Author's post-print of: 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.

Perceptual-Gestural (Mis)mapping in Serial Short-Term Memory: The Impact of Talker Variability

Robert W. Hughes, John E. Marsh, & Dylan M. Jones

School of Psychology, Cardiff University, Cardiff, UK.

RUNNING HEAD: Perceptual-Gestural (Mis)mapping in Serial Recall

Correspondence: Robert W. Hughes, School of Psychology,

Cardiff University, Cardiff CF10 3AT, United Kingdom,

Phone: (+44) 029 20 876904, Fax: +44 029 20874858,

email: HughesRW@cardiff.ac.uk

Perceptual-Gestural (Mis)mapping in Serial Short-Term Memory: The Impact of Talker Variability

The mechanisms underlying the poorer serial recall of talker-variable lists (e.g., alternating female-male voices) as compared with single-voice lists were examined. We tested the novel hypothesis that this *talker variability effect* arises from the tendency for perceptual organization to partition the list into streams based on voice such that the representation of order maps poorly onto the formation of a gestural sequence-output plan assembled in support of the reproduction of the true temporal order of the items. In line with the hypothesis, the presence of a spoken lead-in designed to further promote by-voice perceptual partitioning accentuates the effect (Experiments 1 and 2); the impairment is larger the greater the acoustic coherence between non-adjacent items: alternating-voice lists are more poorly recalled than four-voice lists (Experiment 3); and talker variability combines non-additively with phonological similarity, consistent with the view that both variables disrupt sequence output-planning (Experiment 4). The results support the view that serial short-term memory performance reflects the action of sequencing processes embodied within general-purpose perceptual input-processing and gestural output-planning systems.

KEYWORDS: Short-Term Memory; Talker Variability; Serial Recall; Perceptual-Gestural View; Perceptual Organization; Embodied Cognition.

One major class of accounts of serial short-term memory is centred at the item level and assumes that explanations of serial behavior will flow from an understanding of item-level properties such as the rate of item decay or/and the structural (e.g., phonological) similarity of one item to another (e.g., Baddeley, 1986, 2007; Farrell & Lewandowsky, 2002; Nairne, 1990; Neath, 2000). However, an alternative framework—the *perceptual-gestural account*—focuses on factors that operate at a level superordinate to the item, at the level of sequence formation, both at input (particularly in the auditory modality, in the formation of streams), and at motor output planning in the formation of a sequence of subvocal gestures (Hughes & Jones, 2005; Jones, Macken, & Nicholls, 2004; Jones, Hughes, & Macken, 2006, 2007; Woodward, Macken, & Jones, 2008). In the present article, we sought to illustrate the importance of sequence-level factors in understanding serial short-term memory performance by capitalizing on the disruptive impact on auditory-verbal serial recall of presenting lists in more than one voice, particularly alternating female-male voices (Greene, 1991). Evidence is presented that this *talker variability effect* results from the fact that obligatory formation of voice-based streams (auditory streaming; see Bregman, 1990) produces a poor mapping between the perception of order (within streams) and the need to assemble the items into a gestural (articulatory) sequence output-plan that mimics the canonical order of the items.

Serial Short-Term Memory: A Perceptual-Gestural Account

Current understanding of serial verbal short-term memory is based largely on the serial recall paradigm in which a list of verbal items (e.g., digits, letters, words) must be recalled in strict serial order (Conrad, 1964, Baddeley, 1966). Classically, explanations of serial recall have tended to focus on the properties of the individual list items. For instance, according to the decay-rehearsal approach to verbal short-term memory (e.g., the phonological loop model; Baddeley, 1986, 2007; Baddeley & Hitch, 1974; see also Cowan, 1995), verbal items enter a passive, bespoke, phonological store dedicated to the temporary retention of verbal events (Baddeley, 1986, 2007). Items in the store are subject to decay within a few seconds unless refreshed by a separate articulatory control process and are also susceptible to mutual interference by virtue of their structural (e.g., phonological) similarity (e.g., Baddeley, 1986). Other accounts posit that similarity-based item-interference can account for serial recall phenomena without the notion of item-decay (e.g., Nairne, 1990; Lewandowsky, Geiger, & Oberauer, 2008). An alternative view—the perceptual-gestural account—holds that serial recall performance is parasitic on general-purpose perceptual and motor-planning processes that operate at the level of the sequence, not each item (e.g., Hughes & Jones, 2005; Jones et al., 2004, 2006, 2007; Woodward et al., 2008).

An important feature of the typical serial recall study is that the burden of processing lies with reproducing the order of the items: A familiar closed set of items is typically used on each trial (e.g., permutations of *1-8*) and hence the individual items are known before presentation (hereafter: '*pure* serial recall'). The starting point for the perceptual-gestural view is the characterization of the typical serial recall list as a sequence in which the intrinsic transitional probabilities between successive items—the predictability of an event given the preceding event(s); cf. Miller & Chomsky (1963)—are, by design, very low: Syntax, grammar, and semantics which in natural language constrain temporal order are stripped from the serial recall list (see Jefferies, Lambon Ralph, & Baddeley, 2004; Macken & Jones, 2003). It is this feature that makes serial recall difficult. For example, performance is superior when there is a good match between the list and long-term sequential knowledge such as for lists containing high-frequency letter transitions (Miller & Selfridge, 1950), lists of words that make up a grammatically-legitimate sentence

(Jefferies et al., 2004), or lists of adjective-noun pairings compared to less frequently encountered (in English) noun-adjective pairings (Perham, Marsh, & Jones, in press).

In the absence of strong order cues in the typical serial recall list, how is serial order preserved? We argue that transitional probabilities are grafted onto the material by co-opting the generic skill of gestural sequence-planning (vocal-articulatory in the case of verbal material; Macken & Jones, 2003). The skill of speech-planning lends itself well to this task due to its inherent sequentiality and its range of paralinguistic sequencing sub-skills such as prosody and co-articulation (Sternberg, Wright, Knoll, & Monsell, 1982). For example, prosodic characteristics of timing and intonation deployed in natural phrase and sentence production may be brought to bear to minimize transitional probabilities across group boundaries. This provides strong cues to order at boundaries that constrain migration of items across groups (e.g., Maybery, Parmentier, & Jones, 2002). Moreover, the skill of co-articulation—adjusting the way in which the end of one speech element (e.g., syllable, word) is articulated so as to lend fluency to the transition to the next element (Sternberg et al., 1982) also serves to increase transitional probabilities between successive items (as shown by the finding that serial recall is a positive function of co-articulatory fluency; Murray & Jones, 2002; Woodward et al., 2008). Thus, an articulatory plan is assembled as a surrogate for the lack of correspondence between long-term sequence knowledge and the to-be-reproduced sequence; the motor-plan is the very agent by which item-order is realised and supported (see also Neumann, 1996). Explanations of serial recall phenomena are sought, therefore, by recourse to factors that emerge at the level of the articulated sequence and which are not, by definition, to be found at the local item-level.

The assembly of the articulatory plan is time-critical. In their presentation, sequences are paced and the items evanescent. The particular items need to be loaded successfully and in a timely and orderly fashion onto the necessarily abstract and generic motor sequence-plan, a process involving a number of conversion operations which could include transformation from graphemic form, co-articulation, and so forth. Thus, numerous factors such as spatial or temporal uncertainty during presentation or the complexity of the gestures required to (co)articulate successive items can potentially render the loading process more fraught and hence compromise serial recall performance (e.g., Murray & Jones, 2002). Of particular relevance in the present article, when items are presented auditorily, powerful processes of perceptual organization may also influence the process of picking up and embodying elements in the plan in their correct order.

The perceptual component of the perceptual-gestural account is based largely upon Bregman's (1990) *auditory scene analysis* framework: This refers to the preattentive and obligatory (i.e., non-volitional) partitioning of the mixture of acoustic signals reaching the ears into discrete mental descriptions (streams) corresponding to each individual sound-emitting event contributing to that mixture (for evidence for the obligatory nature of auditory streaming, see, e.g., Deouell, Deutsch, Scabini, Soroker, & Knight, 2008; Macken, Tremblay, Houghton, Nicholls, & Jones, 2003). An aspect of streaming with particular relevance for serial recall is sequential streaming: the computation of whether or not temporally successive auditory stimuli share a common origin, a task accomplished by exploiting Gestalt-like grouping principles such as similarity of frequency, timbre, and good continuation (see Bregman, 1990, Chapter 2; Warren, 1999). Thus, successive events with a relatively low acoustic-level transitional probability—such as would be the case with the alternation of two tones highly distinct in pitch (or fundamental frequency)—are relatively unlikely to be computed as having the same origin and hence will tend to be partitioned to form two distinct streams (e.g., Miller & Heise, 1950; Rogers & Bregman, 1993; van Noorden, 1975). In such cases, "the auditory system is grouping tones that are similar to one another in preference to grouping tones that follow one another immediately in time” (Bregman, 1990, p. 45). In the present article, we sought to demonstrate how such perceptual grouping of elements by spectral similarity may influence—in this case impair—the formation of the articulatory plan assembled in the service of order retention by examining the disruptive impact of talker variability on serial recall (e.g., Greene, 1991).

The Talker Variability Effect in Serial Recall

Auditory-verbal serial recall is impaired if the list is conveyed in more than one voice (e.g., Goldinger, Pisoni, & Logan, 1991; Greene, 1991; Martin, Mullennix, Pisoni, & Summers, 1989; Nygaard, Sommers, & Pisoni, 1995). Previous accounts of this talker variability effect (TVE) in serial recall[1] have appealed to essentially the same explanation that the decay-rehearsal model offers for the word-length effect (e.g., Baddeley, Thomson, & Buchanan, 1975): The increased time taken to encode or/and rehearse talker-variable items—just as with long compared to short words—impairs recall by delaying the opportunity to refresh decay-prone items held in a bespoke phonological store. On one such account, the item-refresh delay results from talker-variable lists imposing a greater burden on talker normalization (cf. Pisoni, 1997) whereby extra-linguistic, indexical, properties such as the accent, gender, speaking style, and emotional state of the particular speaker (see Abercrombie, 1967) needs to be discarded so as to yield abstract, canonical, linguistic (i.e., phonological) item representations (Martin et al., 1989). In another account within this approach, the item-refresh delay is due to a process of incorporating the indexical properties rather than discarding them, a process which would again be under greater duress with more than one talker (Goldinger et al., 1991).

An alternative view based on the perceptual-gestural account is that the TVE reflects a mismapping between perceptual organization and the assembly of an articulatory plan. In a single-voice list, adjacent items are likely to be perceptually grouped together on account of their similarity in fundamental frequency and timbre as well as indexical properties such as accent (Pardo & Remez, 2006). In a talker-variable list, however, the transitional probabilities among immediately adjacent items—and hence the likelihood of the items being grouped together—may be greatly diminished. Indeed, as with alternating tones (e.g., Rogers & Bregman, 1993), when the same two voices (e.g., male and female) alternate in a list (Greene, 1991), the transitional probabilities are likely to be greater between items spoken within each voice, that is, between non-adjacent items. Importantly, several studies have shown that the perception of order is relatively good for a succession of events assigned to the same stream but is notoriously poor for events traversing different streams (e.g., Bregman & Campbell, 1971; Warren, Obuzek, Farmer, & Warren, 1969). For example, if a sequence of high-frequency tones (A,B,C) is alternated with a sequence of low-frequency tones (1,2,3)—e.g., A,1,B,2,C,3—in a repeating loop, the majority of participants inadvertently report the order of the tones by frequency-range (or stream; e.g., A,B,C,1,2,3) rather than according to true temporal order (Bregman and Campbell, 1971; for similar effects using speech tokens, see Lackner & Goldstein, 1974). Such temporal order confusion—due this time to streaming by spatial location (e.g., Handel, 1989)—has also been used to explain the tendency to repeat back a sequence of simultaneous pairs of digits—one digit presented to each ear—by ear rather than by true temporal order (Broadbent and Gregory, 1961) and also of the difficulty of reproducing the true order of a sequence of ear-alternating items (Moray, 1960; see Bregman, 1990; ten Hoopen, 1996). All these cases illustrate how streaming can yield a perception of order pertaining to acoustically proximal events that takes precedence over that for temporally proximal ones. In relation to the TVE, then, we hypothesize that, perceptually, the order of adjacent items in a single-voice list would be relatively well preserved. In contrast, with a talker-variable list, obligatory by-voice streaming produces a mismapping between perceived order and the deliberate attempt to impose articulatory-based transitional probabilities between adjacent items in support of strict serial recall.

We test the perceptual-gestural mismapping account in the present study by modulating in various ways the likely strength of by-voice perceptual partitioning of an alternating-voice list (Experiments 1-3). Our rationale was that if the TVE is driven by the perceptual incoherence of temporally adjacent items in an alternating-voice list, any factor that promotes that incoherence—i.e., promotes by-voice streaming—should accentuate the TVE. We then go on in Experiment 4 to examine the contention that by-voice streaming has its disruptive impact by impairing the articulatory sequence-planning process.

Experiment 1

All experiments reported here follow Greene's (1991) methodology and examine the impact of alternating female-male voices on strict serial recall of a closed set of verbal items (digits or letters). Experiment 1 puts the perceptual-gestural mismapping account to the test by capitalizing on the fact that streaming is not an all-or-none phenomenon and that it, moreover, takes time to build up. We sought to accentuate the perceptual partitioning of alternating-voice lists by capitalizing on a form of *auditory stream biasing* whereby the likelihood of partitioning an alternating-stimulus sequence is strengthened if one of the stimuli in that sequence already forms part of a perceptually stable stream. More concretely, if a sequence of alternating high (H) and low (L) tones (HLHLHL) is preceded by a lead-in of either H or L tones (e.g., LLLLLHLHLHL), the partitioning of the eventual alternating sequence into two separate L and H streams occurs more readily. This is because the stable 'LLLLL…' stream perceptually captures the L tones in the following alternating sequence whilst the remaining, H, tones are more readily 'thrown out' as 'alien' tones to form their own distinct stream (Beavois & Meddis, 1997; Rogers & Bregman, 1993). We have demonstrated elsewhere that the same principles hold also for speech stimuli (Nicholls &

Jones, 2002; although see Remez, Rubin, Berns, Pardo, & Lang, 1994, for the view that for natural speech utterances there must also be domain-specific phonetic organization processes distinct from Gestalt-based ones). Thus, in this experiment, we sought to promote the perceptual partitioning of an alternating-voice list in a serial recall task by presenting a lead-in in the form of a countdown ("8, 7, 6…1") spoken in the same rhythm as the ensuing to-be-remembered (TBR) items and spoken in just one of the two voices making up the ensuing alternating voice list.[2]

We also sought to promote by-voice partitioning of the TBR items by exploiting the fact that such partitioning takes some time to build up; it becomes more emphatic as the evidence that there are indeed two distinct acoustic events accumulates. Thus, "all stimuli [in an alternating-tone sequence] begin by sounding temporally coherent and…the probability of stream formation increases steadily over time as a function of sequence duration" (Beavois & Meddis, 1997, p. 81; see also Anstis & Saida, 1985; Bregman, 1978, 1990; Carlyon, Cusack, Foxton, & Robertson, 2001). This build-up has a similar basis to the biasing effect described earlier: As an alternating sequence (LHLHLH…) continues, the ever-increasing stability of each individual stream—LLL and HHH—gradually increases the likelihood that each stream will capture each new L and H tone, respectively. Thus, we included a further condition that involved effectively increasing the duration of the alternating sequence by preceding an alternating-voice list with an alternating-voice lead-in. Again, this type of lead-in should promote the partitioning of alternating-voice items in the TBR list—and hence the magnitude of the TVE—because the by-voice partitioning process will have already begun during the lead-in and should be stronger therefore (compared to a no lead-in condition) by the time the TBR list starts. Table 1 provides a list of all six conditions contrasted in Experiment 1. Conditions 1 and 2 (i.e., those without a lead-in) are those required to show the standard TVE. The critical contrast thereafter will be between performance with an alternating-voice list (Alt) and that with an alternating-voice list preceded by a lead-in (Single-Alt or Alt-Alt): The perceptual-gestural mismapping account predicts poorer performance in the Single-Alt and Alt-Alt conditions compared to the Alt condition because the lead-in (of either type) should promote by-voice partitioning thereby exacerbating the conflict between order perception and the articulatory assembly of the items in canonical order. Whilst the present experiment serves primarily as a test of our perceptual-gestural mismapping account, it is unclear how an item-decay based approach (Goldinger et al., 1991; Martin et al., 1989) could accommodate any influence of lead-ins on the TVE.

Method

Participants

Twenty-two undergraduates from Cardiff University took part in return for course credits. Each participant reported normal hearing and normal or corrected-to-normal vision.

Apparatus and Materials

The TBR lists comprised 8 items taken without replacement from the digit-set 1-8. Each item was recorded digitally once in a female voice and once in a male voice (the items within each voice were spoken at an approximately even pitch), and sampled with a 16-bit resolution at a sampling rate of 44.1KHz using *Sound Forge 5* software (Sonic Inc., Madison, WI; 2000). The male and female voices clearly differed from one another on account of their distinct fundamental frequency and timbre. Each item's duration was edited to 250 ms using the same software. For each TBR list, the digits were presented in a pseudo-random order with care taken to ensure that there were no more than two occasions across a given TBR list on which there was an ascending or descending run of two or more digits (e.g., 2-3 or 7-6) and that there were no runs of 3 or more digits. This was also the case for non-adjacent items (e.g., those in positions 1 and 3) so that in alternating-voice lists there were no more than two 2-digit runs within a given voice in a given list. The TBR list (and lead-in when present) was presented at approximately 65-70 dB(A) over stereo headphones with an inter-stimulus interval (ISI; offset to onset) of 100ms giving an item presentation rate of 1 item/350ms. The stimuli were presented using the *SuperLab* software (Cedrus Corporation).

Table 1 provides a schematic representation of the 6 conditions assembled. In three of the conditions, the TBR lists were presented in a single voice (i.e., Single, Single-Single, and Alt-Single) whilst in another three, the TBR lists were spoken in an alternating female-male fashion. In conditions involving a single-voice lead-in, a countdown was presented either in the same voice as the ensuing single-voice list (Single-Single) or, for the Single-Alt condition, in the same voice as that conveying the second, fourth, sixth, and eighth items of the ensuing alternating-voice list. In the Alt-Single and Alt-Alt conditions, the lead-in was presented in alternating female-male voices.

Design

The design involved three repeated-measures factors: Lead-in (with three levels: no lead-in, single-voice lead-in, and alternating-voice lead-in), List-type (with two levels: single-voice and alternating-voice), and Serial position (eight levels). There were 84 experimental trials divided into two blocks: The ‘with lead-ins block’ comprised 56 experimental trials made up of 14 Alt-Alt trials, 14 Alt-Single trials, 14 Single-Single trials, and 14 Single-Alt trials. The block was preceded by 4 practice trials, one from each of the four conditions. The other block—the ‘without lead-ins block’—comprised 28 experimental trials made up of 14 single-voice TBR lists and 14 alternating-voice TBR lists preceded by 2 practice trials, one from each condition. In both blocks, the various trial-types were presented pseudo-randomly with the constraint that no condition was presented more than twice in succession. The order in which the two blocks were undertaken was counterbalanced across participants. Further counterbalancing measures incorporated into Experiments 1-3 are provided in the Appendix.

Procedure

Participants were tested in groups of up to four in a sound-attenuated room with each participant placed in a separate cubicle with its own PC and headphone. Participants were to recall the TBR digits in their correct order and to ignore the particular voice(s) conveying the digits. Participants were also told that for one block of trials the spoken list would be preceded by a spoken countdown. They were informed that 100 ms following the offset of the last TBR item of each list, the cue ‘RECALL’ would appear on the screen at which point they were to write down the items in the correct order on response sheets marked with 8 blank spaces for each trial. Participants had 15 s to recall the list and were instructed to do so in a strict left to right fashion such that they should start with position 1, then 2, and so on. They were instructed to guess if they were uncertain of any of the digits’ positions. A 500 ms tone was presented over the headphones 13 s into the 15 s recall-period to signal that the presentation of the first item of the next trial was imminent (in trials with a lead-in, the first item would be the first item of the countdown). The experiment lasted approximately 45 min.

Results and Discussion

For all experiments, the raw serial recall data were scored according to the strict serial recall criterion: To be recorded as correct an item had to be recalled in its original presentation position. Figure 1 shows the percentage of items correctly recalled across the eight serial positions in the six conditions. The pattern of results is clear-cut and can be unpacked initially into two distinct sets of curves: Replicating the basic TVE, performance in conditions involving an alternating TBR list (i.e., Alt, Single-Alt, and Alt-Alt; represented by the triangle symbols) was uniformly poorer than for conditions involving a single-voice list (i.e., Single, Alt-Single, and Single-Single; represented by the square symbols). More importantly, the TVE was markedly accentuated by the presence of a lead-in: Performance with alternating-voice lists was particularly poor when those lists were preceded by either an alternating- or single-voice lead-in (Single-Alt and Alt-Alt). The pattern across conditions thus conforms to that predicted by the perceptual-gestural mismapping account.

A 2 (List-type) by 3 (Lead-in) by 8 (Serial Position) repeated-measures ANOVA revealed a main effect of Serial position $F(7, 147) = 55.27$, $MSE = .06$, $p < .001$, a main effect of List-type, $F(1, 21) = 69.83$, $MSE = .07$, $p < .001$, a main effect of Lead-in, $F(2, 42) = 15.87$, $MSE = .01$, $p < .001$, and, most importantly, a significant List-type by Lead-in interaction, $F(2, 42) = 12.19$, $MSE = .02$, $p < .001$, reflecting the fact that the TVE was larger when an alternating list was preceded by a lead-in (of either type). The only other significant effect was an interaction between List-type and Serial position, $F(7, 147) = 18.37$, $MSE = .01$, $p < .001$, possibly reflecting ceiling effects at primacy and recency serving to obscure differences according to list-type. Follow-up simple effects analyses confirmed that all alternating-voice TBR list conditions produced poorer performance than any of the conditions with a single-voice TBR list (all comparisons $p < .005$). More importantly, they also showed that performance was poorer in both the Single-Alt and Alt-Alt conditions than in the Alt condition (both $p < .001$). A further diagnostically important feature of the data is that the presence of a lead-in per se had no effect on serial recall: There was no significant difference between either the Single-Single or Alt-Single and the Single condition (both comparisons, $p > .05$).[3]

The results of Experiment 1 confirm a prediction of a perceptual-gestural mismapping account of the TVE: The lead-in (of either type) promoted the perceptual incoherence of adjacent TBR items—and at the same time promoted the perceptual coherence of non-adjacent items—thereby accentuating the mismapping between the order suggested by streaming and the action requirements of the serial recall task. Accordingly, the TVE was significantly larger when an alternating-voice list was preceded by a lead-in. A possible alternative interpretation of these data, however, is that the particularly poor performance in the Alt-Alt and Single-Alt conditions was due to the digit lead-in producing proactive interference (PI)—the difficulty of recalling items due to their post-categorical similarity to previously encountered items (e.g., Bunting, 2006; Underwood, 1957)—or producing some general attentional distraction effect (S. Lewandowsky, personal communication). This seems unlikely, however, because the presence of a lead-in had no effect when it preceded a single-voice list: There was no difference between the Alt-Single or Single-Single conditions and the Single condition. Nonetheless, it is possible to counterargue that a PI/distraction effect from a lead-in might only emerge if recall is already made difficult by having to recall an alternating-voice list. In Experiment 2, we examined these alternative interpretations of the data of Experiment 1.

Experiment 2

We sought to rule out the PI/distraction accounts and sought further evidence for the perceptual-gestural mismapping account of the results of Experiment 1 by two convergent means. First, the lead-in on this occasion consisted of letter-names rather than digits. This should minimize the likelihood of PI based on the postcategorical similarity of the items used in the lead-in and the TBR digit list (e.g., Bunting, 2006). If such a lead-in still accentuates the TVE, a PI-based explanation seems unlikely. Second, we manipulated the voice conveying this (single-voice) letter lead-in. According to the perceptual-gestural mismapping account, the TVE was augmented by the presence of a lead-in because the voice(s) conveying that lead-in perceptually captured the same-voice items in the alternating TBR list through their similar acoustic (not post-categorical) attributes. Indeed, it has been shown that for auditory stream biasing to occur, the frequency of the tone(s) in the lead-in must be very similar in frequency to those in an alternating-tone test sequence (Anstis & Saida, 1985). Thus, in Experiment 2, we contrasted a condition involving a single-voice lead-in spoken in one of the voices conveying the alternating TBR list (as in the Single-Alt condition of Experiment 1) and one involving a single-voice lead-in spoken in a 'third voice', a voice different from either of the voices conveying the alternating voice TBR list. Based on the perceptual-gestural mismapping account, this different voice lead-in—due to its relative acoustic dissimilarity to the list voices particularly in terms of frequency/timbre—should be less effective in promoting by-voice partitioning of the alternating-voice TBR list and hence in augmenting the TVE. Thus, in relation to Table 2—which shows all 6 trial-types contrasted—whereas a same-voice lead-in, as in Experiment 1, should further impair recall of an alternating voice list as compared with the Alt condition, such accentuation should be diminished with a different-voice lead-in. On the PI or distraction accounts, there is little reason to suppose that the particular voice conveying the lead-in should modulate its impact.

Method

Participants

Twenty-six undergraduates from Cardiff University each reporting normal hearing and normal or corrected-to-normal vision took part in return for course credits.

Apparatus and Materials

The TBR lists were constructed and presented in the same fashion as Experiment 1. However, in addition to recording the digit-sets 1-8, the letters that now served as the lead-in—*r*, *s*, *t*, *u*, *v*, *x*, *y*, and *z*—were digitally recorded three times, once in the same female voice and once in the same male voice used for the digit-sets (the same voices as used in Experiment 1) and once in a different (male) voice chosen on the basis that it was clearly distinguishable perceptually from both the other two (male and female) voices. Table 2 shows the full set of 6 conditions: All TBR lists were presented in one of the original voices or in both original voices presented in an alternating fashion. In the conditions involving a lead-in, a letter countdown (*r* through *z* excluding *w*) was presented either in the same voice as that conveying an ensuing single voice-list (Same-Voice/Single) or in the 'different voice' (Different-Voice/Single). For the Same-Voice/Alt condition, the lead-in was presented in the same voice as that conveying the second, fourth, sixth, and eighth items of the ensuing alternating-voice list. Finally, in the Different-Voice/Alt condition, the lead-in was presented in the 'different voice' and hence was distinct from both voices conveying the ensuing alternating-voice list.

Design and Procedure

The design and procedure were the same as Experiment 1 except: The three levels of the Lead-in factor were no lead-in, same-voice lead-in, and different-voice lead-in. The 56 trials in the 'with lead-ins block' were composed of 14 Same-Voice/Alt trials, 14 Different Voice/Alt trials, 14 Same-Voice/Single trials, and 14 Different-Voice/Single trials and was preceded by 8 practice trials, two from each of the four conditions. The 'without lead-ins block' comprised 28 trials made up of 14 single-voice lists and 14 alternating-voice lists preceded by 4 practice trials, 2 from each condition. See also Appendix for further counterbalancing measures. The procedure was the same as for Experiment 1.

Results and Discussion

It is clear from Figure 2 that performance was markedly impaired in all talker-variable conditions compared to the Single-Voice conditions (both with and without a lead-in). Moreover, whilst performance was worse in the Same-Voice/Alt condition compared with the Alt condition, this was not the case for the Different-Voice/Alt condition. Thus, in line with the perceptual-gestural account, the TVE was accentuated by a lead-in only when that lead-in was in the same voice as one of the voices conveying the ensuing alternating-voice sequence (Same-Voice/Alt).

A 2 (List-type) by 3 (Lead-in) by 8 (Serial Position) repeated-measures ANOVA revealed a main effect of Serial position, $F(7, 189) = 210.27$, $MSE = .038$, $p < .001$, and of List-type, $F(1, 27) = 111.42$, $MSE = .040$, $p < .001$, but no main effect of Lead-in, $F(2, 54) = 2.82$, $MSE = .024$, $p = .068$. Importantly, the List-type by Lead-in interaction was significant, $F(2, 54) = 3.51$, $MSE = .018$, $p < .05$. Simple effects analyses showed that whereas there was no significant difference between performance in the Different-Voice/Alt condition and the Alt condition ($p > .05$), performance was indeed poorer in the Same-Voice-Alt condition compared to the Alt condition ($p < .001$). Again, there was no difference between any of the conditions with a single voice list (Single, Same-voice/Single, Diff-voice/Single) showing that the lead-in in-and-of-itself did not affect performance. Both Lead-in and List-type interacted with Serial Position ($p < .05$); we refrain from attempting to ascribe any theoretical significance to these interactions however.

The results of Experiment 2 indicate that the lead-in's power to augment the TVE in Experiment 1 was not dependent on the fact that it comprised items that were post-categorically similar (in fact identical) to those to be recalled: A letter lead-in conveyed in the same voice as one of the voices in the ensuing TBR digit list also accentuates the TVE. Moreover, a letter lead-in has no effect if the voice in which it is conveyed is relatively acoustically dissimilar to both voices in the TBR list. This is entirely as would be expected based on previous evidence that auditory stream biasing is highly sensitive to the similarity of the frequency of lead-in and test tones (Anstis & Saida, 1985). Together, these two aspects of the results of Experiment 2 provide strong support for the view that a (same voice) lead-in has its impact not by being post-categorically similar to the TBR list but by capturing the same voice items in the following TBR list through their similar acoustic characteristics. Neither a PI- (e.g., Bunting, 2006) nor a general attentional distraction-based explanation can readily account for this pattern of results.

The results of Experiments 1 and 2 are also not readily explained by previous item-decay accounts of the TVE (Goldinger et al., 1991; Martin et al., 1989): It is far from clear on this approach why the presence of a lead-in generally should accentuate the TVE and, more particularly, why that lead-in has to be acoustically similar to one of the voices in the alternating list to exert this effect. Indeed, one might have expected pre-exposure to the indexical properties (e.g., timbre) of the same voice as that conveying some of the ensuing TBR list (Single-Alt condition in Experiment 1 and same-voice lead-in/Alt in Experiment 2) or both of the voices conveying that list (Alt-Alt condition of Experiment 1)—and particularly being pre-exposed to the temporal pattern of voice-changes (Alt-Alt condition)—to facilitate voice normalization or incorporation processes. Such facilitation should in turn have allowed greater opportunity to refresh decay-prone item-traces via rehearsal and hence reduce, rather than accentuate, the TVE.

Experiment 3

In Experiment 3 we take a different approach to testing the perceptual-gestural mismapping account. The account yields the perhaps counterintuitive prediction that increasing the number of different voices in a TBR list should diminish not augment the TVE. This follows on the grounds that the smaller the number of different voices, the more likely and often it is that non-adjacent items will be more acoustically similar to one another and hence the greater the likelihood of their forming a coherent perceptual stream. Before explaining further, we describe the three conditions contrasted in Experiment 3 in more detail. The TBR list in the ‘four-voice condition’ comprised the female samples used in Experiment 1 (and 2) and another three ‘voices’ generated by pitch-shifting those female (F) samples down by 3 semi-tones (hereafter: F-), up by 3 semi-tones (F+), and up by 6 semi-tones (F++). The eight-digit list was conveyed in the following pattern of voices: F F+ F++ F+ F F- F F+ (or its mirror image: F+ F F- F F+ F++ F+ F). In the single-voice condition, the TBR list was conveyed in either one of the four voices whilst in the alternating-voice condition, the TBR list involved an alternation between F and F+ (or F+ and F).[4] For each list-type, given that we have shown that the TVE is more robust when lead-ins are used, each TBR list was preceded by a lead-in (a digit countdown as in Experiment 1) in which the voice or pattern of voices conformed to that characterizing the ensuing TBR list.

On the basis of the perceptual-gestural mismapping account, we expect poorer recall in both the alternating- and four-voice conditions than in the single-voice condition due to the far greater perceptual incoherence of successive items in the two talker-variable conditions. However, the potentially more telling prediction is that the alternating-voice list should produce the poorest performance. This is because it is well known that the likelihood of temporally non-adjacent items perceptually "capturing" one another into the same stream is a function of both their acoustic similarity, the number of times those similar items are encountered, and the temporal proximity of the acoustically similar events (Bregman, 1990). Thus, given that in the alternating-voice list, there are six instances in which immediately non-adjacent items are in the same voice whereas there are only two such instances in the four-voice list, the propensity for non-adjacent items to perceptually capture one another to form a coherent stream—and hence the degree of perceptual-gestural mismapping—is greater in the alternating- compared to the four-voice list.

Method

Participants

Forty undergraduates from Cardiff University took part in return for course credits. Each participant reported normal hearing and normal or corrected-to-normal vision.

Apparatus, Materials, Design & Procedure

These aspects of the method were the same as Experiment 1 except: Three new sets of voice samples were generated by pitch-shifting the original female-spoken items down by 3 semi-tones, up by 3 semi-tones, and up by 6 semi-tones (without altering each item's duration) using the 'pitch-shift' function in the *Soundforge 7* software. The two repeated-measures factors were List-type (three levels: Single-voice, Alternating-voice, and Four-voice) and Serial position (eight levels). There was one block with 84 trials: 28 Single-voice trials, 28 Alternating-voice trials in which the F voice alternated with the F+ voice, and 28 Four-voice trials (forming either the pattern F F+ F++ F+ F F- F F+ or F+ F F- F F+ F++ F+ F). Each TBR list was preceded by a lead-in that conformed to the same voice-format as that TBR list. One of each of the 8 trial-types was presented as practice trials.

Results and Discussion

It is evident from Figure 3 that performance was markedly impaired in both talker-variable conditions compared to the single-voice condition. More importantly, in line with the perceptual-gestural mismapping account, performance was significantly poorer in the alternating-voice condition than in the four-voice condition. There was main effect of Serial Position, $F(7, 273) = 120.72$, $MSE = .03$, $p < .001$, and of List-type, $F(2, 78) = 49.87$, $MSE = .02$, $p < .001$, as well as an interaction between List-type and Serial Position, $F(14, 546) = 5.57$, $MSE = .006$, $p < .001$. Planned repeated contrasts showed that performance in the Four-voice condition was significantly poorer than in the Single-voice condition, $F(1, 39) = 63.27$, $MSE = .05$, $p < .001$, and, most importantly, that performance in the Alternating-voice condition was slightly (Cohen's $d = .12$) but significantly poorer than in the Four-voice condition, $F(1, 39) = 5.9$, $MSE = .02$, $p = .02$.

The results of Experiment 3 provide further support for the perceptual-gestural mismapping account: Whilst the perception of true temporal order would be impaired by the lack of perceptual coherence in both talker-variable conditions, the non-adjacent items in the alternating-voice list condition would be more strongly grouped on account of their greater acoustic similarity and the greater number of repetitions of those more similar events (voices) compared to the case in the four-voice condition (see, e.g., Bregman & Rudnicky, 1975). Again, it is unclear how the item-decay accounts could accommodate this result: there is no apparent reason why voice normalization (Martin et al., 1989) or incorporation processes (Goldinger et al., 1991) would be slower—hence leading to more item decay—with two different voices in the list as compared with four.

The present results also militate against a suggestion made by Greene (1991) that the TVE may result from participants adopting a deliberate, but counterproductive, strategy of grouping (or rehearsing) the TBR items by voice when presented with an alternating-voice list. This suggestion draws on evidence from free recall paradigms in which participants tend to cluster at output items that share source characteristics such as language (Tulving and Colotla, 1970), modality (Murdock & Walker, 1969), or type-face (Hintzman, Block, & Inskeep, 1972; for a recent review, see Polyn, Norman, & Kahana, 2009). On this deliberate-grouping view, the TVE reflects the cost of having to re-organize the items into their true serial order at output having deliberately organized the items by voice during presentation. This differs from our account in which the non-strategic, obligatory, grouping of items by voice conflicts with the deliberate attempt to assemble the items according to their true temporal order; at no point is there a deliberate attempt to organize the items by voice. Given that it seems unlikely that participants would choose to adopt a deliberate strategy of grouping the items by voice in the four-voice condition (i.e., into four groups comprising two items each), the deliberate grouping account has difficulty in explaining the marked impairment in this condition compared to the single-voice condition. Even if a 4-group strategy (in the four-voice condition) is feasible, it is reasonable to expect that it would be more burdensome to re-organize the items into canonical order at output than with a 2-group strategy (in the two-voice condition), an expectation at odds with the present results.

The findings of Experiments 1-3 converge to support the view that one key component of the TVE is an obligatory auditory perceptual organization of the items that conflicts with the true temporal order of the items. The second key aspect of our account is that the locus of the impairment is, ultimately, the gestural sequence-planning process: The process of assembling the items into a gestural analogue is fed impoverished or incompatible information regarding the order of the items by the obligatory auditory perceptual organization process. In Experiment 4, we turn to focus on this gestural-planning component of the account.

Experiment 4

The analytical device of examining whether two or more variables known to independently affect serial recall combine to produce an additive or non-additive effect has played a key role in the development of short-term memory theory (e.g., Baddeley, Lewis, & Vallar, 1984; Larsen & Baddeley, 2003; Longoni, Richardson, & Aiello, 1993). In Experiment 4, we examine for the first time the possible interplay of talker variability and the phonological similarity effect, a benchmark finding in serial recall whereby phonologically similar items (*b, d, v*….) are more difficult to serially recall than phonologically dissimilar items (*f, r, q*…) (Baddeley, 1966). Within the phonological loop model—the most successful instantiation of the decay-rehearsal based approach—the phonological similarity effect is the chief empirical signature of the passive phonological store (e.g., Baddeley, 2007). However, recent evidence suggests instead that the locus of the effect is the speech-planning process, not a separate passive phonological store: the effect disappears when rehearsal is blocked by articulatory suppression for both visual and auditory lists (Jones et al., 2004, 2006). The residual phonological similarity effect found for auditory lists under suppression—previously taken as evidence against identifying the phonological similarity effect with the rehearsal process (e.g., Baddeley et al., 1984)—is better explained by recourse to acoustic-based perceptual organization processes rather than a phonological store (Jones et al., 2004, 2007; but see Baddeley & Larsen, 2007).

The different accounts of the phonological similarity effect held by the decay-rehearsal model (e.g., Baddeley, 2002) and the perceptual-gestural framework (e.g., Jones et al., 2004) provides a further means of adjudicating between the item-decay accounts and the perceptual-gestural mismapping account of the TVE. A key finding cited in support of the fractionation of the phonological loop into a passive phonological store and an articulatory rehearsal process is that the phonological similarity effect combines additively with the word-length effect: "…articulatory duration and the phonemic confusability of items to be remembered exert additive and independent effects upon performance in immediate serial recall, and hence [this shows] that they reflect distinct components of the working-memory system" (Longoni, Richardson, & Aiello, 1993, p. 14; see also Baddeley et al., 1984). That is, whereas the phonological similarity effect reflects a confusion-during-retrieval between similar item-traces in the passive store, the word-length effect reflects a race between articulatory rehearsal and item decay. Thus, given that the item-decay accounts of the TVE appeal to the same mechanism as the decay-rehearsal model offers for the word-length effect, they also predict that phonological similarity and talker variability should exert independent (i.e., additive) effects. In contrast, the perceptual-gestural view posits that both phenomena affect the speech-planning process and hence predicts that the two effects will interact (i.e., will be non-additive): Phonological similarity will have less impact because the speech-planning process is already impaired by talker variability. In Experiment 4, therefore, participants serially recalled 6 letters which could either be phonologically similar or dissimilar and these two types of list were presented either in a single voice or in alternating voices.

Method

Participants

Twenty undergraduates from Cardiff University took part in return for course credits. Each participant reported normal hearing and normal or corrected-to-normal vision.

Apparatus and Materials

Four list-types were generated. Each list comprised 6 letters that were either phonologically dissimilar (*k, q, h, y, r, m*) or similar (*p, d, t, v, b, g*) and these two list-types could be presented either in a single voice or in alternating voices. Female and male-spoken versions of the items were recorded, edited, and presented in the same manner as in Experiment 1.

Design and Procedure

There were 3 repeated-measures factors: Serial position; Phonological similarity (similar vs dissimilar); and Voices (single vs alternating). There were 2 blocks of trials: 20 phonologically dissimilar lists and 20 phonologically similar lists. Block-order was counterbalanced across participants. Within each block, 10 lists were presented in a single voice (5 female; 5 male) and 10 in alternating female-male voices (5 starting with a female-spoken item; 5 starting with a male-spoken item). No trial-type was presented more than twice in succession within a block. (Note that none of the lists were preceded by a lead-in in this experiment). The procedure was the same as Experiment 1 except that before each block, the 6 consonants to be used in that block were presented in a circle on the screen for 2 minutes to allow the participant to familiarize themselves with the closed item-set (cf. Henson, Norris, Page, & Baddeley, 1996). There were 4 practice trials (one from each condition) before the experimental trials. The experiment took approximately 25 min.

Results and Discussion

It is apparent from Figure 4 that the phonological similarity effect is diminished in the context of alternating-voice lists. Thus, the data exhibit a non-additivity of phonological similarity and talker variability in line with the perceptual-gestural mismapping account but at variance with an item-decay approach. A repeated-measures ANOVA showed a main effect of Serial position, $F(5, 95) = 83.39$, $MSE = 0.03$, $p < .001$, a main effect of Phonological similarity, $F(1, 19) = 68.51$, $MSE = 0.11$, $p < .001$, a main effect of Voice, $F(1, 19) = 16.15$, $MSE = 0.04$, $p < .001$, and, most importantly, a significant interaction between Phonological similarity and Voice, $F(1, 19) = 13.16$, $MSE = 0.03$, $p < .01$. The interactions between Serial position and each of the other two variables were also significant, which were subsumed within a significant three-way interaction between all three variables, $F(5, 95) = 3.72$, $MSE = 0.01$, $p < .01$. Again, we will not ascribe any theoretical significance to these interactions at this point given that neither of the accounts being contrasted here make clear predictions with regard to serial position.

Experiment 4 showed that talker variability and phonological similarity interact (i.e., are non-additive) consistent with the view that the two effects share a functional locus. The results are therefore at odds with the predictions of item-decay accounts based on the decay-rehearsal model, at least as exemplified by the phonological loop model (Baddeley, 1986): If the phonological similarity effect is the empirical signature of a passive phonological store, there is no reason to expect it to interact with the TVE which, from this perspective, has been attributed to the articulatory control process (as with the word-length effect; e.g., Martin et al., 1989). In contrast, the non-additivity of talker variability and phonological similarity effects is consistent with the perceptual-gestural account: If, as we suppose, talker variability impairs the sequence output-planning process, the expression of another effect also associated with that process (Jones et al., 2004, 2006; see also Ellis, 1980; Page, Madge, Cumming, & Norris, 2007) would be expected to be diminished with talker-variable lists. Further evidence that the by-voice perceptual organization of an alternating-voice list—as demonstrated in Experiments 1-3—has its impact by conflicting with the assembly or/and rehearsal of a sequence-output plan comes from studies manipulating the nature of the memory task: Tasks that call for the retention of order—and hence promote the use of an articulatory sequence-planning strategy—are far more susceptible to a talker variability effect than those that only require the retention of item information (Hughes, Marsh, & Jones, 2009)[5].

General Discussion

The impact of the present series may be summarized as follows: Experiments 1 and 2 showed that promoting the perceptual partitioning of talker-variable lists by voice by presenting a (same, but not different, voice) single- or alternating-voice lead-in before an alternating-voice list accentuates the TVE in serial recall. Experiment 3 showed that the recall of a two-voice (i.e., alternating) list was poorer than that of a four-voice list, again implicating the action of perceptual streaming processes. Experiment 4 showed that the phonological similarity effect is far weaker in the context of talker-variable lists in line with the perceptual-gestural view that both effects are located in the gestural-sequencing process rather than reflecting distinct components of a bespoke memory system.

Implications for the Decay-Rehearsal Approach to Serial Short-Term Memory

Previous accounts of the TVE have been set within the framework of the classical decay-rehearsal approach to verbal short-term memory (e.g., Baddeley, 1986). The process of normalizing out or incorporating indexical attributes (e.g., accent, emotional state) of spoken words delays the encoding or/and rehearsal of each item thereby promoting item-decay from a phonological short-term store (Goldinger et al., 1991; Martin et al., 1989). The present results are not easily accommodated within this approach. One potential objection, however, is that the item-decay accounts were developed in the context of a 'non-pure' serial recall task in which, unlike the present experiments, a new set of items were presented on each trial (e.g., Goldinger et al., 1991; Martin et al., 1989). However, there is little reason why the accounts should not apply in principle to the TVE in pure serial recall: On the decay-rehearsal model, the same item-decay approach is used to explain the word-length effect found in pure serial recall (e.g., Baddeley et al., 1975). Nevertheless, the present results do not rule out the possibility that item-decay plays a role in the TVE in 'non-pure' serial recall where there is a large burden on item as well as order retention, unlike the task used in the present series. An alternative possibility consistent with the perceptual-gestural mismapping account is that although a non-pure serial recall task places a large burden on item memory, talker variability nevertheless impairs only that component of the task that taps the efficacy of a sequence output-planning process. In line with this analysis, in non-pure serial recall (e.g., Goldinger et al., 1991) and also under free recall instructions (Watkins & Watkins, 1980), the TVE is confined to that part of the list—the early part—that has been shown (even in free recall) to be supported by serial rehearsal (Bhatarah, Ward & Tan, 2008; Beaman & Jones, 1998; Kahana, 1996). Clearly, further research that contrasts the TVEs found in pure and non-pure serial recall will be needed to determine more conclusively whether they share a common mechanism.

It is also important to recognize that TVEs have been observed in many other tasks that seem not to require order retention at all such as long-term word recognition (Mullennix et al., 1989), vowel perception (Verbrugge, Strange, Shankweiler, & Edman, 1976), and the speeded classification of words (Mullennix & Pisoni, 1990). Such effects suggest strongly that the indexical and acoustic features of spoken words are not normalized out during perception but incorporated as part of the encoding episode (e.g., Goldinger 1996). The present findings lend further support to this general view; if list-items are represented in an abstract-phonological form (e.g., Baddeley, 2007), it is difficult to account for the streaming-by-voice effects observed here. However, our findings suggest that this voice incorporation process does not cause the TVE in (at least pure) serial recall by promoting item-decay but may indeed do so through its influence on supra-item perceptual organization.

Whilst current item-decay based accounts do not fare well in relation to the present findings, there may be other means by which the decay-rehearsal model within which they are conceptually embedded (e.g., the phonological loop model; Baddeley, 1986, 2007) could potentially account for the TVE. One possibility might be to appeal to computational models of how the phonological loop represents serial order, a consideration that has typically been seen as complementary, but secondary, to the core item-based architecture of the underlying functional-level theory (Baddeley, 2007). Thus, talker variability might be seen as disrupting the mechanism supporting serial order in the phonological loop (e.g., a primacy gradient, Page & Norris, 1998; an oscillator-based timing signal; Burgess & Hitch, 1999). However, the interaction between phonological similarity and talker variability in Experiment 4 still seems to pose difficulties for an approach that appeals to an order-based account of the TVE but nevertheless adheres to the concept of a passive phonological store. In computational models of the phonological loop, the phonological content (and hence phonological similarity) of the items only has an impact at a second item-retrieval stage that is independent of whatever mechanism is responsible for representing order (e.g., Burgess & Hitch, 1999; Page & Norris, 1998). Thus, given that the phonological similarity effect is identified with the second item-retrieval stage on this account, its magnitude should not be influenced by a factor—talker variability—presumed to affect a separate order-storing stage, contrary to our data. Moreover, such models rarely consider pre-phonological perceptual organization processes in explanations of serial recall performance (but see Page & Norris, 1998). Yet the present results add to a growing body of evidence suggesting that perceptual organization coupled with output-planning processes may alone account for verbal serial recall phenomena without the encumbrance of a bespoke phonological store (e.g., Jones et al., 2004, 2006; Woodward et al., 2008; but see Baddeley & Larsen, 2007; Page et al., 2007).

Implications for Item-Interference Accounts of Serial Short-Term Memory

We have focused on the decay-rehearsal model (e.g., Baddeley, 1986, 2007) of short-term memory as the main theoretical counterpoint to our perceptual-gestural account (but see Experiment 2). However, the TVE seems problematic also for accounts of serial short-term memory in which the relative distinctiveness of items—and hence their relative immunity from interference—serves as the primary explanatory mechanism (e.g., Brown, Neath, & Chater 2007; Lewandowsky et al., 2008; Nairne, 1990). Although these accounts, like the perceptual-gestural view, eschew a dedicated short-term store, they still invoke a specifically mnemonic process, namely, interference between similar memory traces. For example, according to the feature model (Nairne, 1990; Neath, 2000), serial recall performance is assumed to bear a simple positive relationship to the distinctiveness—and hence immunity from being overwritten—of the items in a TBR list in terms of both their modality-dependent features (e.g., pitch) and their modality-independent features (those that do not vary with modality of presentation). As pointed out previously by Greene (1991), it follows quite straightforwardly that each item in a talker-variable list should be less prone to being overwritten by its successor because there is less overlap in terms of modality-dependent features (e.g., pitch, timbre) than with a single-voice list. Such models cannot, therefore, readily explain why the greater distinctiveness of items in a talker-variable list leads to much poorer, not better, recall (Greene, 1991). Moreover, even if such models could be modified to accommodate the basic TVE, it is not immediately apparent how they might explain the accentuating effect of a lead-in (Experiments 1 and 2), or the non-monotonic relationship between the level of impairment and the number of different voices (Experiment 3).

Perceptual-Gestural Account: Towards an Embodied View of Short-term Memory

The present results are in line with previous studies demonstrating the powerful effects of auditory perceptual organization on order processing generally (e.g., Bregman & Campbell, 1971; Creel, Newport, & Aslin, 2004; Warren et al., 1969) and on serial recall in particular (e.g., Nicholls & Jones, 2002; Macken et al., 2003). In the present case, the obligatory by-voice perceptual organization of talker-variable lists gives rise to a perception of order that conflicts with the process of assembling the items into an articulatory plan designed to increase the transitional probabilities between adjacent items. Our appeal to general-purpose processes involved in perception and action that are co-opted to meet the demands of a short-memory task (for similar views, see Glenberg, 1997; Reisberg, Rappaport, & O'Shaughnessy, 1984; Wilson & Fox, 2007) resonates with a current shift in cognitive science towards embodying cognition (e.g., Clark, 2006). This shift has emerged as a reaction to the received view of cognition as the action of static, central, and context-free processing and storage structures/resources that are divorced from the so-called "peripheral" processes of perception and action (e.g., Clark, 2006; Hurley, 2001). Instead, an embodied analysis focuses on the dynamic processes involved in goal-directed and coherent engagement with the environment given the constraints and capacities of the organism's sensori-motor apparatus.

Couched within the embodied cognition approach, one way of fleshing out the gestural component of our account is to suppose that sequence output-planning (or 'rehearsal') reflects the operation of a motor-action emulator. Recent work on motor control suggests that in order for motor-action to be executed in a fluent manner, a 'forward model' of the action—consisting of both the instructions to the effectors and, importantly, the sensory sequelae of the action—is generated so that the imminent action can be compared with the intended action (e.g., Grush, 2004; Shubotz, 2007). An important feature of these models that aligns with the phenomenology of subvocal rehearsal during serial recall is that they can be run without being implemented, that is, they may be run in emulation mode without necessarily resulting in any overt action (e.g., Jordan & Rumelhart, 1992). Thus, we contend that articulatory rehearsal during verbal serial recall is not in the service of refreshing decay-prone items in a labile phonological store (see also Reisberg et al., 1984); rather, the necessarily sequential nature of the emulation of the movements of the vocal tract endows the individual with an ideal medium for taking a series of largely sequentially-unrelated verbal items and placing them onto a common carrier, that is, a single, relatively more sequentially-coherent, motor-plan. The TVE on this view reflects the impairment of an auditory-motor mapping process (cf. Buchsbaum & D'Esposito, 2008) whereby the process of populating the necessarily abstract motor-output emulator system with specific content is ill-informed by auditory-perceptual organization. Another hypothesis that flows from our account, therefore, is that conditions under which perceptual organization yields information that corresponds particularly well to the way in which items need to be assembled into the motor-plan should facilitate rather than impair performance. For example, one might contrast two versions of a pure serial recall task, one in which the recall of a sub-set of adjacent items is pre-cued ("recall items 1-5 [or 6-10] of the following 10-item list") and one in which a sub-set of non-adjacent items is pre-cued ("recall items 2, 4, 6, 8, and 10 [or 1, 3, 5, 7, and 9], cf. Penney & Butt, 1986). The perceptual-gestural account predicts that alternating-voice lists would impair the first version of the task but facilitate the second. The account might also provide a way of reconstruing the impairment of serial recall when successive items or pairs of items are presented to different sensory modalities (Penney & Butt, 1986) or to different ears

(Moray, 1960; Treisman, 1971). For example, ear alternation effects, like the TVE, have typically been attributed to a delay in the processing of each item: The "attention shifting time would reduce the time available for perception and storage" (Treisman, 1971, p. 164). However, as suggested previously by ten Hoopen (1996), these effects may instead be driven by streaming-by-location, again giving rise to a perceptual-gestural mismapping. The device of introducing (on this occasion ear-alternating) lead-ins may provide one means of testing this possibility.

Encouragingly, conclusions from several research programmes are now converging on an embodied conceptualization of short-term memory (Acheson & MacDonald, 2009; Buchsbaum & D'Esposito, 2008; Postle, 2006; Jones et al., 2004; 2006; Wilson & Fox, 2007). For example, Wilson and Fox (2007) recently found that serial recall of novel sequences of hand gestures exhibits several of the effects that are, putatively, hallmarks of a specifically verbal short-term memory system, namely, the "phonological" similarity effect, the "articulatory" suppression effect and the "word" length effect. The authors concluded that "(r)ather than involving hard-wired and dedicated components, working memory may instead consist of the strategic recruitment of cognitive resources, determined on the fly by the immediate demands of the task" (p. 473). We suggest that the present findings illustrate further the fruitfulness of this shift in research focus away from delineating the properties of bespoke stores and mechanisms and toward examining instead how perceptual and action-planning processes support and constrain the retention and reproduction of serial order over the short-term.

References

Abercombie, D. (1967). Introduction. In D. Abercrombie (Ed.), Elements of general phonetics (pp. 1-17). Chicago: Aldine Publishing Company.

Acheson, D. J., & MacDonald, M. C. (2009). 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,* 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. (2007). *Working memory, thought and action*. Oxford: 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: Academic Press.

Baddeley, A. D., & Larsen, J. D. (2007). The phonological loop unmasked? A comment on the evidence for a “perceptual-gestural” alternative. *Quarterly Journal of Experimental Psychology, 60,* 497-504.

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology, 36,* 233–252.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575-589.

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 and Cognition, 23*, 459-471.

Beauvois, M. W., & Meddis, R. (1997). Time decay of auditory stream biasing. *Perception & Psychophysics, 59,* 81-86.

Bhatarah, P., Ward, G., & Tan, L. (2008). Examining the relationship between free recall and immediate serial recall: The serial nature of recall and the effect of test expectancy. *Memory and Cognition, 36,* 20-34.

Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 380-387.

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

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

Bregman, A. S., & Rudnicky, A. I. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception and Performance, 1,* 263-267.

Broadbent, D. E., & Gregory, M. (1961). On the recall of stimuli presented alternately to two sense-organs. *Quarterly Journal of Experimental Psychology, 13,* 103-109.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114,* 539-576.

Bunting, M. (2006). Proactive interference and item similarity in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 183-196.

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

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

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 115–127.

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences, 10*, 370-374.

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, England: Oxford University Press.

Creel, S. C., Newport, E. L., & Aslin, R. (2004). Distant melodies: Statistical learning of nonadjacent dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1119-1130.

Deouell, L. Y., Deutsch, D., Scabini, D., Soroker, N., & Knight, R. T. (2008). No disillusions in auditory extinction: Perceiving a melody comprised of unperceived notes. *Frontiers In Neuroscience, 1:15.*

Ellis, A. W. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior, 19,* 624–634.

Farrell, S., & Lewandowsky, S. (2002). An endogenous model of ordering in serial recall.

Psychonomic Bulletin & Review, 9, 59-79.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20*, 1-55.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance, 22,* 1166-1183.

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effect s in recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 152-162.

Greene, R. L. (1991). Serial recall of two-voice lists: Implications for theories of auditory recency and suffix effects. *Memory & Cognition, 19*, 72-78.

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*, *27*, 377-442.

Handel, S. (1989). *Listening.* Cambridge, MA: MIT Press.

Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology, 49A,* 80–115.

Hintzman, D. L., Block, R. A., & Inskeep, N. R. (1972). Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior, 11,* 741–749.

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 and Performance, 31,* 316-327.

Hughes, R. W., Marsh, J. E., & Jones, D. M. (2009). *One at a time, please! The impact of talker variability on short-term memory.* Manuscript submitted for publication.

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,* 1050-61.

Hurley, S. (2001). Perception and action: Alternative views. *Synthese, 129,* 3-40.

Jefferies, E., Lambon Ralph, M. A., & Baddeley, A. D. (2004). Automatic and controlled processing in sentence recall: The role of long-term and working memory. *Journal of Memory and Language*, *51*, 623-643.

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

Jordan, M. I., & Rumelhart, D. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science, 16,* 307-354.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory and Cognition*, *24*, 103-109.

Lackner, J. R., & Goldstein, L. M. (1974). Primary auditory stream segregation of repeated consonant-vowel sequences. *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, 56A,* 1249–1268.

Lewandowsky, S., Geiger, S. M., & Oberauer, K. (2008). Interference-based forgetting in verbal short-term memory. *Journal of Memory and Language, 59,* 200-222.

Longoni, A. M., Richardson, J. T. E., & Aiello, A. (1993). Articulatory rehearsal and phonological storage in working memory. *Memory & Cognition, 21,* 11–22.

Macken, W. J., & Jones, D. M. (2003). Reification of phonological storage. *Quarterly Journal of Experimental Psychology, 56A*, 1279-1288.

Macken, W. J., Tremblay, S., Houghton, R., Nicholls, A. P., & Jones, D. M. (2003). Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 43-51.

Martin, C. S. Mullennix, J. W., Pisoni, D. B., & Summers, W.V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15,* 676-684.

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.

Miller, G. A., & Chomsky, N. (1963). Finitary Models of Language Users. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of Mathematical Psychology, Vol. II.* New York: Wiley.

Miller, G. A., & Heise, G. A. (1950). The trill threshold. *Journal of the Acoustical Society of America*, *22,* 637-638.

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology, 63,* 176–185.

Moray, N. (1960). Broadbent's filter theory: postulate H and the problem of switching time. *Quarterly Journal of Experimental Psychology, 12,* 214-220.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, *47*, 379-390.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85,* 365-378.

Nairne, J. S. (1990) A feature model of immediate memory. *Memory & Cognition*, *18,* 251-269.

Neath, I. (2000). Modelling the effect of irrelevant speech on memory. *Psychonomic Bulletin and Review, 7*, 403-423.

Nicholls, A. P., & Jones, D. M. (2002). Capturing the suffix: Cognitive streaming in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28,* 12-28.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics, 57,* 989-1001.

Page, M. P. A., Madge, A., Cumming, N., & Norris, D. G. (2007). Speech errors and the phonological similarity effect in short-term memory: Evidence suggesting a common locus. *Journal of Memory and Language, 56*, 49-64.

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

Pardo, J. S., & Remez, R. E. (2006). The perception of speech. In M. Traxler and M. A. Gernsbacher (Eds.), *The Handbook of Psycholinguistics, 2nd Edition* (pp. 201-248). New York: Academic Press.

Penney, C. G., & Butt, A. K. (1986). Within- and between-modality associations in probed recall: A test of the separate-streams hypothesis. *Canadian Journal of Experimental Psychology, 40,* 1-11.

Perham, N. R., Marsh, J. E., & Jones, D. M. (in press). Syntax and serial recall: How language supports short-term memory for order. *Quarterly Journal of Experimental Psychology.*

Pisoni, D. B. (1997). Some thoughts on “normalization” in speech perception. In K. A. Johnson & J. Mullennix & (Eds.), *Talker variability in speech processing* (pp. 9-32). New York: Academic Press.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116,* 129-156.

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

Reisberg, D., Rappaport, I., & O’Shaughnessy, M. (1984). Limits to working memory: The digit digit-span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10,* 203–221.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review, 101,* 129-156.

Rogers, W. L., & Bregman, A. S. (1993). An experimental evaluation of three theories of auditory stream segregation. *Perception & Psychophysics, 53,* 179-189.

Schubotz, R. I. (2007). Prediction of external events with our motor system: Towards a new framework. *Trends in Cognitive Sciences, 11,* 211-218.

Sternberg, S., Wright, C. E., Knoll, R. L., & Monsell, S. (1982). Motor programs in rapid speech: Additional evidence. In R. A. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, N.J: Erlbaum.

ten Hoopen, G. (1996). Auditory attention. In O. Neumann & A. F. Sanders (Eds.), *Attention* (Vol. 3, pp. 79–122). London: Academic Press.

Treisman, A. M. (1971). Shifting attention between the ears. *Quarterly Journal of Experimental Psychology, 23,* 157-167.

Tulving, E., & Colotla, V. A. (1970). Free recall of trilingual lists. *Cognitive Psychology, 1,* 86-98.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64,* 49-60.

van Noorden, L. P. A. S. (1975). Temporal coherence in the perception of tone sequences. Unpublished doctoral thesis, Eindhoven University of Technology.

Verbrugge, R. R., Strange, W., Shankweiler, D. P., & Edman, T. R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America, 60,* 198-212.

Warren, R. M. (1999). *Auditory Perception: A New Analysis and Synthesis.* New York: 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.

Watkins, O. C., & Watkins, M. J. (1980). Echoic memory and voice quality: Recency recall is not enhanced by varying presentation voice. *Memory & Cognition, 8,* 26-30.

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

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 and Language, 58*, 48-65.

Appendix: Additional Counterbalancing Measures Incorporated Into The Design of Experiments 1-3

Experiment 1

The 14 trials within each condition were further sub-divided as follows:

‘With lead-in block’: 14 Alt-Alt: 7 trials in which the TBR list started with a female item and 7 in which it started with a male item; 14 Alt-Single trials: 7 in which the TBR list was female-spoken and 7 in which it was male-spoken; 14 Single-Single trials: 7 in which the TBR list was female-spoken and 7 in which it was male-spoken; 14 Single-Alt trials: 7 in which the TBR list started with a female item and 7 in which it started with a male item.

‘Without lead-in block’: 14 single-voice TBR lists: 7 female lists, 7 male lists; 14 alternating-voice TBR lists: 7 female item first, 7 male item first.

Experiment 2

The 14 trials within each condition were further sub-divided as follows:

‘With lead-in block’: 14 Same-Voice/Alt trials**:** 7 in which the lead-in was presented in a female voice and the TBR list started with a female item and 7 in which the lead-in was presented in a male voice and the TBR list started with a male item; 14 Different Voice/Alt trials: 7 in which the TBR list started with a female item and 7 in which it started with a male item; 14 Same-Voice/Single trials: 7 in which both the lead-in and TBR list was female-spoken and 7 in which both the lead-in and TBR list was male-spoken; and 14 Different-Voice/Single trials: 7 in which the TBR list was female spoken and 7 in which it was male spoken.

Experiment 3

The 28 trials in each condition were further sub-divided as follows:

28 Single-voice trials: 7 in each voice: F, F-, F+, and F++; 28 Alternating-voice trials: 14 started with the F voice and 14 started with the F+ voice; 28 Four-voice trials: 14 forming the pattern F F+ F++ F+ F F- F F+ and 14 forming the pattern F+ F F- F F+ F++ F+ F.

Footnotes

1. Talker variability effects have been found in several domains other than short-term memory (long-term word recognition; e.g., Mullennix, Pisoni, and Martin, 1989). However, the present article is concerned specifically with the putative mechanisms underpinning the impact of talker variability in serial recall and we use the term 'talker variability effect' in this restricted sense unless indicated otherwise (see also General Discussion).

2. The countdown did not, of course, have to be reproduced. The start of the list was marked by the end of the countdown, i.e., once the countdown reached '1' (or 'z' in Experiment 2), the participant would know that the next item was the first item of the TBR list.

3. Given that the presentation rate used in Experiment 1 was relatively fast (1 item/350ms), we ran a supplementary experiment to check that the same interaction between talker variability and lead-in is found also with a slower rate typical of some serial recall experiments (1 item/750ms, e.g., Henson, Burgess, Hitch, & Flude, 2003; Hughes, Vachon, & Jones, 2007). Other than the presentation rate—which we increased to 1item/750ms by changing the inter-stimulus interval to 500ms—the experiment was essentially identical to the main Experiment 1 except we only included the Single, Alt, Single-Single, and Alt-Alt conditions. The same pattern was found: There was a main effect of Serial position, $F(7, 175) = 89.07$, $MSE = .02$, $p < .001$, a main effect of List-type, $F(1, 25) = 19.90$, $MSE = .03$, $p < .001$, no main effect of Lead-in, $F < 1$, but again a significant interaction between Lead-in and List-type, $F(1, 25) = 9.15$, $MSE = .02$, $p < .01$, whereby the TVE was larger with a lead-in. The results indicate that using a relatively fast presentation rate to

investigate the functional characteristics of the TVE is unlikely to compromise the generalizability of the results.

4. We purposefully pitch-shifted the original female voice rather than using the original sets of male and female voices and recording another two additional talkers so that the degree of acoustic difference between each successive pair of adjacent items was roughly equal for the two and four-voice conditions (this was also the reason for choosing the particular pattern of voice-changes used in the four-voice condition). If we had used recordings from four different talkers (or used a different pattern), it would be very difficult to manipulate ‘number of different voices’ without confounding this factor with differences in the degree of acoustic difference between successive items.

5. During the peer-review process, it was suggested that the perceptual-gestural mismapping account might also predict an increase in a particular type of order error whereby non-adjacent items are transposed more frequently in an alternating-voice list than in a single-voice list (whereas adjacent transpositions are by far the most usual type of order error; see, e.g., Henson et al., 1996). This prediction does not necessarily follow however: With an alternating voice list, it is possible that it is a tendency to mistakenly *recall* (rather than transpose) the items by voice that is increased. This would in itself be an error given that recall is scored according to strict temporal order. Importantly, such by-voice recall could potentially result in an increase in the most common, that is, immediately-adjacent, transposition error. Consider the list ABCDEFGH. If alternating voices tend to promote by-voice output then the list might be recalled as AB-DC-E-GF-H (by-voice output indicated by the dashes). Thus, recalling by voice would have led to adjacent transposition errors between C and D and between G and F in this case. Thus,

whilst it is plausible that participants might also show an increase in non-adjacent transpositions (due to transposing by voice), whether this would be expected to be greater than the increase in adjacent transpositions (due to recalls-by-voice) is unclear. In the event, transposition analyses of the data from Experiments 1-4—not reported fully here for the sake of brevity—showed that whilst alternating-voice lists increased the number of transpositions generally there was no consistent evidence that such lists disproportionately increased a particular type of (e.g., non-adjacent) transposition error.

Author note

Robert W. Hughes, John E. Marsh, and Dylan M. Jones, School of Psychology, Cardiff University, UK. Dylan Jones is also adjunct professor at the Department of Psychology, University of Western Australia. The research reported in this article received financial support from the United Kingdom's Economic and Social Research Council in the form of a grant awarded to Dylan Jones, Robert Hughes, and Bill Macken. The results of some of the studies in the current article have been presented at the 5th Annual *Auditory Perception Cognition and Action Meeting* (APCAM), Houston, Texas, November, 2006; the Meeting of the *Experimental Psychology Society*, Cardiff, UK, April, 2007, and at the 48th Annual Meeting of the *Psychonomic Society*, Long Beach, California, November, 2007. Thanks are due to Todd Bailey for statistical advice and to Bill Macken for several useful discussions about the work reported here. Correspondence can be addressed to Robert W. Hughes at the School of Psychology, Cardiff University, PO Box 901, Cardiff CF10 3AT, United Kingdom. Email may be sent to HughesRW@cardiff.ac.uk.

Table captions

Table 1. A schematic representation of the six conditions contrasted in Experiment 1. Single = Single voice; Alt = Alternating voices. For conditions 3-6, the first part of each condition-label refers to the voice presentation-format of the lead-in whilst the second refers to that for the TBR list.

Table 2. A schematic representation of the six conditions contrasted in Experiment 2. Single = Single voice; Alt = Alternating voices. For conditions 3-6, the first part of each condition-label refers to the voice presentation-format of the lead-in whilst the second refers to that for the TBR list.

Figure captions

Figure 1. Mean percentage of items correctly recalled at each serial position in the six conditions of Experiment 1 (see Table 1 for an illustration of the six conditions).

Figure 2. Mean percentage of items correctly recalled at each serial position in the six conditions of Experiment 2 (see Table 2 for an illustration of the six conditions).

Figure 3. Mean percentage of items correctly recalled at each serial position in the Single, Alt, and Four-voice conditions in Experiment 3.

Figure 4. Mean percentage of items correctly recalled in the Single and Alt conditions for phonologically dissimilar (Diss) and phonologically similar (Sim) lists in Experiment 4.

Table 1

Condition	**Voice**	**Lead-in**	**To-be-remembered list**
1. Single	Female (or male) voice:		6 5 2 7 1 4 8 3
2. Alt	Female (or male) voice:		6 2 1 8
	Male (or female) voice:		5 7 4 3
3. Single-Single	Female (or male) voice:	8 7 6 5 4 3 2 1	6 5 2 7 1 4 8 3
4. Alt-Alt	Female (or male) voice:	8 6 4 2	6 2 1 8
	Male (or female) voice:	7 5 3 1	5 7 4 3
5. Alt-Single	Female (or male) voice:	8 6 4 2	6 5 2 7 1 4 8 3
	Male (or female) voice:	7 5 3 1	
6. Single-Alt	Female (or male) voice:		6 2 1 8
	Male (or female) voice:	8 7 6 5 4 3 2 1	5 7 4 3

Table 2

Condition	**Voice**	**Lead-in**	**To-be-remembered list**
1. Single	Female (or male) voice:		6 5 2 7 1 4 8 3
2. Alt	Female (or male) voice:		6 2 1 8
	Male (or female) voice:		5 7 4 3
3. Same-Voice/Single	Female (or male) voice:	R S T U V X Y Z	6 5 2 7 1 4 8 3
4. Same-Voice/Alt	Female (or male) voice:		6 2 1 8
	Male (or female) voice:	R S T U V X Y Z	5 7 4 3
5. Diff-Voice/Single	Different (male) voice:	R S T U V X Y Z	
	Male (or female) voice:		6 5 2 7 1 4 8 3
6. Different-Voice/Alt	Different (male) voice:	R S T U V X Y Z	
	Male (or female) voice		6 2 1 8
	Male (or female) voice:		5 7 4 3

Figure 1

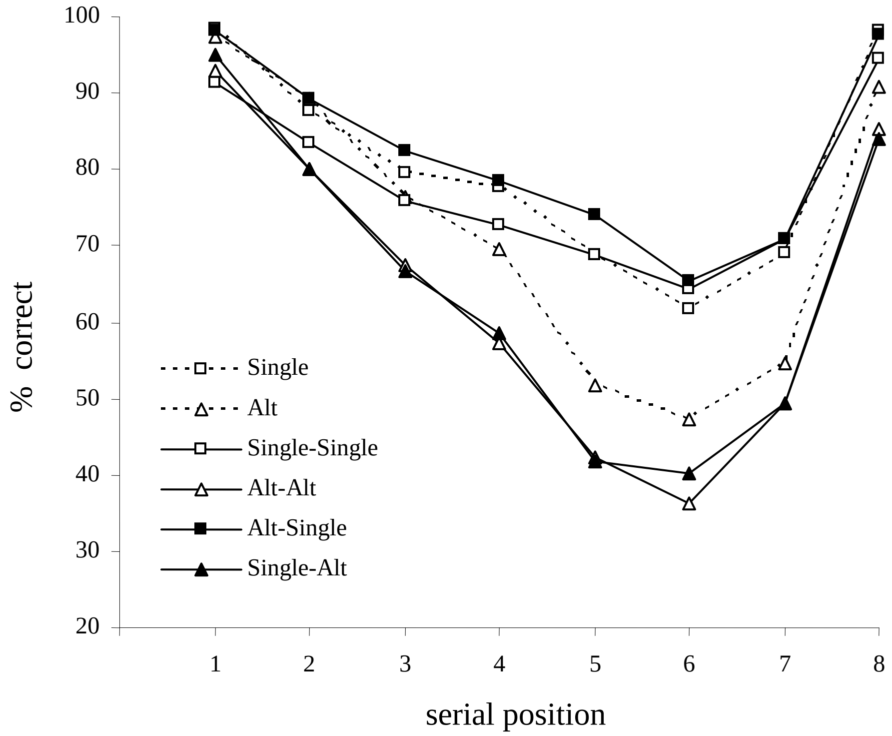

100
90
80
70
60
50
40
30
20
% correct
Single
Alt
Single-Single
Alt-Alt
Alt-Single
Single-Alt
1
2
3
4
5
6
7
8
serial position

Figure 2

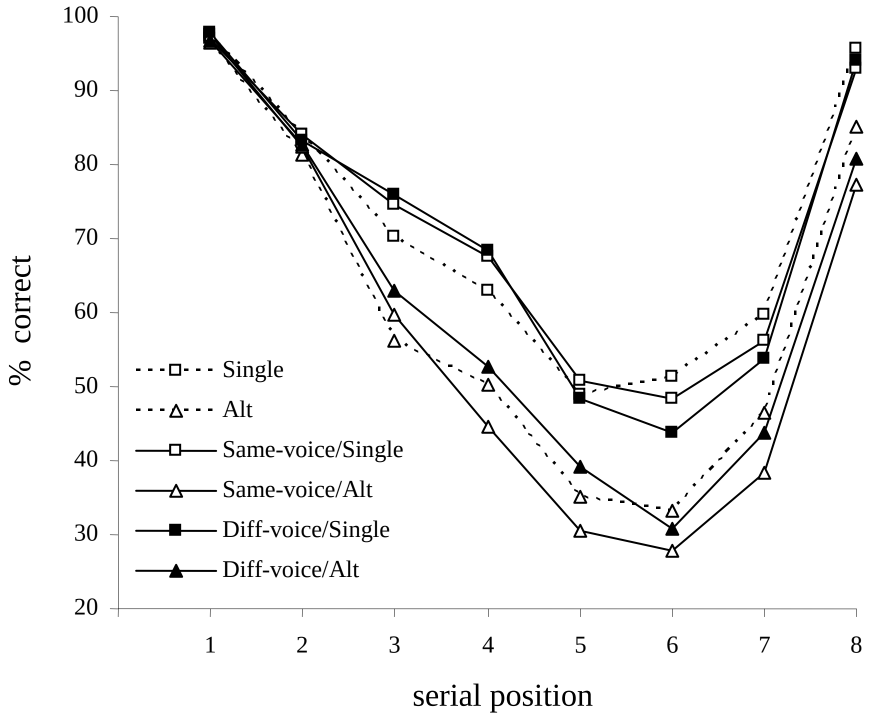

100
90
80
70
60
50
40
30
20
% correct
Single
Alt
Same-voice/Single
Same-voice/Alt
Diff-voice/Single
Diff-voice/Alt
1
2
3
4
5
6
7
8
serial position

Figure 3

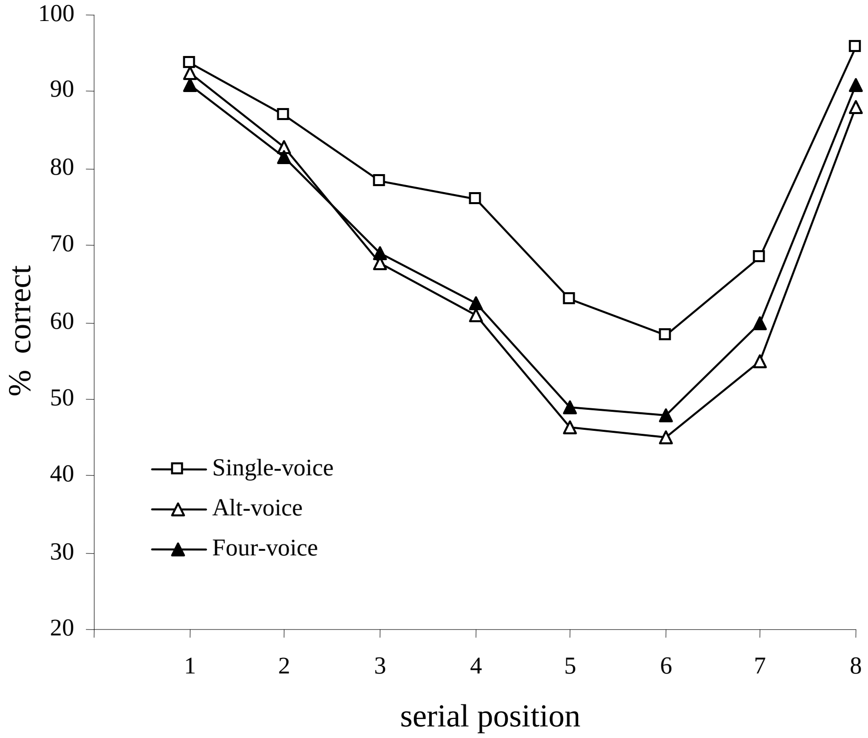
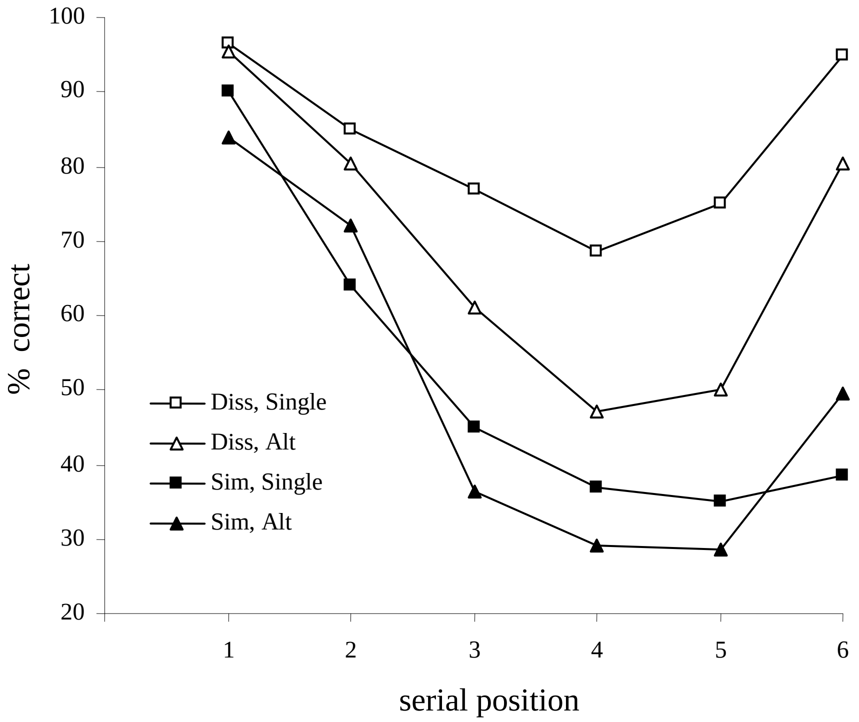

100
90
80
70
60
50
40
30
20
% correct
Single-voice
Alt-voice
Four-voice
1
2
3
4
5
6
7
8
serial position

Figure 4


100
90
80
70
60
50
40
30
20
% correct
Diss, Single
Diss, Alt
Sim, Single
Sim, Alt
1
2
3
4
5
6
serial position