell et al. (2011) in which an order-incongruence effect was found with a lag of 2 but not 5. However, several aspects of their method differed from that used to first demonstrate the effect (Hughes & Jones, 2005) and thus it seemed prudent to check whether the effect does indeed generalize to a lag of 5 when the original methodology (other than lag) is adhered to.

Within each articulatory suppression condition there were 12 trials in each auditory condition except for the letters condition in which there were 24 (so that the number of trials with irrelevant letters equalled that with irrelevant digits), making 60 trials within each suppression condition and a grand total of 120 trials. The articulatory suppression factor was blocked and the order of blocks counterbalanced across participants. Within each suppression block, the four auditory conditions were presented in a pseudorandom order with the constraint that each condition was presented once per 5 trials except for the letters condition which was presented twice every 5 trials.

Procedure. The procedure was the same as for Experiment 1 except in relation to the under-suppression block. In that block, participants were required to engage in concurrent whispered articulation of the word “saxophone” repeated approximately twice per second throughout the list-presentation period. The Experimenter demonstrated the required rate of suppression and compliance with the requirement was then monitored via a microphone link. Before each suppression block, two practice trials, one from the letters condition and one from the digits-incongruent order condition, were undertaken before the experimental trials.

Results and Discussion

Figure 3 shows proportion correct recall in the eight conditions (4 auditory conditions × 2 suppression conditions) at each serial position as well as collapsed across serial positions. With regard to the no-suppression data, the first thing to note is that there is a general effect of changing-state sound: Performance is poorer in all conditions in which there was irrelevant sound compared to quiet. The order incongruence was also replicated: Recall was poorer in the digits-incongruent order condition compared to both the digits-congruent order condition and the letters condition and there was no difference between these latter two conditions. Also apparent is that articulatory suppression eliminated all differences between auditory conditions; thus, it abolished not only the general effect of irrelevant sound but also the order incongruence effect.

This impression of the data was confirmed statistically with a 4 (Auditory condition) × 2 (Articulatory suppression) × 8 (Serial position) repeated-measures ANOVA: There was a main effect of Auditory condition, $F(3, 87) = 5.83$, $MSE = .034$, $p = .001$, $\eta_P^2 = 0.17$, a main effect of Articulatory suppression, $F(1, 29) = 44.47$, $MSE = .47$, $p < .001$, $\eta_P^2 = 0.77$, and a main effect of Serial position, $F(7, 203) = 35.25$, $MSE = .106$, $p < .001$, $\eta_P^2 = 0.55$. Critically, there was also a significant interaction between Auditory condition and Articulatory suppression, $F(3, 87) = 5.08$, $MSE = .031$, $p = .003$, $\eta_P^2 = 0.15$. Simple effects analyses confirmed that, in the no-suppression condition, recall was significantly poorer in each of the sound conditions compared to quiet (all $p = .037$ or smaller). Of particular importance, the difference between the digits order-incongruent and digits order-congruent conditions was significant, $F(1, 29) = 6.83$, $MSE = .003$, $p < .02$, $\eta_P^2 = 0.19$ (again, a “large” effect; cf. Cohen, 1988), as was the difference between the digits order-incongruent condition and the letters condition, $F(1, 29) = 8.57$, $MSE = .002$, $p < .01$, $\eta_P^2 = 0.23$. In contrast, there was no difference between the digits order-congruent and the letters condition, $F < 1$, $p = .98$. Under articulatory suppression, there were no significant differences between any of the four auditory conditions, with the mean proportion of items correctly recalled collapsed across serial position for each condition being: Quiet – .34, Letters – .33, Order-congruent – .34, and Order-incongruent – .33, $p > .22$, for all pairwise contrasts).

In sum, Experiment 2 established that the order incongruence effect is eliminated under articulatory suppression. This finding supports the notion that when the deliberate process of sequencing the to-be-remembered items via a subvocal motor-plan is impeded or blocked, the incongruent transitions processed obligatorily from the digits-incongruent sound-sequence can no longer produce perceptual-motor interference. We can now also be fairly confident that the order incongruence effect is not restricted to a particular lag, having now been shown with a lag of 2 (Hughes & Jones, 2005), a lag of 4 (current Experiment 1) and a lag of 5 (current Experiment 2; cf. Bell et al., 2011; we return to the implications of this observation in the General Discussion).

Analysis of Intrusions from the Incongruent Sequence

Here we report an additional analysis of the data from both Experiments 1 and 2 to examine the extent to which the order incongruence effect is underpinned at least in part by the intrusion of the concurrently presented irrelevant order-incongruent sequence (or sub-elements of that sequence) into the response output. An alternative possibility is that it is the mere competition from the incongruent sequence, or the attempt to inhibit it, that underpins the effect, not its actual intrusion (e.g., Hughes & Jones, 2003; Tipper, 2001). On this latter hypothesis, such intrusions would be expected to be relatively rare or absent.

Method

For this analysis, then, we counted, for all trials in each order incongruent condition across the two experiments, the number of instances across a response-sequence that any successive pair of items matched a successive pair of items within the irrelevant sequence (e.g., outputting *23681754* when the to-be-remembered list was *26831754* and the irrelevant sequence was "4, 8, 2, 7, 3, 6, 1, 5" would be recorded as one pairwise intrusion). Note that a pairwise intrusion-error was recorded as such regardless of the position of the pair within the outputted sequence and regardless of its position within the irrelevant sequence. With a list-length of 8 items, the maximum number of such pairwise intrusions per trial was seven (i.e., transitions 1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8).

Results and Discussion

In relation to Experiment 1, the mean number of item-pairs within a response-sequence that matched those in a concurrent order-incongruent (no-alternation) sequence was 4.67 (SD = 4.59) out of a possible 112 (i.e., 7 possible item-pairs × 12 order-incongruent trials), that is, 4.17% of all possible successive pairs of responses. While this number is low in absolute terms it was nevertheless significantly higher than that in the order-incongruent (with-alternation) condition (M = 3.2, SD = 2.71; a rate of 2.85%), $t(1, 45) = 2.75$, $p < .01$. This result further bolsters our supposition that the alternation altered the perceptual organization of the objective order-incongruent sequence such that, subjectively, the sequence was stripped of that order incongruence.

One issue with the analysis of the data from Experiment 1 in terms of the overall intrusion rate, however, is that some of the response-pairs that matched those in the order-incongruent sequence may simply have occurred by chance and not constitute actual intrusions from that sequence. For the corresponding intrusion analysis on the data from Experiment 2, therefore, we matched a given response-sequence against not only the concurrent order-incongruent sequence that participants actually heard but also against a second order-incongruent sequence that they did not (it was also ensured that, for each trial, the control sequence had no pairwise overlap with the actual, heard, order-incongruent sequence). The number of actual intrusions could then be estimated by subtracting the number of pairs matching the actually-heard sequence from the number matching the second, not-heard, order-incongruent sequence (which would be by-chance matches).[2] For the data from Experiment 2, then, the mean number of matching pairs in the order-

[2] This approach would not have worked in the context of Experiment 1 because the order-incongruent sequences in that experiment (but not in Experiment 2) were still systematically related to the to-be-remembered list (because non-successive items were, as part of the design of that experiment, order-congruent with it). Thus, a second set of not-

incongruent (no-suppression) condition was 5.87 (*SD* = 3.7) out of a possible 84 (i.e., 7 possible intrusions × 12 order-incongruent trials), a rate of 7%. When compared against a not-heard order-incongruent sequence, the number of matching pairs was 3.5 (SD = 1.94) or 4.2%. Thus, the rate of actual intrusions (number of matches in the order-incongruent condition minus the number of by-chance matches) was 2.37 (or 2.82%). The corresponding set of values for the order-incongruent (with-suppression) condition was 6.13 (SD = 3.79) or 7.3% of pairs matching the heard order-incongruent sequence and a by-chance number of 4.9 (SD = 2.88) or 5.83%, giving a mean actual intrusion number of 1.23 (or 1.47%). A 2 (Status: Heard *vs*. Not-heard order-incongruent sequence) × 2 (Articulatory suppression) repeated-measures ANOVA showed a main effect of the status of the order-incongruent sequence (Heard *vs*. Not-heard), $F(1, 29) = 12.16$, $MSE = 7.99$, $p < .005$, $\eta_P^2 = .3$, indicating that while the rate of intrusions was very low it was significantly greater than would be expected by chance. There was no main effect of Articulatory suppression, $F(1, 29) = 1.97$, $MSE = 10.59$, $p > .05$, $\eta_P^2 = .06$, nor an interaction between this factor and Status, indicating that suppression did not significantly modulate the rate of intrusions, $F(1, 29) = 1.73$, $MSE = 5.56$, $p > .05$, $\eta_P^2 = .06$.

In sum, the pairwise intrusions analyses indicated that there is a tendency for order-incongruent transitions from the irrelevant sequence to intrude into participants' responses. However, the rate of such intrusions is very low generally and thus it is far from the case that the order incongruence effect is underpinned by such intrusions. Indeed, this assumption is supported by the finding that the intrusion rate was not modulated by articulatory suppression: While suppression eliminated the order incongruence effect, the intrusion rate was not affected. This dissociation suggests that the intrusions may be due to occasional slippages of attention (e.g., Lachter, Forster, & Ruthruff, 2004) that are independent of the perceptual-motor interference

---

heard, order-incongruent, sequences would also differ systematically from the heard order-incongruent sequences and not in fact provide an appropriate control.

process we argue underpins the order incongruence effect. Any such slippages would not result in intrusions from the order-incongruence sequence, however, when the perceptual organization of that sequence renders it no longer order-incongruent (cf. Experiment 1). Thus, the order incongruence effect appears to be driven by the competition per se, and possibly the cost of trying to inhibit that competition, not the actual loss of that competition (i.e., the sound *assuming* the control of the motor system; see General Discussion).

General Discussion

To summarize the key findings of the present experiments, we showed that the disruptive effect on verbal serial recall of spoken distractors that are in an order that is incongruent with a to-be-remembered sequence is eliminated if the distractor sequence is presented in such a way as to demote the likelihood of successive items being organized into a single perceptual stream (Experiment 1). The order incongruence effect, as well as the more general effect of changing-state sound, was also eliminated if subvocal motor-planning is impeded via articulatory suppression (Experiment 2). There was also a tendency for a particular type of error wherein pairwise transitions present within the to-be-ignored order-incongruent sequence intruded into the recalled sequence but such intrusions are independent of the order incongruence effect per se. Our findings converge to support the view that the order incongruence effect is the result of interference between perceptual and motor processes that are both axiomatically supra-item, sequence-level, processes. We argue that auditory perceptual organization processes obligatorily yield information pertaining to the serial transitions between irrelevant spoken distractors which in turn conflicts with the serial transitions that must be embodied in a motor-plan of the to-be-reproduced list. Thus, if the involvement of the motor system is reduced or if the organization of the auditory sequence is altered, the disruption is modulated despite not altering the content of either the relevant or irrelevant material. More generally, the results are in line with an approach that emphasizes the action of general-purpose perceptual and motor processes in short-term memory performance rather than mechanisms and representations that are specifically mnemonic (e.g., Hughes et al., 2009, 2011; Maidment, Macken, & Jones, 2013).

Implications for the Specificity of Obligatory Perceptual-Motor Mapping

One implication of the perceptual-motor interference effect studied here is that it suggests a great deal of specificity in the degree to which obligatory auditory perceptual analysis flows into the motor system (cf. Buchsbaum & D'Esposito, 2008). It has long been argued on the basis of studies of the irrelevant sound effect that the content of the sound is not important; only the presence of acoustic change between successive segmentable entities matters (e.g., Hughes & Jones, 2001; Jones & Macken, 1993; Jones & Tremblay, 2000). The changing-state effect has thus been used to argue that a perceptual sequencing process applied involuntarily to sound will interfere with a motor sequencing process applied deliberately to task-relevant material regardless of the nature of either the changing-state sound (e.g., speech, nonspeech) or the nature of the focal-sequencing task (e.g., verbal, spatial; see Jones et al., 1995). However, a novel implication of the present experiments is that such perceptual-motor interference is not confined to sequencing at this gross, non-content-specific, level: An overlap in the specific content of the irrelevant and to-be-remembered material does indeed exacerbate the disruption of verbal serial recall. Critically, however, overlap at the item-level is necessary but not sufficient for this additional disruption: The overlap in item-content only assumes importance to the extent that the items are organized by perceptual streaming processes into a particular order and that order conflicts with that in which the to-be-remembered items need to be assembled into a motor-sequence plan.

More specifically, the finding that pairwise transitions between irrelevant words interfere with the motor-planning of a different pairwise transition (two-seven) suggests that, at the very least, the offset of one spoken word and the onset of the next ("two-four…") is processed and, uninvitedly, this transition makes itself available to the motor-plan assembly process. A consideration of the present data coupled to previous studies suggests further that at least one item in each successive pair must be lexically intact (i.e., “two-four…” or “two-four…”). Specifically, most studies that have included a condition in which the irrelevant items were identical (but order-incongruent) to those in the to-be-remembered list showed greater disruption in such a condition compared to an unrelated condition (present Experiment 1; Bell et al., 2011; Hughes & Jones, 2005; Jones & Macken, 1995b; Salamé & Baddeley, 1982; but see Bridges & Jones, 1996, Experiment 4; note that other than Hughes and Jones, 2005, none of these studies identified order incongruence as the key factor underpinning such disruption). However, evidence for disruption when there is between-sequence overlap in onset-offset information but not lexical identity (e.g., 1-9 as to-be-remembered items and “tun”, “gnu”, “tee”, “sore”, “thrive”, “fix”, “heaven”, “fate”, and “sign” as distractors) is much weaker: Whereas Salamé and Baddeley (1982, Experiment 5) found such an effect, many subsequent studies have since failed to replicate their finding or observed only a weak and typically non-significant effect (Bridges & Jones, 1996; Jones & Macken, 1995b; LeCompte & Shaibe, 1997). Several other studies have also shown that overlap at the sub-lexical, phonemic, level does not contribute to sound’s disruptive impact on verbal serial recall (Larsen et al., 2000; Marsh, Vachon, & Jones, 2008; Saito & Baddeley, 2004). Thus, whilst experiments aimed directly at this issue would be desirable, the available evidence suggests that at least one of the irrelevant words making up a pairwise transition must be intact. This suggests, more broadly, that preattentive processing at the level of seriating lexical entities assumes particular functional significance in the mapping of auditory input onto speech-production processes. Recent theorising based on neuroimaging research suggests the existence of a system dedicated to auditory-motor integration (Buchsbaum & D’Esposito, 2008), in line with several psycholinguistic models of speech perception and production (e.g., Levelt, Roelofs, & Meyer, 1999; Nadeau, 2001). From this standpoint, the interference from an order-incongruent auditory sequence brings into relief the action of a usually functionally-adaptive system whereby auditory sequences are preattentively integrated with systems capable of generating a motoric version of that sequence.

The outcome of the additional pairwise-intrusion analyses provides some further valuable information about the nature of how the access of a perceptually-derived sequence into the motor-planning system impairs performance. The finding that intrusions from the order-incongruent sequence were rare and, moreover, did not underpin the order incongruence effect suggests that it is the mere competition from the order-incongruent sequence, or the attempt to inhibit that competition, that impedes performance. Evidence that the competing sequence is indeed subject to inhibition comes from a negative priming study by Hughes and Jones (2003): If an auditory sequence that was order-incongruent with the concurrent to-be-remembered list was re-presented as a to-be-remembered list on the next trial (now visually-presented), recall of that sequence was poorer than for a control list. This impairment of recall was attributed to the fact that inhibition applied to the same sequence presented recently as a competitor-sequence carried over to the next trial (cf. Tipper, 2001). Thus, the order incongruence effect may primarily be due to an overhead-cost of *preventing* the intrusion of competing pairwise transitions into the overt execution of the motor-plan via an inhibitory process; the small number of pairwise-intrusions that are observed, then, would reflect the rare occasions on which the inhibitory process has been unsuccessful (see also Marsh, Hughes, Beaman, & Jones, 2012).

Implications for Alternative Theoretical Accounts

Phonological Store-Based Models. That verbal serial recall is disrupted by irrelevant speech generally has, historically, been cited as one of the empirical cornerstones of the phonological loop component of Baddeley and colleagues' Working Memory model (e.g., Baddeley & Hitch, 1974; Baddeley, 1986, 2007, 2012; Repovs & Baddeley, 2006; Page & Norris, 2003). The phonological loop comprises a passive, decay-prone, phonological store and an articulatory rehearsal process that can revivify decaying phonological representations in the store as well as giving visually presented items access to the store. An initial account of irrelevant speech disruption based on the phonological loop construct posited that "it is the *degree of phonological similarity* between the irrelevant material and the memory items that underlies the irrelevant speech effect" (Gathercole & Baddeley, 1993, p. 13; Salamé & Baddeley, 1982). However, the fact that nonspeech disrupts verbal serial recall (Jones & Macken, 1993) and the observation that, when speech is used, overlap at the phoneme-level does not in fact dictate the magnitude of the disruption (e.g., Jones & Macken, 1995b; LeCompte & Shaibe, 1997) has led to the abandonment of this early account (see, e.g., Baddeley, 2007; Repovs & Baddeley, 2006) and to the development of alternative phonological store-based accounts. The most prominent of these, based on the primacy model, has now adopted the changing-state hypothesis (Jones & Macken, 1993) and posits, like the perceptual-motor account, that disruption from changing-state irrelevant speech (and nonspeech) represents a conflict of two sequencing processes rather than item- or sub-item level interference (Norris, Baddeley, & Page, 2004; Page & Norris, 2003). It is assumed that a changing-state (but not steady-state) sound yields a primacy gradient of activation that usurps attentional resources required to generate a primacy gradient that represents the order of the to-be-remembered items. However, this account, unlike the perceptual-motor account, adheres to the core assumption of the original phonological loop account (Salame & Baddeley, 1982), namely, that the locus of the disruption is the passive phonological short-term store, not a subvocal-motor process (Norris et al., 2004; see also Henson, Hartley, Burgess, Hitch, & Flude, 2003).

Critical for present purposes is that the representation of order in the phonological store—the primacy gradient—is blind to the post-categorical content of items making up a sequence; item-content is coupled to a given item's point on the primacy gradient at a later, separate, stage of processing. Thus, any effect that is assumed on this model to be located at the first stage (order) and not the second stage (item-to-order coupling)—such as the impact of irrelevant sound—will be insensitive to the post-categorical content of the input (whether task-relevant or task-irrelevant). This would explain, for example, “why the IS [irrelevant sound] effect is insensitive to the phonological overlap between the to-be-remembered items and the IS” (Norris et al., 2004, p. 1103). However, this same assumption means that the primacy model (as well as other phonological store-based models that share its two-part structure; Burgess & Hitch, 2006; Hitch et al., 2003), like early incarnations of the changing-state hypothesis (e.g., Jones & Macken, 1993), cannot accommodate the fact that an overlap in the post-categorical content of the to-be-remembered and irrelevant sequence does indeed exacerbate the disruption (as long as the sequences are also order-incongruent). Thus, as discussed earlier, the present perceptual-motor interference effect suggests instead that the processes we argue are involved in the encoding of order—perceptual organization (with auditory sequences) and motor-planning—cannot be divorced from the specific content being ordered nor can they be divorced from one another (for other findings that are problematic for phonological-store based accounts of the irrelevant sound effect more generally, see Hanley & Hayes, 2012; Hanley & Shah, 2012; Jones et al., 2004).

Attentional Diversion Accounts. Another class of accounts supposes that irrelevant sound disrupts serial recall because it draws attention away from the focal task rather than because it competes specifically with the processes involved in that task (Bell et al., 2011; Cowan, 1995; Elliott, 2002; Röer, Bell, & Buchner, 2015; see also Neath, 2000). We have also argued that sound can disrupt performance due to attentional diversion—such as when a single deviating sound is detected—but that such a mechanism does not underpin the changing-state effect (e.g., Hughes, Hurlstone, Marsh, Vachon, & Jones, 2013; Hughes, Vachon, & Jones, 2005, 2007) nor the order incongruence effect studied here (Hughes & Jones, 2005). The attentional diversion account of the changing-state effect faces a number of difficulties, including its failure to explain why the effect is only found in tasks that involve or encourage a subvocal motor sequence-planning strategy (Beaman & Jones, 1997; Elliott et al., 2016; Hughes et al., 2007). It also does not explain why the effect is eliminated if such planning is precluded via articulatory suppression (Jones et al., 2004; Hanley, 1997; present Experiment 2).

Of most relevance in the present context is that the attentional diversion account has also been applied to the order incongruence effect. Bell et al. (2011) argued that recall is disrupted more by irrelevant spoken items that are postcategorically identical to the to-be-remembered items because the activation of such irrelevant items is primed by the to-be-remembered hence rendering the irrelevant items particularly salient and attention-diverting (see Cowan, 1995). An immediate problem for this account, however, is that the order incongruence effect, by definition, refers to the finding that item-similarity is not sufficient for disruption; the items must also be in an incongruent sequence, that is, the very same items fail to cause additional disruption compared to dissimilar items if they are order-congruent with the to-be-remembered list. Bell et al. (2011), however, sought to explain away the apparent importance of order incongruence. They argued that the poorer performance found in an order-incongruent condition compared to an order-congruent condition occurs because order *congruence facilitates* the subvocal serial rehearsal of the to-be-remembered list, not because order incongruence impairs it. Thus, on their attentional diversion account, the general claim is that items that are similar to the to-be-remembered items divert attention more than dissimilar items. If the similar items are order-congruent with the to-be-remembered list, such diversion facilitates performance; if not, the diversion impairs performance. Thus, in this view, order *in*congruence does not play any role in the effect.

A difficulty for Bell et al.'s (2011) account, however, is that if an order-congruent irrelevant sequence facilitates rehearsal then recall in the presence of such a sequence should not only be better than in an order-incongruent condition but also better than under control (i.e., dissimilar-items or quiet) conditions. This is not the case: A defining empirical feature of what we maintain is an order *in*congruence effect is that—as replicated in the present Experiment 2—no difference is found between an order-congruent and a dissimilar-items condition. Moreover, the account cannot explain why performance is still markedly poorer in an order-congruent condition than in a quiet condition (current Experiment 2; Hughes & Jones, 2005) while, on the perceptual-motor account, this can be readily attributed to a changing-state effect (cf. Jones & Macken, 1993) that would be operative regardless of any additional order-incongruence based interference. There is no evidence, therefore, that order-congruence facilitates rehearsal. The only evidence provided by Bell et al. (2011) that it does is very indirect: The authors found a difference between order-congruent and order-incongruent conditions with a lag between the irrelevant and relevant sequences of two items (as used by Hughes and Jones, 2005) but not five items, and suggested that in the latter case the order-congruent sequence would not be expected to facilitate rehearsal (although no independent evidence for this assumption was provided). In the present Experiment 2, however, we showed that an order incongruence effect is indeed produced with a lag of five (and also a lag of four in Experiment 1). Furthermore, while the account correctly predicts that blocking rehearsal via articulatory suppression should reduce performance in an order-congruent condition (because the order-congruent sequence could no longer be used to facilitate rehearsal), it does not seem capable of explaining why similar items (regardless of congruence) no longer disrupt performance compared to dissimilar items or compared to quiet.[3] In sum, whilst there is little doubt that attentional diversion is one mechanism of distraction by irrelevant sound (e.g., Hughes, 2014; Hughes et al., 2013; Hughes & Marsh, 2015; Röer et al., 2013), we suggest that it does not provide an adequate way of accounting for the changing-state effect or the order

[3] Bell et al. (2011, Experiments 3 and 4) also reported that in the context of serial recall of letter-lists, spoken letter-distractors—but ones from a set different from that used for the to-be-remembered lists—were more disruptive than digit-distractors. Whereas this would seem difficult to explain in terms of our perceptual-motor interference account, their study involved a ‘non-pure’ serial recall task in which any given list comprised 8 letters out of a possible 16 and in which participants were free to respond with any letter on the keyboard. Both these features of the design are likely to promote the probability of item (as opposed to serial order) errors such as omissions and the intrusion of distractor-items. And indeed, as noted by the authors themselves, the greater disruptive effect of same- *vs.* different-category distractors was attributable entirely to such item errors. As such, the category-similarity effect they reported may reflect long-term memory item-retrieval processes manifesting in a putative short-term memory task, not an effect that speaks specifically to short-term memory (cf. Baddeley, 2012; note that a similar difficulty arises in relation to a study by Eagen and Chein, 2012, which showed an exacerbation of disruption by irrelevant sound attributable to subphonemic, featural, overlap).

incongruence effect studied here. Instead, we argue that these latter two effects reflect different levels of the same competition process between order cues yielded by the perceptual organization of the sound and the sequence-planning processes involved in the focal serial short-term memory task.

A Functional Approach to Short-Term Memory

The present results add to a growing body of work supporting an approach to short-term memory performance that emphasizes the contribution of general-purpose perceptual and motor functions as opposed to dedicated mnemonic structures or mechanisms (e.g., Hughes et al., 2009, 2011, 2016; Jones et al., 2006, 2004, 2007; Maidment & Macken, 2012). There is now a good deal of converging evidence, for example, that the primary evidence for the notion of a specific verbal short-term store—the phonological similarity effect—has a motoric not phonological basis (Acheson & MacDonald, 2009a; Jones et al., 2004, 2006, 2007; Maidment & Macken, 2012). The phonological similarity effect refers to the finding that a list of items that sound alike (e.g., B G D C […]) is far more poorly recalled than a list of items that sound dissimilar (e.g., F H J Q […]; Baddeley, 1966; Conrad, 1964). The classical, structuralist, explanation of the phenomenon is that the similar items are more easily confused during retrieval from a dedicated phonological store (Baddeley, 1986, 2007). The possibility that the effect instead has a motor output-planning locus has, historically, been dismissed on the basis that the effect is still found even when motor-planning is impeded via articulatory suppression (e.g., Baddeley et al., 1984) so long as the to-be-remembered items are presented auditorily. Classically, the interpretation of this interaction has been that there exists a phonological store separate from articulatory processes and to which auditory, but not visual, input enjoys direct, obligatory, access (Baddeley, 1986, 2007).

It turns out, however, that the residual phonological similiarity effect under suppression with auditory presentation is located primarily in recency (the last one or two items in the list), an effect that can be explained in terms of the opportunistic use of the acoustic (not phonological) processing of the end-boundary of an auditory sequence (see Jones et al., 2006, 2004; Maidment & Macken, 2012). The fact that motor-planning is, after all, a prerequisite for the 'phonological' similarity effect regardless of input-modality (notwithstanding the residual acoustic-based effect at recency) indicates strongly that the locus of the effect is that motor-planning process. More direct evidence for this conclusion comes from the observation that the precise pattern of item-exchanges when recalling a phonologically similar list mimics very closely the 'slips of the tongue' produced occasionally during natural speech production, or that can be induced in the laboratory, despite little or no memory load (Acheson & MacDonald, 2009a; Ellis, 1980; Page, Madge, Cumming, & Norris, 2007).

In conclusion, the present findings suggest that the inherent openness of a motor-plan assembled opportunistically in the face of an under-specification of action-parameters problem (i.e., a novel sequence to be reproduced) renders short-term serial memory vulnerable to interference from extraneous sequences derived from the passive organization of the auditory scene into streams. We argue that the sort of interference effects studied here do not therefore, contrary to a commonly held view, reveal that there are specific short-term memory mechanisms that are short-term precisely because they are quintessentially fragile. Rather, they reflect the action of a perceptual-motor integration process that is, in general, perfectly functional. More generally, we suggest that limits on short-term serial memory performance reflect the use of processes that are perfectly well-designed to fulfil the functions for which they evolved—those of perceptual organization of the environmental input into objects and the planning of coherent motor-actions—coming into play in the context of a task-setting for which they are not necessarily well suited.

.References

Acheson, D. J., & MacDonald, M. C. (2009a). Twisting tongues and memories: Explorations of the relationship between language production and verbal working memory. *Journal of Memory & Language, 60*, 329-350.

Acheson, D. J., & MacDonald, M. C. (2009b). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, *135*, 50-68.

Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology*. *Human Perception and Performance*, *11*(3), 257-271.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Quarterly Journal of Experimental Psychology, 18,* 362–365.

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

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189-208.

Baddeley, A. D. (2007). *Working memory, thought and action.* Oxford, England: Oxford University Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York, NY: Academic Press.

Baddeley, A. D. (1992). Working memory. *Science*, *255,* 556-559.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63,* 1–29.

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly*

*Journal of Experimental Psychology: Human Experimental Psychology, 36*A, 233–252.

Beaman, C. P. (2005). Auditory distraction from low-intensity noise: A review of the consequences for learning and workplace environments. *Applied Cognitive Psychology, 19*(8), 1041-1064.

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

Bell, R., Mund, I., & Buchner, A. (2011). Disruption of short-term memory by distractor speech: Does content matter? *Quarterly Journal of Experimental Psychology, 64*, 146–168.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organisation of sound.* Cambridge, MA: MIT Press.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89,* 244–249.

Bridges, A. M., & Jones, D. M. (1996). Word dose in the disruption of serial recall by irrelevant speech: Phonological confusions or changing-state? *Quarterly Journal of Experimental Psychology, 49A,* 919-939.

Buchsbaum, B. R., & D'Esposito, M. (2008). The search for the phonological store: From loop to convolution. *Journal of Cognitive Neuroscience, 20,* 762–778.

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

Campbell, T. A., Beaman, C. P., & Berry, D. C. (2002a). Auditory memory and the irrelevant sound effect: Further evidence for changing-state disruption. *Memory, 10*, 199-214.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning & Verbal Behavior, 15*, 17-32.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55*, 75–84.

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

Divin, W., Coyle, K., & James, D. T. T. (2001). The effects of irrelevant speech and articulatory suppression on the serial recall of silently presented lipread digits. *British Journal of Psychology, 92*, 593-616.

Ellermeier, W., Kattner, F., Ueda, K., Doumoto, K., & Nakajima, Y. (2015). Memory disruption by irrelevant noise-vocoded speech: Effects of native language and the number of frequency bands. *The Journal of the Acoustical Society of America, 138*(3), 1561-1569.

Elliott, E. M. (2002). The irrelevant-speech effect and children: Theoretical implications of developmental change. *Memory & Cognition, 30,* 478–487.

Elliott, E. M., Hughes, R. W., Briganti, A., Joseph, T. N., Marsh, J. E., & Macken, B. (2016). Distraction in verbal short-term memory: Insights from developmental differences. *Journal of Memory and Language*, *88,* 39-50.

Gathercole, S. E., & Baddeley, A. D. (1993). Working memory and language. Hillside, NJ: Lawrence Erlbaum Associates.

Hanley, J. R. (1997). Does articulatory suppression remove the irrelevant speech effect? *Memory*, *5,* 423–431.

Hanley, J. R., & Hayes, A. (2012). The irrelevant sound effect under articulatory suppression: Is it a suffix effect? *Journal of Experimental Psychology: Learning, Memory & Cognition*, *38,* 482-487.

Hanley, J. R., & Shah, N. (2012). The irrelevant sound effect under articulatory suppression is a suffix effect even with 5-item lists. *Memory*, *20*, 415-419.

Henson, R., Hartley, T., Burgess, N., Hitch, G., & Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing signal hypothesis. *Quarterly Journal of Experimental Psychology, 56*A, 1307–1334.

Hommel, B. (2010). Grounding attention in action control: The intentional control of selection. In B. Bruya (Ed.), *Effortless attention: A new perspective in the cognitive science of attention and action* (pp. 121-140). Cambridge, MA: MIT Press.

Hughes, R. W. (2014). Auditory distraction: A duplex-mechanism account. *PsyCh Journal*, *3*(1), 30-41.

Hughes, R. W., Chamberland, C., Tremblay, S., & Jones, D. M. (2016). Perceptual-motor determinants of auditory-verbal serial short-term memory. *Journal of Memory & Language*, *90*, 126-146.

Hughes, R. W., Hurlstone, M., Marsh, J. E., Vachon, F., & Jones, D. M. (2013). Cognitive control of auditory distraction: Impact of task difficulty, foreknowledge, and working memory capacity supports duplex-mechanism account. *Journal of Experimental Psychology: Human Perception & Performance, 39*, 539-553.

Hughes, R., & Jones, D. M. (2001). The intrusiveness of sound: Laboratory findings and their implications for noise abatement. *Noise & Health*, *4*(13), 51-70.

Hughes, R. W., & Jones, D. M. (2003). A negative order-repetition priming effect: Inhibition of order in unattended auditory sequences? *Journal of Experimental Psychology: Human Perception & Performance, 29,* 199-218.

Hughes, R. W., & Jones, D. M. (2005). The impact of order incongruence between a task-irrelevant auditory sequence and a task-relevant visual sequence. *Journal of Experimental Psychology: Human Perception & Performance, 31,* 316–327.

Hughes, R. W., & Marsh, J. E. (2015). *Auditory Attentional Capture During Offline Cognition: A Reprieve for Classical Attention Theory and the Sentinel of the Senses.* Manuscript submitted for publication.

Hughes, R. W., Marsh, J. E., & Jones, D. M. (2009). Perceptual-gestural (mis)mapping in serial short-term memory: The impact of talker variability. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35,* 1411–1425.

Hughes, R. W., Marsh, J. E., & Jones, D. M. (2011). Role of serial order in the talker variability effect in short-term memory: Testing a perceptual organization-based account. *Memory & Cognition, 39,* 1435-1447.

Hughes, R. W., Tremblay, S., & Jones, D. M. (2005). Disruption by speech of serial short-term memory: The role of changing-state vowels. *Psychonomic Bulletin & Review*, *12,* 886–890.

Hughes, R. W., Vachon, F., & Jones, D. M. (2005). Auditory attentional capture during serial recall: Violations at encoding of an algorithm-based neural model? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 736-749.

Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 33,* 1050-1061.

Jones, D. M., Alford, D., Bridges, A., Tremblay, S., & Macken, W. J. (1999). Organizational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25,* 464–473.

Jones D. M., Farrand, P., Stuart, G., Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology:*

Learning, Memory, & Cognition, 21, 1008–1018.

Jones, D. M., Hughes, R. W., & Macken, W. J. (2006). Perceptual organization masquerading as phonological storage: Further support for a perceptual– gestural view of short-term memory. *Journal of Memory & Language, 54,* 265–281.

Jones, D. M., Hughes, R. W., & Macken, W. J. (2007). The phonological store abandoned. *Quarterly Journal of Experimental Psychology, 60,* 497–504.

Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19,* 369-381.

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

Jones, D. M., & Macken, W. J. (1995b). Phonological similarity in the irrelevant speech effect: Within- or between-stream similarity? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21,* 103-115.

Jones, D. M., Macken, W. J., & Murray, A. C. (1993). Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Memory & Cognition, 21,* 318–328.

Jones, D. M., & Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological, and is it a store? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30,* 656-674.

Jones, D. M., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *Quarterly Journal of Experimental Psychology, 44A*, 645-669.

Jones, D. M., & Tremblay, S. (2000). Interference by process or content? A reply to Neath (2000). *Psychonomic Bulletin & Review, 7,* 550-558.

Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition, 144,* 1-13.

Klatte, M., Kilcher, H., & Hellbrück, J. (1995). Wirkungen der zeitlichen Struktur von Hintergrundschall auf das Arbeitsgedächtnis und ihre theoretischen und praktischen Implikationen. *Zeitschrift für Experimentelle Psychologie, 42,* 517-544.

Koffka, K. (1935). *Principles of Gestalt Psychology,* Harcourt Brace.

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review, 111*(4), 880-913.

Lackner, J. R., & Goldstein, L. M. (1974). Primary auditory stream segregation of repeated consonant–vowel sequences. *The Journal of the Acoustical Society of America, 56,* 1651–1652.

Larsen, J. D., & Baddeley, A. D. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source? *Quarterly Journal of Experimental Psychology, 56*(8), 1249-1268.

Larsen, J. D., Baddeley, A. D., & Andrade, J. (2000). Phonological similarity and the irrelevant speech effect: Implications for models of short-term verbal memory. *Memory, 8,* 145-157.

LeCompte, D. C., & Shaibe, D. M. (1997). On the irrelevance of phonological similarity to the irrelevant speech effect. *Quarterly Journal of Experimental Psychology, 50A,* 100-118.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences, 22,* 1-38.

MacDonald, M. C. (2016). Speak, act, remember: The language-production basis of serial order and maintenance in verbal memory: *Current Directions in Psychological Science, 25* (1), 47-53.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109,* 35–54.

Macken, W. J., Phelps, F. G., & Jones, D. M. (2009). What causes auditory distraction? *Psychonomic Bulletin & Review, 16*(1), 139-144.

Macken, B., Taylor, J., & Jones, D. (2015). Limitless capacity: a dynamic object-oriented approach to short-term memory. *Frontiers in Psychology*, *6*.

Macken, W. J., Taylor, J. C., & Jones, D. M. (2014). Language and short-term memory: The role of perceptual-motor affordance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1257-1270.

Macken, B., Taylor, J., Kozlov, M., Hughes, R. W., & Jones, D. M. (2016). *Memory as Embodiment: The Case of Modality and Short-Term Memory.* Manuscript submitted for publication.

Maidment, D.W., & Macken, W. J. (2012). The ineluctable modality of the audible: perceptual determinants of auditory-verbal short-term memory. *Journal of Experimental Psychology: Human Perception & Performance*, *38,* 989-997.

Maidment, D.W., Macken, B., & Jones, D.M. (2013). Modalities of memory: Is reading lips like hearing voices? *Cognition, 129,* 471-493.

Marsh, J. E., Vachon, F., & Jones, D. M. (2008). When does between-sequence phonological similarity promote irrelevant sound disruption? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 243-248.

Marsh, J. E., Beaman, C. P., Hughes, R. W., & Jones, D. M. (2012). Inhibitory control in memory: Evidence for negative priming in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1377-1388.

Marsh, J. E., Beaman, C. P., Hughes, R. W., & Jones, D. M. (2012). Inhibitory control in memory:

Evidence for negative priming in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(5), 1377-1388.

Maybery, M. T., Parmentier, F. B. R., & Jones, D. M. (2002). Grouping of list items reflected in the timing of recall: Implications for models of serial verbal memory. *Journal of Memory & Language, 47*, 360–385.

Moore, B. C., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica, 88,* 320-333.

Melby-Lervåg, M & Hulme, C. (2010). Serial and free recall in children can be improved by training: Evidence for the importance of phonological and semantic representations in immediate memory tasks. *Psychological Science, 21,* 1694-1700.

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology, 78*, 679-684

Nadeau, S. E. (2001). Phonology: A review and proposals from a connectionist perspective. *Brain & Language, 79,* 511–579.

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

Neisser, U. (1967). *Cognitive Psychology.* New York: Appleton Century Crofts.

Neumann, O. (1987). Beyond capacity: A functional view of attention. In H. Heuer and A. F. Sanders (Eds.), *Perspectives on perception and action*. Hillsdale, NJ: Erlbaum.

Neumann, O. (1996). Theories of attention. In O. Neumann and A. F. Sanders (Eds.), *Handbook of perception and action* (Vol. 3) (pp. 389-446). London, England: Academic Press.

Norris, D. G., Baddeley, A. D., & Page, M. P. A. (2004). Retroactive effects of irrelevant speech on serial recall from short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30,* 1093-1105.

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

Page, M. P. A., & Norris, D. G. (2003). The irrelevant sound effect: What needs modelling, and a tentative model. *Quarterly Journal of Experimental Psychology, 56(A),* 1289-1300.

Perham, N., & Vizard, J. (2011). Can preference for background music mediate the irrelevant sound effect? *Applied Cognitive Psychology*, *25*, 625–631.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*, 23–38.

Repovs, G., & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience*, *139*, 5-21.

Röer, J. P., Bell, R., & Buchner, A. (2015). Specific foreknowledge reduces auditory distraction by irrelevant speech. *Journal of Experimental Psychology: Human Perception and Performance, 41*(3), 692.

Rosenbaum, D. A. (2009). *Human motor control*. Academic Press.

Saito, S., & Baddeley, A. D. (2004). Irrelevant sound disrupts speech production: Exploring the relationship between short-term memory and experimentally induced slips of the tongue. *The Quarterly Journal of Experimental Psychology*, *57*, 1309-1340.

Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning & Verbal Behavior*, *21*, 150-164.

Salamé, P., & Baddeley, A. D. (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology, 41A,* 107-122.

Salamé, P., & Baddeley, A. D. (1990). The effects of irrelevant speech on immediate short-term memory. *Bulletin of the Psychonomic Society, 28,* 540-542.

Schlittmeier, S., Hellbrück, J., & Klatte, M. (2008). Does irrelevant music cause an irrelevant sound effect for auditory items? *European Journal of Cognitive Psychology*, *20*, 252-271.

Sjöblom, A. M., & Hughes, R. W. (2016). *Verbal Sequence Learning Without a Phonological Store: A Perceptual-Motor Approach.* Manuscript in preparation.

Sörqvist, P. (2010). High working memory capacity attenuates the deviation effect but not the changing-state effect: Further support for the duplex-mechanism account of auditory distraction. *Memory & Cognition*, *38,* 651-658.

Sternberg, S., Wright, C. E., Knoll, R. L., & Monsell, S. (1980). Motor programs in rapid speech: Additional evidence. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 507–534). Hillsdale, NJ: Erlbaum.

Sussman, E. S., Bregman, A. S., & Lee, W. W. (2014). Effects of task-switching on neural representations of ambiguous sound input. *Neuropsychologia*, *64*, 218-229.

Sussman, E. S., Horváth, J., Winkler, I., & Orr, M. (2007). The role of attention in the formation of auditory streams. *Perception & Psychophysics*, *69*, 136-152.

Taylor, J. C., Macken, B., & Jones, D. M. (2015). A matter of emphasis: Linguistic stress habits modulate serial recall. *Memory & Cognition*, 43(3), 520-537.

Tipper, S. P. (2001). Does negative priming reflect inhibitory mechanisms? A review and integration of conflicting views. *Quarterly Journal of Experimental Psychology, 54*(2), 321-343.

Treisman, A. M. (1971). Shifting attention between the ears. *Quarterly Journal of Experimental Psychology, 23*(2), 157-167. doi:10.1080/14640747108400236

Tremblay, S., & Jones, D. M. (1998). Role of habituation in the irrelevant sound effect: Evidence from the effects of token set size and rate of transition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*, 659–671.

Tremblay, S., Macken, W. J., & Jones, D. M. (2001). The impact of broadband noise on serial memory: Changes in band-pass frequency increase disruption. *Memory, 9*, 323-331.

Tremblay, S., Nicholls, A. P., Alford, D., & Jones, D. M. (2000). The irrelevant sound effect: Does speech play a special role? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 26, 1750-1754.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of*

Van Noorden, L. (1975). *Temporal coherence in the perception of tone sequences* (Doctoral dissertation, Technische Hogeschool Eindhoven).

Warren, R. M. (1999). *Auditory perception: A new analysis and synthesis.* New York, NY: Cambridge University Press.

Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. (1969). Auditory sequence: Confusion of patterns other than speech or music. *Science, 164,* 586–587.

Waugh, N. C., & Norman, D. A. (1965). Primary Memory. *Psychological Review, 72,* 89-104.

Wilson, M., & Fox, G. (2007). Working memory for language is not special: Evidence for an articulatory loop for novel stimuli. *Psychonomic Bulletin & Review, 14*(3), 470-473.

Winkler, I., Denham, S.L., & Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences, 13,* 532–540.

Woodward, A. J., Macken, W. J., & Jones, D. M. (2008). Linguistic familiarity in short-term memory: A role for (co-)articulatory fluency? *Journal of Memory & Language, 58*, 48–65.

Author note

Robert W. Hughes, Department of Psychology, Royal Holloway, University of London, Egham, UK. John E. Marsh, School of Psychology, University of Central Lancashire, UK. John E. Marsh is also a Visiting Researcher at the Department of Building, Energy, and Environmental Engineering, University of Gävle, Gävle, Sweden. The two authors contributed equally to the work presented here. We are grateful to Christopher Thomas, Eleanor Wong and Eleanor Coles for data collection. The results of Experiment 2 were presented at the *Auditory Perception Cognition and Action Meeting*, Toronto, Canada, November, 2013, and the results of both experiments were presented at the joint *Experimental Psychology Society/SEPEX* meeting, Oxford, UK, July, 2016. The research received financial support from an *Experimental Psychology Society* grant awarded to R. W. Hughes and a bursary awarded by the *Centre for Research-Informed Teaching* at the University of Central Lancashire to J. E. Marsh which provided Christopher Thomas and Eleanor Wong with internships. Correspondence may be addressed to R. W. Hughes (Rob.Hughes@rhul.ac.uk).

Figure 1

TO-BE-REMEMBERED LIST (visual): 5 1 3 7 8 4 6 2

Auditory condition:

<u>No-Alternation</u> (Female or male voice/right or left ear)

1. ***Order-congruent*** 8 4 6 2 5 1 3 7

2. ***Order-incongruent*** 8 5 4 1 6 3 2 7

<u>With-Alternation</u>

3. ***Order-congruent*:** Male voice-left ear: 8 6 5 3

Female voice-right ear: 4 2 1 7

4. ***Order-incongruent*:** Male voice-left ear: 8 4 6 2

Female voice-right ear: 5 1 3 7

Figure 2

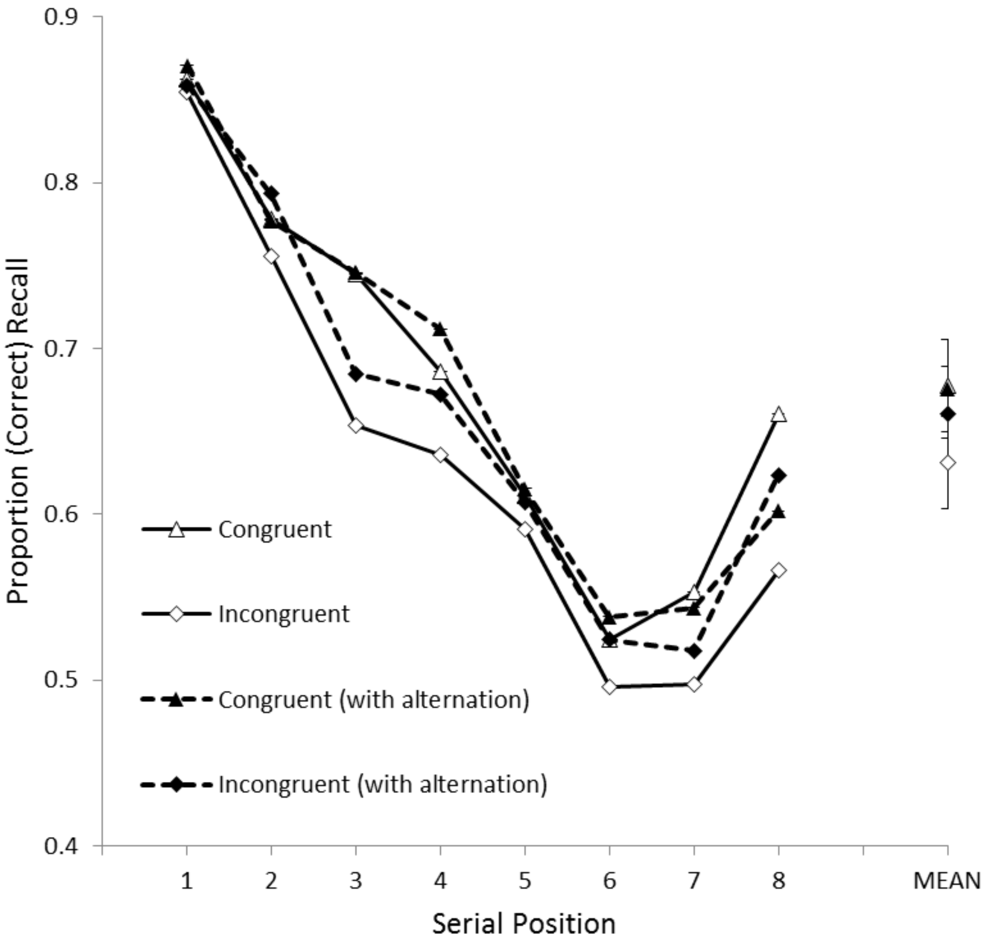
0.9
0.8
0.7
0.6
0.5
0.4
Proportion (Correct) Recall
Congruent
Incongruent
Congruent (with alternation)
Incongruent (with alternation)
1
2
3
4
5
6
7
8
MEAN
Serial Position

Figure 3

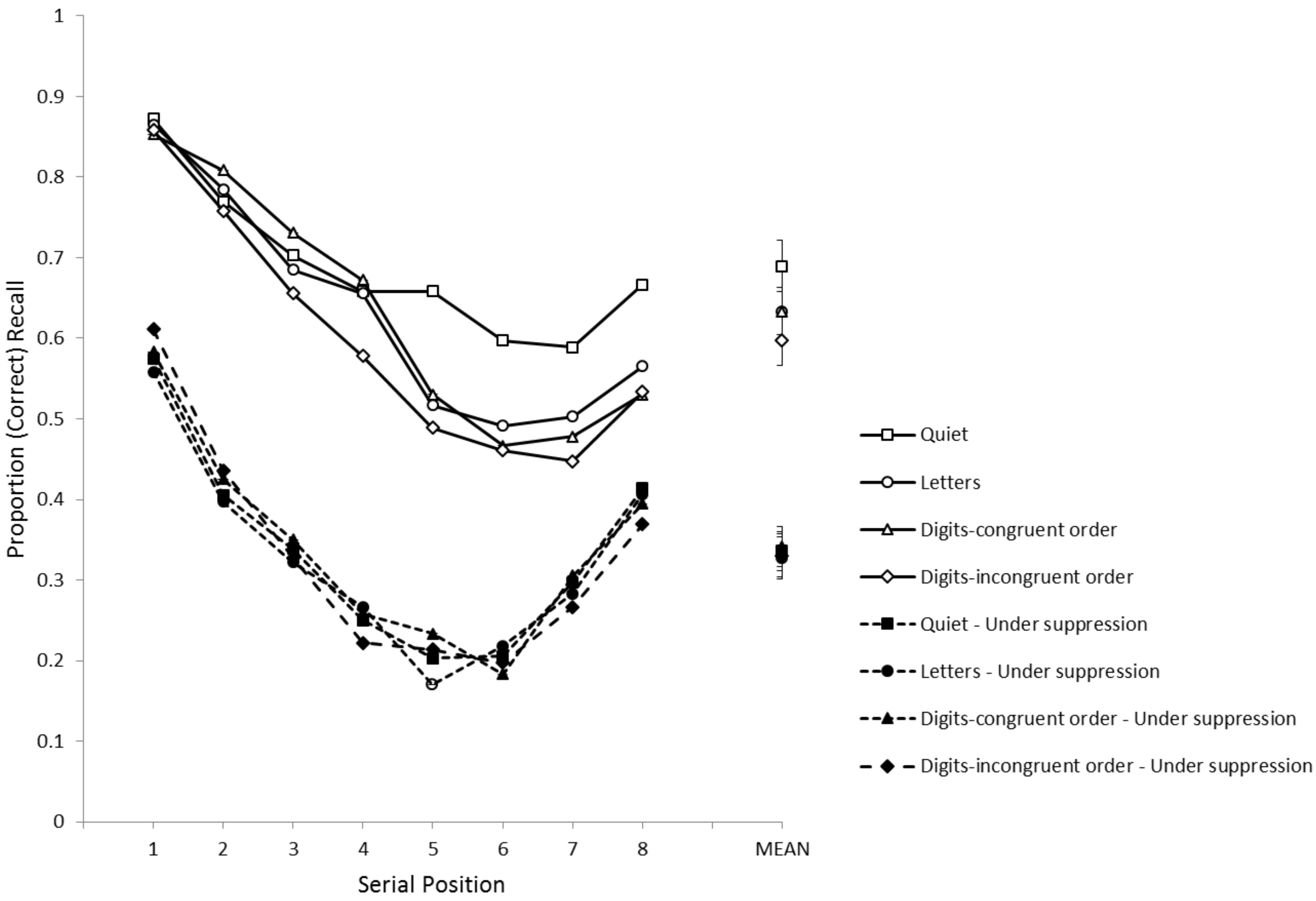
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Proportion (Correct) Recall
1
2
3
4
5
6
7
8
MEAN
Serial Position
Quiet
Letters
Digits-congruent order
Digits-incongruent order
Quiet - Under suppression
Letters - Under suppression
Digits-congruent order - Under suppression
Digits-incongruent order - Under suppression

Figure Captions

Figure 1. Schematic illustration of the four conditions of Experiment 1.

Figure 2. Proportion correct serial recall at each serial position in the four conditions (order-congruence × alternation)—as well as the mean (and standard error) collapsed across serial positions—in Experiment 1.

Figure 3. Proportion correct serial recall at each serial position in the eight conditions (auditory condition × articulatory suppression)—as well as the mean (and standard error) collapsed across serial positions—in Experiment 2.