Journal of Memory and Language 90 (2016) 126–146

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Perceptual-motor determinants of auditory-verbal serial short-term memory ☆

Robert W. Hughes [a,*], Cindy Chamberland [b], Sébastien Tremblay [b], Dylan M. Jones [c,1]

[a] Department of Psychology, Royal Holloway, University of London, Egham, Surrey, UK
[b] École de psychologie, Université Laval, Québec, Canada
[c] School of Psychology, Cardiff University, Cardiff, UK

ARTICLE INFO

Article history:
Received 25 July 2015
revision received 20 April 2016
Available online 8 May 2016

Keywords:
Perceptual organization
Motor-planning
Short-term memory
Serial recall
Perceptual variability

ABSTRACT

The role of the compatibility between obligatory perceptual organization and the active assembly of a motor-plan in auditory-verbal serial recall was examined. The classic finding that serial recall is poorer with ear-alternating items was shown to be related to spatial-source localization, thereby confirming a basic tenet of the perceptual-motor account and disconfirming an early account characterizing the two ears as separate input-channels (Experiment 1). Promoting the streaming-by-location of ear-alternating items—and therefore the incompatibility between perceived and actual order—augmented the ear-alternation effect (Experiment 2) whereas demoting streaming-by-location by reducing the regularity of the alternation attenuated it (Experiment 3). Finally, increasing the perceptual variability of an ear-alternating list while demoting the likelihood of streaming-by-location—by adding uncorrelated voice changes—also reduced the ear-alternation effect as did articulatory suppression for that part of the list (pre-recency) associated with motor-planning (Experiment 4). The results are incompatible with theories in which perceptual variability impairs serial recall due to a deficit in encoding items into a limited-capacity short-term memory space and instead point to a central role for perceptual and motor processes in serial short-term memory performance.

Introduction

The ability to retain and reproduce a sequence of stimuli over the short-term has long been recognized as a fundamental aspect of cognition, playing a critical role in many higher-level functions including problem-solving, reasoning, speech processing, and language learning (e.g., Baddeley, 1986, 2007; Hurlstone, Hitch, & Baddeley, 2014; Lashley, 1951; Rosenbaum, Cohen, Jax, Weiss, & van der Wel, 2007). Classically, the study of verbal short-term memory has been wedded conceptually to the formalist, psycholinguistic, tradition (cf. Chomsky & Halle, 1968) in which the key unit of analysis is modality-independent, phonological, representation (e.g., Baddeley, 2007, 2012; Baddeley & Hitch, 1974; Broadbent, 1958,

☆ Author note: We would like to thank Al Bregman for valuable comments on an early version of the paper. Thanks are due also to Mark Hurlstone for programming Experiments 1 and 3 and Marie-Josée Côte, Laurence Dumont, Ben Jones, and Marie-Kim Côté for data collection. The research was supported in part by an *Economic and Social Research Council* grant awarded to Dylan M. Jones, Bill Macken, & Robert W. Hughes. The results of Experiments 1 and 2 were presented at the Meeting of the *Experimental Psychology Society*, University College London, January, 2014, and at the *International Conference on Working Memory*, Cambridge, July, 2014.

* Corresponding author.
E-mail address: Rob.Hughes@rhul.ac.uk (R.W. Hughes).
[1] Also at the University of Western Australia.

http://dx.doi.org/10.1016/j.jml.2016.04.006
0749-596X/© 2016 The Authors. Published by Elsevier Inc.

1984; Pashler, 1998). In this view, the capacity to recall a series of verbal items is understood in terms of the operation of a dedicated but highly fragile system or representational space in which central (e.g., phonological) representations of such items outlast their physical presence but are subject to inexorable decay or/and interference from other structurally similar items (Baddeley, 2007). Accordingly, research set within this centralist view has tended to focus on structural factors such as the duration of the short-term representation (e.g., Baddeley, Thomson, & Buchanan, 1975; Barrouillet & Camos, 2014), the particular mechanism by which it might be corrupted by other events occupying the same memorial space (e.g., Neath, 2000; Oberauer, 2002), or the overall capacity of that space (e.g., Cowan, 2001, 2015).

We argue here that an emphasis on the structural properties assumed to underpin short-term memory performance has obscured a key role for general-purpose perception and motor functions that have often been seen as peripheral; merely providing the input to, and output from, central short-term memory mechanisms. Indeed, while some centralist accounts now make strong links between short-term memory mechanisms and perception and action processes (e.g., Page, Madge, Cumming, & Norris, 2007), there is a burgeoning body of work suggesting that a consideration of perceptual organization, motor-planning, and the mapping between them may go a long way to accounting for short-term memory phenomena without having to invoke dedicated storage mechanisms (e.g., Guérard & Tremblay, 2011; Hughes & Jones, 2005; Hughes, Marsh, & Jones, 2009, 2011; Jones, Hughes, & Macken, 2006, 2007; Jones, Macken, & Nicholls, 2004; Macken, Taylor, & Jones, 2014, 2015; Maidment & Macken, 2012; for related views, see Acheson & MacDonald, 2009; MacDonald, 2016; Reisberg, Rappaport, & O'Shaughnessy, 1984; Wilson & Fox, 2007). The aim of the present research is to examine the way in which the passive perceptual organization of the auditory scene into coherent temporally-extended perceptual objects or streams (cf. Bregman, 1990) is a key determinant of the short-term reproduction of a spoken sequence. More specifically, we study the importance of the compatibility between the obligatory, non-volitional, organization of spoken items according to their perceived spatial-source and the active assembly of those items in their true temporal order in (subvocal) articulatory form. As a theoretical counterpoint, we contrast the predictions of our perceptual-motor account with ones allied to the centralist view in which factors such as variation in spatial-source are seen as compromising the initial encoding of items into short-term memory rather than ones that affect processes that are integral to short-term memory performance.

Serial short-term memory: a perceptual-motor approach

Present understanding of serial short-term memory is based primarily on performance in a verbal serial recall task in which, typically, around 5–8 verbal items (e.g., digits, letters, words) are presented one at a time and which must be reproduced in strict serial order following the last item (Baddeley, 1966; Conrad, 1964). A key observation at the heart of the perceptual-motor account is that the post-categorical identities of the items in a serial recall list, by design, exhibit very low transitional probabilities (i.e., the predictability of an item given the preceding event(s) is very low or zero; e.g., Miller & Chomsky, 1963): That is, the Experimenter typically strips the list of supra-item features—syntax, grammar, and semantic relations—that in a normal linguistic sequence constrain the serial order of its constituent elements (see, e.g., Jefferies, Lambon Ralph, & Baddeley, 2004; Macken & Jones, 2003). At least in relatively 'pure' serial recall tasks, in which the burden falls entirely or primarily on reproducing item order rather than individual item identity (cf. Baddeley, 2012), it is this characteristic (alone) that makes serial recall challenging. Accordingly, performance is superior when there is a good match between the list and long-term sequential knowledge (i.e., when the list-items exhibit relatively high transitional probabilities) such as with a list containing high-frequency letter transitions (Miller & Selfridge, 1950), a list of words that make up a grammatically legitimate sentence (Jefferies et al., 2004), or a list that contains sub-sequences already unitized in long-term memory due to repeated exposure to, for instance, telephone or personal identification numbers (Jones & Macken, 2015).

Given the lack of serial order constraints within the presented list, we argue that the motor-sequence planning system (vocal-articulatory in the case of verbal serial recall) is co-opted to impose such constraints. The skill of speech (or more accurately, speech-planning) provides a particularly effective medium for this purpose on account of its inherent sequentiality, continuity, and prosodic and co-articulatory nature. For example, the very act of (covertly) co-articulating the items—whereby the exact manner in which the end of one speech element is articulated depends on the next speech element (Sternberg, Wright, Knoll, & Monsell, 1980)—generates new sequential information (and hence constraints) not present in the list itself (e.g., Woodward, Macken, & Jones, 2008). Importantly, then, in the perceptual-motor account, the function of articulatory rehearsal (or vocal-motor planning) is not the refreshing of decaying representations within a distinct verbal (phonological) store (or the conversion of visually-presented items into phonological form; e.g., Baddeley, 2007) but rather the motoric sequential binding of otherwise post-categorically unbound items.

A second key facet of the perceptual-motor approach is a consideration of obligatory perceptual processes that organize sensory input into coherent perceptual objects (e.g., Wertheimer, 1923/1938). Of particular relevance are the especially powerful processes of auditory scene analysis that generate objects or 'streams'; representations that, unlike representations of visual stimuli, are inherently sequential given the temporally-unfolding nature of sound (e.g., Bregman, 1990). To elaborate, auditory sequential streaming involves the computation of whether or not successive sounds are emanating from the same environmental event and whether they should therefore be assigned to the same stream (stream integration) or to different streams (stream segregation), respectively. This is widely assumed to be an obligatory, non-volitional, process involving the preattentive processing of a variety of cues well characterized by Gestalt principles such as grouping by similarity and 'good continuity' (e.g., Bregman, 1990; Sussman, Bregman, & Lee, 2014; Sussman, Horváth, Winkler, & Orr, 2007; Wertheimer, 1923/1938; Winkler, Denham, & Nelken, 2009).

The relevance of auditory perceptual organization to the ostensibly separate domain of serial short-term memory derives from the well-established relation between auditory sequential streaming and serial order processing. When stimuli that are distinct from one another but are nevertheless relatively perceptually invariable on account of some broader common ground (e.g., different words spoken from the same location)—and hence are likely to be integrated into a single stream—the order of those stimuli can be readily discerned and reported. Once non-temporally-successive elements share a stronger common ground than temporally-successive elements, however—leading to stream segregation—the true temporal order of elements becomes very difficult to identify. For example, Bregman and Campbell (1971), in a study of stream-segregation by fundamental-frequency, found that given a looping sequence of three different high-frequency tones (ABC) alternating with three different low-frequency tones (123)—i.e., A1B2C3—listeners typically reported the order of the tones as ABC123 (or 123ABC) but could not apprehend the order of tones from different frequency-ranges (i.e., the actual temporal order of the tones, e.g., that 1 followed A). This can be understood by supposing that the two distinct frequency-ranges provided the basis for assignment of the ABC tones to one stream and the assignment of the 123 tones to another stream. However, the perception of the changes carried on each stream (e.g., A–B–C) yielded cues as to the order of the stream's constituent elements such that within-stream, but not across-stream, order was readily reportable.

Thus, the perceptual-motor account posits that rather than reflecting dedicated systems or mechanisms, performance in serial short-term memory tasks is based on exploiting any matches between the list (or parts thereof) and long-term semantic memory (see, e.g., Jones & Macken, 2015) and—to the extent that there is no perfect match—on an episodic record of the active motoric reconstruction of the list (regardless of list-modality). In addition, with auditory presentation, the episodic record of the way in which the sequence was automatically perceptually organized may also support (though may also hinder; see below) the reproduction of the list (for a more extensive theoretical discussion, see Macken et al., 2015). More generally, we assume that all sensory processing—including processing of the simulated sensory consequences of imminent motor-output (cf. Clark, 2013)—produces an episodic record that outlasts the physical presence of the external stimuli or the internal simulation and whose contents reflects the way in which the sensory input was organized. To the degree that all sensory and motor processes lead to lasting effects, we deem it unnecessary to demarcate a subset of mental mechanisms that 'store' information: storage is the by-product of other (perceptual and motor-planning) processes (see also Craik & Lockhart, 1972).

There is already a relatively large body of work documenting a key role for auditory perceptual organization in numerous serial short-term memory phenomena, including the disruptive effect of task-irrelevant sound on serial recall (Jones, Alford, Bridges, Tremblay, & Macken, 1999), the advantage in the recall of the last few items of an auditorily-compared to visually-presented list (i.e., the modality effect; Nicholls & Jones, 2002; see also Macken, Taylor, Kozlov, Hughes, & Jones, in preparation), and the residual detrimental effect of phonological similarity on auditory-verbal serial recall under conditions in which articulatory rehearsal is impeded through articulatory suppression (Jones et al., 2004, 2006, 2007; Maidment & Macken, 2012). Our interest in the present paper centers on the influence on serial recall of the degree of compatibility between the non-volitional perceptual streaming of a spoken sequence and the deliberate assembly of a motor-plan designed to support the serial reproduction of the sequence.

Perceptual variability and auditory-verbal serial recall

We examine the role of perceptual-motor mapping here in the context of a relatively neglected phenomenon: the impairment of auditory-verbal serial recall when the list-items are perceptually variable. Specifically, it has been shown that serial recall is impaired markedly if the items are presented in different voices (Goldinger, Pisoni, & Logan, 1991; Greene, 1991; Hughes et al., 2009, 2011; LeCompte & Watkins, 1993; Martin, Mullennix, Pisoni, & Summers, 1989) or from different spatial locations (or vary in terms of ear-of-presentation; e.g., Treisman, 1971). Within the classical centralist approach to short-term memory, these perceptual variability effects have been interpreted in terms of a difficulty in initial item-encoding that usurps resources needed for central short-term storage. For example, the detrimental impact of presenting items in different voices has been explained in such terms. One variant of this *item-encoding account* invokes the process of speech normalization whereby the pre-categorical features of a spoken item (e.g., fundamental frequency, accent, speaking-rate) must be discarded at the point of input in order to realize an idealized abstract-phonological representation (e.g., Stevens, 1960). In this view, when such precategorical features differ from one item to the next in a serial recall task, the normalization process is particularly stressed thereby impairing the categorical encoding of each item and in turn impairing the storage of items already in short-term memory (e.g., Martin et al., 1989; Mogensen, Miller, & Roodenrys, 2015). Another item-encoding account posits that precategorical features of a spoken item are incorporated with its abstract-linguistic content into the short-term memory representation. To the extent that this incorporation process, like speech-normalization, is resource-demanding, it also impairs central short-term storage processes (Goldinger et al., 1991; Nygaard, Sommers, & Pisoni, 1995). Of particular interest given the form of perceptual variability focused upon in the present study, the detrimental effect on serial recall of presenting successive items alternately to the two ears (e.g., Moray, 1960; Treisman, 1971; see also Broadbent, 1954; Broadbent & Gregory, 1964) has also been explained in terms of an item-encoding problem, though this time by recourse to an attentional mechanism: "...there is a limit to the rate at which attention can be shifted between the ears. This attention shifting time would reduce the time available for perception and storage..." (Treisman, 1971, p. 164).

It is notable that most models of short-term memory have not explicitly addressed perceptual variability effects. While we think this is remiss, it is understandable given the traditional emphasis on post-perceptual, modality-independent, mechanisms. Indeed, a core construct within arguably the most prominent short-term memory models is a distinct modality-independent phonological short-term store (e.g., Baddeley, 2007; Burgess & Hitch, 1999, 2006; Henson, 1998; Page & Norris, 1998). As speech normalization is a necessary assumption of all such phonology-based models, they align naturally with a speech-normalization account of the ear-alternation effect. However, these and other models of short-term memory models also retain the concept of a central executive that controls the entry of stimuli into the short-term store (or one of several such stores; Baddeley, 1996) or a limited 'focus of attention' (Cowan, 1999; Oberauer, 2002). Thus, the idea that ear alternation demands a costly process of selective attention-switching to encode each item could also be readily adopted by such models.

The goal of the present research is to show that perceptual variability effects instead reveal a core, not peripheral, role for perception and, more specifically, the role of the compatibility between perceptual organization and motor-planning, in auditory-verbal serial recall. Our analysis begins with the observation that the centralist, item-encoding, approach focuses on the deficit produced under perceptually variable conditions. That is, the impairment serves as an empirical referent for the central or/and limited-capacity character of the cognitive structures assumed to support item-storage. In contrast, on the perceptual-motor account, perceptual variability effects bring into relief the accomplishment of obligatory perceptual object-formation processes that, as a by-product of their primary function, play a role in the representation and reproduction of the order of items in the typical, (relatively) perceptually invariable, spoken serial recall list (e.g., items spoken in the same voice from the same location). Specifically, we contend that the effect of perceptual variability serves to reveal the fact that, in the typical list in which successive items enjoy a great deal of common ground (e.g., same spatial-source, same voice), stream integration is a key mechanism by which serial order is encoded and hence an important determinant of performance. In particular, when the constituents of a sequence are integrated passively into a single stream, there is a relatively high level of compatibility between the order of items as yielded by that integration process and the order in which items need to be assembled into the subvocal motor-plan. Thus, when the sequence contains a high degree of perceptual variability between successive items—to the extent that the variability causes the sequence to be segregated into two (or more) streams—the perception of order is incompatible with the required motor-plan and hence recall is impaired.

Present study

In previous work in which we began to reconceptualize perceptual variability effects as a perceptual-motor mapping problem, we first replicated the finding that presenting items in an alternating female-male voice fashion (e.g., "**3**, 1, **7**, 4, **6**, 2, **8**, 5"; bold and underlined items in female voice; remainder in male voice) impairs serial recall appreciably (see also Greene, 1991; Mogensen et al., 2015). We argued that this was due to the obligatory perceptual segregation of items by voice yielding two sets of by-voice order cues (**3-7-6-8** and 1-4-2-5) that mismatched the motor-planning of the true temporal order of the items (i.e., 3 1 7 4 [...]). In line with this account, promoting such stream segregation by allowing the segregation process to 'build up' during a voice-alternating lead-in (a countdown into the to-be-remembered list, i.e., **8**, 7, **6**, 5, **4**, 3, **2**, 1 [...]) increased the cost of voice-alternation (see also current Experiment 2). In the present article, we apply the perceptual-motor account for the first time to the effects of perceptual variability on the spatial dimension, specifically, variability in the ear to which successive spoken items are presented (cf. Treisman, 1971).

As noted, classical accounts of the effect of ear-alternation on verbal serial recall have been couched in terms of the cost of switching a selective attention mechanism that protects a central limited-capacity short-term memory system from overload. In this view, the selection mechanism is too slow to shift between the ears to 'catch' each item to give them access to the short-term store (e.g., Broadbent, 1958; Treisman, 1971). The normalization/incorporation accounts of the impact of talker variability on serial recall (Martin et al., 1989; Mogensen et al., 2015; Nygaard et al., 1995) could also be extended readily to the effect of ear alternation: In this view, rather than a failure of timely attention-switching, the impairment could either be attributed to having to strip each item of ear-of-entry or spatial-location information—which would be a nuisance variable in the task of achieving a canonical abstract representation—or attributed to a cost of incorporating that information into the short-term memory representation. (Conversely, the attention-switching account of the ear-alternation effect could be readily applied to talker variability effects; for a discussion, see Hughes et al., 2009, 2011.)

Our alternative, perceptual-motor, account of the ear-alternation effect begins with a consideration of how the auditory-perceptual system determines the spatial-source of a sound. In a natural auditory scene, two main cues are used for sound localization, which in turn is one of several cues the auditory system uses to determine whether a succession of sounds are emanating from the same environmental object (e.g., a given talker) and hence whether they should be assigned to the same stream. One is interaural intensity difference (IID; e.g., Culling & Summerfield, 1995): The greater the signal-energy received by, say, the left ear compared to the right, the more the stimulus is assigned to the left of auditory space (from the listener's perspective). A second cue is inter-aural time-difference (ITD; Darwin & Hukin, 1999): If, for instance, the right ear receives the same spectral input later than the left ear, again the stimulus will tend to be assigned to the left portion of space as the sound-emitting event must be closer to the left than the right ear.[2] In the context of serial recall, then, when items are presented from one spatial location (e.g., all presented to both ears or all presented to one ear), the items will tend to be assigned to a single coherent stream and hence perceived order will map relatively well onto the required motor-plan (assuming the absence of other cues promoting stream segregation, e.g., alternating voices; Hughes et al., 2009). With ear-alternating items, however, the auditory system is likely to deduce that there are two environmental events emanating from two different loci and thereby partition the sequence into two streams, each containing serial order information at odds with the required motor-plan.

We begin with a test of an axiomatic assumption of the perceptual-motor account, namely, that the ear-alternation effect is a by-product of streaming-by-location. From this perspective, the ears per se are only important in the ear-alternation effect insofar as IID can be used to locate a sound in space, which in turn is just one of several possible streaming cues. Showing that it is spatial-source alternation that is key to the 'ear'-alternation effect, not alternation between the ears per se, would also rule out an early variant of the item-encoding hypothesis in which the effect is due to a difficulty in switching attention between the ears as two discrete structural input-channels (Treisman, 1971; see also Broadbent, 1954; Cherry & Taylor, 1954). Specifically, then, we test the prediction of the perceptual-motor account that if items alternating between the ears were perceived as coming from a common spatial source—and hence likely to be assigned to the same stream—the ear-alternation effect should be reduced or eliminated. To achieve this apparently paradoxical set of affairs we adapt a technique used by Deutsch (1979) in a study examining the difficulty of recognizing a melody when its constituent tones are alternated between the ears. Similar to the logic underpinning the present experiment, Deutsch (1979) investigated whether the difficulty was due to the fact that non-adjacent tones in the melody were streamed-by-location—based on their sharing a common IID—such that the order relations between successive tones (critical for melody recognition) could not be readily apprehended (cf. Bregman & Campbell, 1971), rather than a problem of switching attention between the ears. Of most relevance here, she found that melody recognition with ear-alternating tones improved dramatically if a drone was presented along with each tone to the other (i.e., contralateral) ear (see also Judd, 1979; Schubert & Parker, 1956). This can be explained in terms of a mislocalization illusion arising from a phenomenon known as *auditory contralateral induction* (Warren & Bashford, 1976): As each tone was presented, the contralateral drone served to reduce the IID—including the intensity difference at the frequency characterizing the target tone—such that the auditory system was 'fooled' into perceiving the target tone as if it was also present at the ear receiving the noise. As a result, the listener would hear the target-tone as if it were presented from a more frontal–central source. Thus, a succession of ear-alternating tones would now have been perceived as sharing a similar frontal–central spatial source and be assigned, therefore, to a single coherent stream, restoring thereby the perception of the temporal order between successive tones of the melody (see Judd, 1979). In line with the general argument forwarded here, Deutsch (1979) argued that "when a decrement in integrating input to the two ears occurs, this is due not to capacity limitation, but rather to a mechanism which is imposed to prevent confusion" (p. 3). Here we apply the contralateral induction-based technique for the first time in the context of the ear-alternation effect in serial recall.

Experiment 1

The upper section of Table 1 illustrates the four conditions included in Experiment 1 (note that the 'Lead-in' column is only relevant to Experiment 2). In the *Binaural* condition, each item in the list was presented to both ears. In the *Alt* condition, they were presented to the two ears in an alternating fashion. The main novel condition introduced in this experiment is condition 4, *Alt + noise*: This was identical to the *Alt* condition except each to-be-remembered item was accompanied by a white noise-burst presented to the contralateral ear (note that noise rather than a drone has been shown to produce contralateral induction in numerous studies; Judd, 1979; Schubert & Parker, 1956; Warren & Bashford, 1976). Finally, a *Binaural + noise* condition was included to control for any effect of the mere presence of alternating noise-bursts; here, the to-be-remembered items were presented binaurally together with noise-bursts alternating between the ears.

The perceptual-motor account predicts that introducing the alternating noise-bursts to the Alt list (i.e., *Alt + noise*) should induce a 'to-the-centre' (mis)localization of each of the ear-alternating to-be-remembered items such that those items would now tend to be assigned to a single coherent stream in contrast to the tendency toward a two-stream organization in the *Alt* (i.e., without noise) condition. We predicted, therefore, that the noise should reduce the ear-alternation effect: Whereas performance in the *Alt* condition should be poorer than in the *Binaural* condition (i.e., replicating the basic ear-alternation effect), performance in the *Alt + noise* condition should be more comparable to that in the *Binaural* condition. This prediction contrasts with any item-encoding account of perceptual variability effects that invokes the notion of attention switching between two input-channels: The noise, despite providing a mislocalization cue, does not contain any information that can be used for

[2] ITD is not relevant when a signal is presented solely to one headphone or the other as was the case in the ear-alternating conditions of the current as well as previous studies of the ear-alternation effect. This is because ITD depends on the same signal reaching both ears (at some point) which, to all intents and purposes, does not occur when signals are panned fully left or right through headphones.

Table 1
A schematic representation of the conditions contrasted in each of Experiments 1–4. Single = Single ear; Alt = Alternating ears; reg = regular; irr = irregular; # = noise burst. The different fonts used for each digit in the multi-voices conditions (Experiment 4) represent different voices. The four conditions of Experiment 4 were undertaken in both a no-articulatory-suppression condition and a with-articulatory-suppression condition. See Method section of each experiment for several other details not depicted in the table.

Experiment/*Condition*	Ear	Lead-in																List							
Experiment 1																									
1. Binaural	Both:																	6	5	2	7	1	4	8	3
2. Alt	Left:																	6		2		1		8	
	Right:																		5		7		4		3
3. Binaural+noise	Left:																	#		#		#		#	
	Both:																	6	5	2	7	1	4	8	3
	Right:																		#		#		#		#
4. Alt+noise	Left:																	6	#	2	#	1	#	8	#
	Right:																	#	5	#	7	#	4	#	3
Experiment 2																									
1. Single	Left (or right):																	6	5	2	7	1	4	8	3
2. Alt	Left:																	6		2		1		8	
	Right:																		5		7		4		3
3. Single+lead-in	Left:	8		6		4		2		8		6		4		2		6	5	2	7	1	4	8	3
	Right:		7		5		3		1		7		5		3		1								
4. Alt+lead-in	Left:	8		6		4		2		8		6		4		2		6		2		1		8	
	Right:		7		5		3		1		7		5		3		1		5		7		4		3
Experiment 3																									
1. Single	Left (or right):																	6	5	2	7	1	4	8	3
2. Alt-reg	Left:																	6		2		1		8	
	Right:																		5		7		4		3
3. Alt-irr	Left:																	6	5		7			8	
	Right:																			2		1	4		3
Experiment 4																									
1. Single-ear—Single-voice	Left (or right):																	6	5	2	7	1	4	8	3
2. Single-ear—Multi-voices	Left (or right):																	6	5	2	7	1	4	8	3
3. Alt-ear—Single-voice	Left:																	6		2		1		8	
	Right:																		5		7		4		3
4. Alt-ear—Multi-voices	Left:																	6		2		1		8	
	Right:																		5		7		4		3

discriminating the identity of each to-be-remembered item. Thus, the need to switch attention between two putative input-channels is the same in the *Alt + noise* as in the *Alt* condition, leading to the prediction that, in this view, the cost of alternation should not vary as a function of the presence of noise-bursts.

Method

Participants

Thirty-six undergraduates from Cardiff University took part in return for course credit. All reported normal hearing and normal or corrected-to-normal vision.

Apparatus & materials

The experiment was controlled by a PC computer using the *E-Prime 1.1* software (Psychology Software Tools). The to-be-remembered stimuli were sequences of eight items taken without replacement from the digit set 1–8. Each item was recorded digitally in a male voice and sampled with a 16-bit resolution at sampling rate of 44.1 kHz using *Sound Forge 7.0* software (Sony Corporation, 2003). The duration of each item was edited to 250 ms using the same software. For each sequence, the digits were presented in a pseudorandom order with the constraint that there were no ascending or descending runs of more than two digits (e.g., 3-4-5 or 8-7-6). The sequences were presented at approximately 65 dB(A) over headphones with an inter-stimulus interval (ISI; offset to onset) of 100 ms, giving a presentation rate of one item per 350 ms.

Design

The experiment had a 2 (List-type: Binaural, Alt) by 2 (Noise: Present, Absent) by 8 (Serial position) fully repeated-measures design. Note that we used a binaural sequence for the non-alternating condition in this experiment (rather than single-ear presentation as we used in subsequent experiments) to avoid the risk with single-ear presentation—unlike binaural presentation—that the addition of an alternating noise-burst would mask every other to-be-remembered item. In the *Binaural + noise* and *Alt + noise* conditions, a white noise-burst was presented at the same intensity as the to-be-remembered items. Each white noise-burst had the same duration (250 ms) and the same onset as each to-be-remembered item; thus, the ISI for the noise-bursts was also the same as for the to-be-remembered items (100 ms). There were 80 trials in total comprising a block of 40 trials in which the lists were accompanied by alternating noise-bursts and another 40 trials in which they were not. Within each of these blocks, there were 20 *Binaural* lists and 20 *Alt* lists intermixed with the constraint that the same condition was not presented more than twice in succession. Before each block, there were two practice trials corresponding to the two conditions represented in that block. The order of blocks was counterbalanced across participants.

Procedure

Participants were tested individually in a dimly lit and sound-attenuated room. For all conditions, after the presentation of the last to-be-remembered item of each list the word *recall* was displayed for 100 ms. Participants were told that regardless of the type of list (binaural or alternating) that they should try to recall the to-be-remembered digits in the actual temporal order in which they were presented by writing them on a response sheet in a strict left-to-right fashion. They were instructed to guess if they were uncertain of any digit's position. Participants had 15 s to recall each list. Three seconds before the end of the response period, a 500 ms tone was presented over the headphones to indicate that the presentation of the first item of the next trial was imminent. The experiment took approximately 45 min.

Results and discussion

For all experiments reported in the present article, the raw data were scored according to the strict serial recall criterion; an item was scored as correct only if it was recalled in the same absolute position in which it was presented. Fig. 1 shows recall performance in each of the four conditions of Experiment 1.

The results are clear-cut: The impairment caused by ear alternation was markedly reduced when a noise burst was presented to the ear contralateral to that receiving a to-be-remembered item. A 2 (List-type: Binaural, Alt) × 2 (Noise: Present, Absent) × 8 (Serial position) repeated-measures ANOVA revealed a main effect of List-type, $F(1,35) = 24.94$, $MSE = .028$, $p < .001$, $\eta_p^2 = .42$, reflecting a replication of the basic ear-alternation effect and, of greater interest, the ear-alternation effect interacted with Noise, $F(1,35) = 12.08$, $MSE = .020$, $p = .001$, $\eta_p^2 = .26$. Planned contrasts confirmed that the difference between *Binaural* and *Alt* was considerably greater, $F(1,35) = 28.95$, $MSE = .060$,

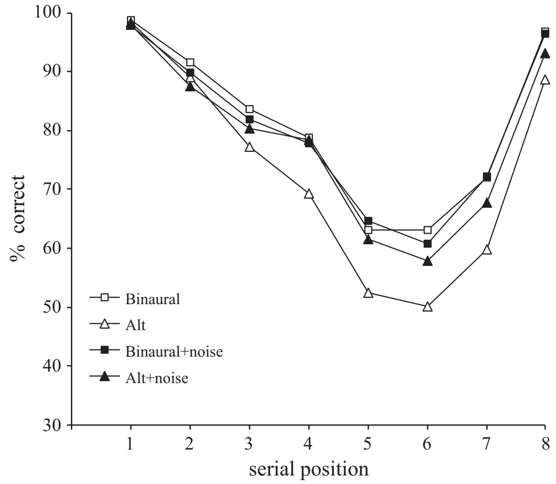

Fig. 1. Mean percentage of items correctly recalled at each serial position in each of the four conditions of Experiment 1, showing the reduction of the ear-alternation effect in the presence of contralateral noise-bursts.

$p < .001$, than when those same list-types were accompanied by noise-bursts; indeed this latter difference did not reach significance, $F(1,35) = 3.47$, $MSE = .035$, $p = .071$. A further contrast showed that the improvement in performance with ear-alternating items in the presence of noise (i.e., the difference between *Alt* and *Alt + noise* conditions) was indeed significant, $F(1,35) = 10.19$, $MSE = .070$, $p < .005$. There was also a main effect of Noise, $F(1,35) = 4.27$, $MSE = .030$, $p = .046$, $\eta_p^2 = .11$, a by-product of the fact that one of the without-noise conditions (*Alt*) resulted in poorer recall than the other conditions [noise per se clearly had no direct impact on performance as indicated by the absence of a significant difference, $F < 1$, between *Binaural* (81% correct recall averaged across serial positions) and *Binaural + noise* (80% correct)].[3]

The results of Experiment 1 provide definitive evidence that the ear-alternation effect in serial recall is not caused by the alternation of the items between the ears per se but is related to the fact that the degree of discrepancy in intensity across the two ears is a major cue for localizing sound in distal space (e.g., Culling & Summerfield, 1995). When the IID for an item presented to one ear is reduced by presenting a noise-burst to the contralateral ear, it is mislocalized perceptually as having a more central spatial source (Warren & Bashford, 1976). A succession of such

[3] It might be suggested that the perceptual-motor account would predict that ear-alternation should lead to a particular type of error whereby non-adjacent pairs are output (e.g., "8-1" or/and "3-5" given the list "8-3-1-5…"), reflecting the partitioning of the alternate items into two streams. This is not, however, a necessary or strong prediction of the account: The impairment due to ear alternation may well be as much due to the additional burden of having to resist such by-stream recall as it is due to actual by-stream recall. Nevertheless, for completeness, for each condition in Experiment 1, we calculated the frequency of such non-adjacent pairings regardless of the absolute serial positions at which they occurred in the output (so for the example above, "3-5" would be scored as a non-adjacent pairing even if it occurred late in the output) and divided this by the total number of errors. No significant differences were found between any of the four conditions.

centrally mislocalized ear-alternating items would thus tend to be integrated into a single, coherent, stream. As a result, perceived order—despite ear-alternation—is now compatible with the to-be-reproduced order and hence ear-alternation has little effect. The results also rule out any item-encoding account of the ear-alternation effect that appeals to early notions of a difficulty in switching attention sufficiently rapidly between the ears as two structurally discrete input-channels (e.g., Treisman, 1971): The presence of contralateral noise bursts would not have altered the extent to which only one ear/input-channel was receiving information required to identify each item and hence would not have altered the need for ear-switching.

The results of Experiment 1—while confirming a necessary prediction of the perceptual-motor account and refuting an early, ear/channel-switching, account—do not rule out the item-encoding hypothesis more generally. The hypothesis could readily accommodate the results of Experiment 1 by supposing that attention needs to switch between spatial sources (rather than ears per se) to encode each item in the *Alt* but not in the *Alt + noise* condition. Similarly, the normalization/incorporation variants of the hypothesis could assume that the irrelevant information to be discarded/incorporated is spatial-source information rather than ear-of-entry per se. The three remaining experiments of the present series do, however, speak to the adequacy of the item-encoding hypothesis more generally as well as providing further tests of the perceptual-motor account.

Experiment 2

In Experiment 2, we capitalize on the fact that the perceptual-motor account and the item-encoding hypothesis make different predictions in relation to the impact of pre-exposure to the ear-alternating pattern on the ear-alternation effect. It is well established that stream segregation takes some time to 'build up': given that stream assignment involves a comparison of a stimulus to a previously-established regularity, there must necessarily be some accumulation of evidence for that regularity (Bregman, 1990). For example, when listeners are presented with a sequence of alternating low- and high-frequency tones, the auditory system's default is to assume initially the presence of a single (fluctuating) sound-source such that all stimuli are assigned to one (albeit rather unstable) stream. As the regularity of the repeating low-tone and repeating high-tone pattern continues, however, the auditory system eventually accrues sufficient evidence that it is more probable that the stimulation is caused by two distinct events, hence the low and high tones are split into separate streams (Anstis & Saida, 1985; Beauvois & Meddis, 1997; Bregman, 1978; Rogers & Bregman, 1993).

Based on the cumulative nature of stream segregation, we sought here to promote the partitioning of ear-alternating items into two streams by introducing a (task-irrelevant) induction sequence (or 'lead-in') that preceded the to-be-remembered list. The lead-in was in the form of a countdown ("8, 7, 6 [...] 1" presented twice in succession) presented in the same voice, and in the same ear-alternating fashion, as the ear-alternating to-be-remembered items. Our rationale was that the partitioning of ear-alternating items into distinct streams would begin during the lead-in such that an ensuing sequence of ear-alternating to-be-remembered items would be more clearly partitioned than would be the case in the absence of a lead-in. Thus, if the effect is indeed driven by stream segregation, the impairment of serial recall due to ear-alternation (compared to a list presented to a single ear) should be greater when such segregation is promoted by a lead-in. The 'Experiment 2' section of Table 1 depicts the four conditions contrasted. The *Single* and *Alt* conditions (conditions 1 and 2) were supplemented by conditions 3 and 4 in which a *Single* list and an *Alt* list, respectively, was preceded by an alternating-ear lead-in. The perceptual-motor account makes the unique prediction that the impairment usually found in the *Alt* compared to *Single* condition (i.e., the basic ear-alternation effect) will be greater when a lead-in is introduced (note that the *Single + lead-in* condition provides the most appropriate comparison-condition to *Alt + lead-in* as it controls for a possible effect of the mere presence of a lead-in).

Turning to predictions of the item-encoding hypothesis, on the view that attention fails to 'keep up' with the changes in spatial location in an ear-alternating list, there are good reasons to expect that pre-exposure to the extent and rhythm of the spatial changes should serve to entrain attention such that it is better able to track those changes by the time the to-be-remembered list begins. For example, the time taken to identify the features of a target-sound is facilitated if attention is entrained via a preceding sequence of irrelevant sounds to the rhythm of the sounds and hence the temporal onset of the target-sound (Jones, Moynihan, MacKenzie, & Puente, 2002). Similarly, the provision of a cue regarding the probable frequency (Mondor & Bregman, 1994) or location (Mondor & Zatorre, 1995) of an upcoming target-sound facilitates responding to other dimensions of that sound (e.g., its duration). Thus, on the item-encoding hypothesis, an ear-alternating lead-in should, if anything, facilitate the timely encoding of the ear-alternating to-be-remembered items and hence serial recall should be impaired to a lesser extent (than with an ear-alternating list in the absence of the lead-in). The same prediction follows from the perceptual normalization/incorporation accounts: It is well established that perceptual processes become attuned over time to whatever perceptual variability (e.g., talker, accent, speaking rate, location, and so on) is to be eliminated (normalization account) or retained (incorporation account) from the speech signal (e.g., Diehl, Souther, & Convis, 1980; Kidd, 1989). Again, therefore, pre-exposure to the pattern of ear/location-alternation via a lead-in should, if anything, facilitate item-encoding and thereby reduce the ear-alternation effect.

Finally, to the extent that we have demonstrated previously that a (voice-) alternating lead-in accentuates the voice-alternation effect (Hughes et al., 2009), the present experiment will speak to our assumption that the ear- and voice-alternation effect are underpinned by common (perceptual-motor) mechanisms. Such identicality in the empirical character of the two effects would also buttress our supposition that item-encoding accounts of the voice-alternation effect (Mogensen et al., 2015) should also, in principle, apply to the ear-alternation effect.

Method

Participants

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

Apparatus and materials

As illustrated in Table 1, in the *Single* condition, the to-be-remembered lists were presented to a single ear (for half these lists, all the items were presented to the left ear; for the other half, all items were presented to the right ear). In the *Alt* condition, the to-be-remembered lists were presented in an ear-alternating fashion (whether the first item was presented to the left or right ear was counterbalanced). The *Single + lead-in* and *Alt + lead-in* conditions were identical to the *Single* and *Alt* conditions, respectively, except that a lead-in (an "8-1" countdown presented twice) was presented in an ear-alternating fashion before the to-be-remembered list. The last item of the lead-in was always presented to the ear opposite to that receiving the first item of the to-be-remembered list (regardless of whether that to-be-remembered list was a single-ear or ear-alternating list). The soundfiles used for the lead-in digits were the same as those for the to-be-remembered lists and the timing for the lead-in (item-duration and ISI) was also identical to that for the to-be-remembered lists.

Design and procedure

These aspects of the method were identical to those for Experiment 1 except that, effectively, the manipulation of the presence (or not) of noise-bursts was replaced with that of the presence (or not) of a lead-in. Participants were told that for one block of trials the spoken list would be preceded by two countdowns.

Results and discussion

The percentage of items correctly recalled in order as a function of the four conditions [2 (Lead-in: Present, Absent) × 2 (List-type: Single, Alt)] across the eight serial positions is shown in Fig. 2. The pattern confirms the prediction of the perceptual-motor account and can be summarized in terms of two main results: First, the basic ear-alternation effect was again replicated: Recall was poorer in both conditions involving an ear-alternating to-be-remembered list (i.e., *Alt* and *Alt + Lead-in*) than in conditions involving a single-ear to-be-remembered list (i.e., *Single* and *Single + Lead-in*). Second, and of most interest, recall of an ear-alternating list (but not a single-ear list) was particularly poor when it was preceded by a lead-in; that is, a lead-in increased the magnitude of the ear-alternation effect.

A 2 (Lead-in: Present, Absent) × 2 (List-type: Single, Alt) × 8 (Serial position) analysis of variance (ANOVA) showed a main effect of Serial position, $F(7,133) = 79.25$, $MSE = .129$, $p < .001$, $\eta_p^2 = .81$. There was no main effect of Lead-in, $F < 1$, indicating that a lead-in, in and of itself, did not affect recall. However, there was a main effect of List-type, $F(1,19) = 28.97$, $MSE = .044$, $p < .001$, $\eta_p^2 = .60$, replicating the ear-alternation effect and, of most interest, a significant interaction between Lead-in and List-type, $F(1,19) = 7.05$, $MSE = .013$, $p < .02$, $\eta_p^2 = .27$, reflecting the fact that the impairment with an ear-alternating list was greater when it was preceded by a lead-in. For completeness, we note that there were also significant interactions between Lead-in and Serial position, $F(7,133) = 2.68$, $MSE = .013$, $p < .05$, and between List-type and Serial position, $F(7,133) = 7.13$, $MSE = .020$, $p < .001$, though we will not attempt to attach any functional significance to these.

Experiment 2 showed that the deleterious effect of ear-alternation on serial recall is increased by pre-exposure to the ear-alternating pattern of stimuli. Importantly, the lead-in per se did not have any effect in and of itself as indicated by the fact that it had no effect on single-ear list recall; thus, its effect on an ear-alternating list is selective and cannot be attributed to some general processing load. We argue, therefore, that the ear-alternating lead-in provided an accumulation of evidence that there were two different sources of information such that by the time the alternating to-be-remembered list started, the partitioning of the items into two streams was stronger than in the absence of the lead-in (cf. Rogers & Bregman, 1993). This stronger partitioning into streams would exacerbate the difficulty of mapping the items onto a motor-plan designed to support the reproduction of the items in their actual, not perceived, temporal order. At the same time, the results are at odds the item-encoding hypothesis which would have predicted that pre-exposure to the pattern of ear/location alternation should ease the burden on attention-switching or perceptual normalization/incorporation and hence reduce the impact of ear alternation. We note, finally, that the fact that the ear-alternation effect is influenced in the same way by an alternating lead-in sequence as reported previously in the context of voice alternation (Hughes et al., 2009) reinforces our view that they are underpinned by a common mechanism.

Experiment 3

In Experiment 3, we use the converse analytic device to that of Experiment 2 to garner convergent evidence to adjudicate between the two theoretical approaches: We now sought to demote stream segregation (thereby predicting, on the perceptual-motor account, a reduction of the ear-alternation effect as in Experiment 1) using a manipulation that should at the same time increase the burden on item-encoding (thereby predicting, from the standpoint of the item-encoding cost hypothesis, an increase in the ear-alternation effect). Specifically, we introduced an ear-alternating condition in which the alternation between left (L) and right ears (R) was irregular within a trial and where that irregular pattern was also unpredictable from trial to trial (e.g., LLRLRRLR followed by, e.g., RLRRLRLL, etc.) and contrasted this

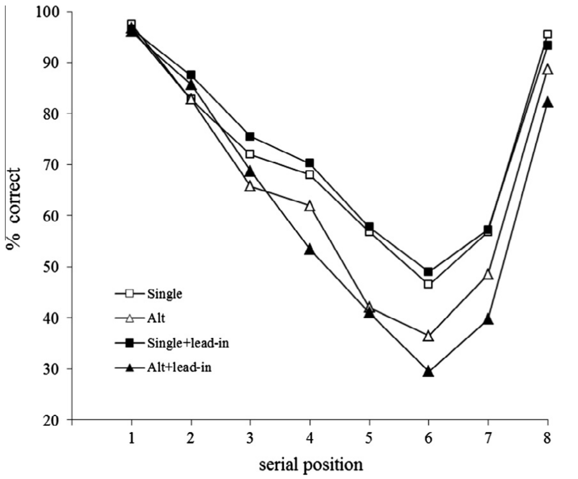

Fig. 2. Mean percentage of items correctly recalled at each serial position in the four conditions of Experiment 2, showing the increase in the magnitude of the ear-alternation effect in the presence of a lead-in.

Alternation-irregular (*Alt-Irr*) condition with a condition in which the alternation was regular and fixed (and hence predictable) across trials (i.e., the same as the *Alt* conditions of Experiments 1 and 2; now labelled *Alt-Reg* for this experiment; for an illustration of the three conditions, see the 'Experiment 3' section of Table 1). Thus, in the *Alt-Irr* condition, stream segregation of successive items would now be demoted on the grounds that segregation is a function of the regularity and frequency with which non-successive sounds exhibit mutual belongingness (Bregman, 1990) and hence the impact of ear-alternation should be reduced in this condition. In relation to the item-encoding hypothesis, in contrast, attention-switching or perceptual normalization/incorporation would be expected to be under greater duress in the *Alt-Irr* compared to the *Alt-Reg* condition: Irregular and unpredictable alternation should reduce markedly the opportunity to anticipate the need (or not) to switch attention for each item or to anticipate the spatial information to be discarded/incorporated (cf. Mondor & Bregman, 1994; Mondor & Zatorre, 1995). Thus, the item-encoding accounts predict a greater impairment with irregular compared to regular ear-alternating items.

Method

Participants

Seventeen undergraduates from Cardiff University took part in return for course credit.

Apparatus, materials, design & procedure

All these aspects of the method were the same as in Experiments 1 and 2 except for the following. The design comprised two within-participant factors. The first was List-type with three levels implemented across three separate blocks of trials: (1) *Single* (as in Experiment 2); (2) *Alt-Reg* in which the eight to-be-remembered items were alternated between the ears in a regular fashion (e.g., LRLRLRLR) for all trials; and (3) *Alt-Irr* in which items alternated between the ears in a different, irregular, fashion (e.g., LRRLRLLR) for each trial across a block (cf. Table 1). The irregularity was in fact pseudo-random given that we imposed the constraints that no more than two items could be presented to the same ear in succession and that each ear would, as in the *Alt-Reg* condition, receive four items over the course of the trial. The second factor was Serial position. There were 60 experimental trials divided into three blocks: 20 trials for each of the three levels of List-type. Each block was preceded by two practice trials corresponding to the condition represented in that block. The order in which the three blocks were undertaken was counterbalanced across participants. The experiment took approximately 35 min.

Results and discussion

Fig. 3 shows serial recall performance in the *Single*, *Alt-Reg*, and *Alt-Irr* conditions across the eight serial positions. It is clearly evident that the effect of ear-alternation was larger in the *Alt-Reg* condition than in the *Alt-Irr* condition. A 3 by 8 repeated-measures ANOVA confirmed this impression: There was a main effect of Serial position, $F(7,112) = 87.70$, $MSE = 9.51$, $p < .001$, $\eta_p^2 = .85$, and a main effect of List-type, $F(2,32) = 13.42$, $MSE = 9.23$, $p < .001$, $\eta_p^2 = .46$. Planned contrasts revealed a significant difference between the *Single* and *Alt-Reg* conditions, $F(1,16) = 23.07$, $MSE = 20.24$, $p < .001$, $\eta_p^2 = .59$, but only a non-significant trend for a difference between *Single* and *Alt-Irr* conditions, $F(1,16) = 2.68$, $MSE = 14.23$, $p = .12$, $\eta_p^2 = .14$. The difference between *Alt-Reg* and *Alt-Irr* was significant, $F(1,16) = 11.38$, $MSE = 20.94$, $p < .005$, $\eta_p^2 = .42$.

The results of Experiment 3 provide further support for the perceptual-motor account. In an *Alt-Irr* condition, the temporal irregularity and relative infrequency with which non-successive items were presented from the same location would have weakened any segregation of the items into two streams. At the same time, the pattern of results is opposite to that predicted by the item-encoding hypothesis: The uncertainty in the *Alt-Irr* condition as to when attention would need to be switched or what information would need to be discarded/incorporated during item-encoding should have accentuated not attenuated the effect. One potential counterargument to this interpretation from the standpoint of the attention-switching account might be that the smaller ear-alternation effect in the *Alt-Irr* condition may have been due to the fewer number of attention-switches required in that condition (five switches) compared to the *Alt-Reg* condition (seven switches). However, this argument would be difficult to sustain: It seems reasonable to suppose that such an account would predict a large difference when contrasting a 'no switches' condition (i.e., the *Single* condition) and a 'five switches' condition (i.e., the *Alt-Irr* condition) and a comparatively much smaller difference when comparing 'five switches' (*Alt-Irr*) and 'seven switches' (*Alt-Reg*). In fact, the difference between no-switches and five-switches was not significant.

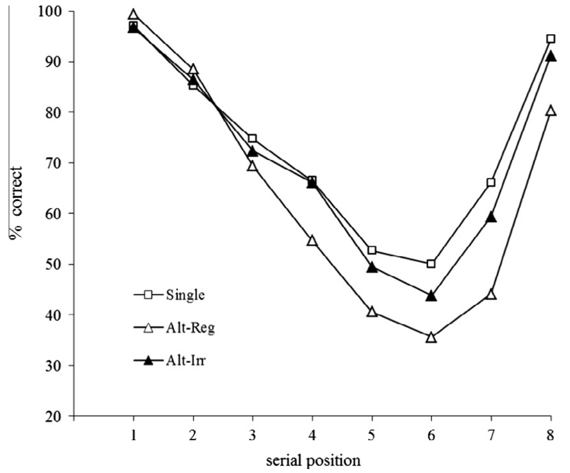

Fig. 3. Mean percentage of items correctly recalled at each serial position in the three conditions of Experiment 3, showing the greater effect of regular and predictable ear-alternation (Alt-Irr) compared to irregular and unpredictable ear alternation (Alt-Irr).

The results of Experiment 3 present perhaps the greatest challenge thus far for the item-encoding hypothesis: Despite a clear requirement in the *Alt-Irr* condition to switch attention between locations to encode the majority of the items or to discard/incorporate the location information for each item, serial recall was only slightly affected compared to the *Single* condition.

Experiment 4

In this experiment, we capitalize on a fundamental conceptual distinction between the perceptual-motor account and the item-encoding hypothesis: On the perceptual-motor account, the critical property of an ear-alternating sequence is that *non-successive* sounds are more *similar* to one another than are successive sounds and hence the non-successive sounds cohere together into two streams, over-riding the integration of successive items into one stream. In contrast, the critical property of an ear-alternating sequence from the standpoint of the item-encoding hypothesis is that *successive* sounds are *dissimilar* to one another, thereby increasing the burden on item-encoding via a need for attention-switching or perceptual normalization/incorporation; the relation between non-successive items is irrelevant on these accounts. In a standard ear-alternating sequence (LRLR[...]), these two properties are perfectly confounded. In the present experiment, we deconfound them by including a condition in which the perceptual dissimilarity between successive items was made particularly great by having each successive item not only shift between ears/locations (in a regular pattern) but also change in voice. Specifically, in the *Alt-ear—multi-voices* condition, the eight ear-alternating items were presented in eight different 'voices' (created by pitch-shifting recordings from a single voice; see Method for details). This therefore increased the dissimilarity between successive items but at the same time decreased the similarity between non-successive items. Thus, the item-encoding hypothesis predicts very straightforwardly that recall should be particularly poor in such a condition compared to when an ear-alternating sequence is relatively less variable (all items presented in the same voice): Attention would need to shift not only between ears/locations but also between different fundamental-frequency ranges (which were unpredictable from trial to trial). Indeed, it has been shown that responding to sounds that vary in both fundamental frequency and location is slower and less accurate than when they vary on only one of these dimensions (Mondor, Zatorre, & Terrio, 1998). Similarly, on the perceptual normalization/incorporation accounts, there would for each item be two dimensions upon which features would need to be discarded/incorporated, hence serial recall should again be particularly poor in the *Alt-ear—multi-voices* condition according to these accounts. Again, the perceptual-motor account makes the opposite prediction: In the *Alt-ear—multi-voices* condition, due to the relative incoherence of non-successive items (compared to the standard, *Alt-ear—single-voice*, sequence), streaming-by-location should be demoted. That is, a cue for segregation (ear-alternation) will be offset by a cue—uncorrelated changes in voice—that demotes the integration of those 'same-ear' sounds. Thus, the perceptual-motor account predicts that the effect of ear-alternation will be diminished with the addition of uncorrelated changes of voice.

Experiment 4 also tests the assumption of the perceptual-motor account that the effect of ear-alternation is located primarily in the mapping of the percept of the sequence onto a motor-plan. That is, serial recall of the typical (single-voice) ear-alternating sequence is impaired because the perception of order is at odds with the need to assemble the items into a motor-plan according to their actual order. If so, then reducing the capacity for motor-planning should attenuate the ear-alternation effect. We test this in the present experiment by examining whether the ear-alternation effect is attenuated under articulatory suppression whereby participants are required to utter an irrelevant sequence during list-presentation (e.g., Baddeley, Lewis, & Vallar, 1984; Jones et al., 2004; Murray, 1968); thus, one group of participants undertook all four [i.e., voice(2) × ear(2)] conditions while articulating an irrelevant sequence during list-presentation.

The perceptual-motor account makes detailed predictions regarding precisely how articulatory suppression would be expected to interact with the ear-alternation effect. Specifically, the attenuation should only be apparent for that portion of a spoken list—the pre-recency portion—that has been shown in previous studies to be associated with the assembly of items into a motor-plan. There is now ample evidence that the serial recall of the last few (e.g., 2 or 3) items of a spoken sequence (i.e., auditory recency) is not as reliant on motor-planning as pre-recency items, as evidenced by the fact that auditory recency (unlike pre-recency) survives articulatory suppression (Hitch, Flude, & Burgess, 2009; Jones et al., 2004, 2006; Macken et al., 2015; Maidment & Macken, 2012). Thus, recall of items at the end-boundary of a spoken sequence has been attributed instead to the direct use of the highly perceptually-salient end-boundary of an auditory sequence, by-passing much if any motoric recoding (e.g., Jones et al., 2004). Further evidence for this more passive route to the recall of list-end spoken items is that auditory recency is impaired by a spoken suffix regardless of articulatory suppression but only if that suffix is presented in such a way as to alter the passive organization of the end of the list (e.g., Maidment & Macken, 2012; Nicholls & Jones, 2002). Thus, to the extent that ear-alternation exerts its effect by modulating the perception of order, its effect should remain apparent under articulatory suppression at those points where recall of the sequence by-passes motor-planning processes (i.e., at recency).

To recapitulate, then, the present experiment involved four conditions undertaken by both a no-suppression group and a with-suppression group: The sequence was presented in an ear-alternating fashion or to a single ear and these two types of sequence could either be presented such that all items were presented in a single voice (as in Experiments 1–3) or in eight different voices.

Method

Participants

Fifty-two undergraduate students from Université Laval took part in return for a small honorarium. Half were assigned to a no-suppression group and half to a with-suppression group.

Apparatus and materials

The apparatus and materials were the same as for Experiments 1–3 except for the following details. A new set of the spoken digits 1–8 was recorded in French in a female voice and sampled with a 16-bit resolution at a sampling rate of 44.1 kHz using Sound Forge 5.0 (Sony Creative Software). The duration of each digit was again edited to 250 ms using the same software. Pitch-shifting was then applied to each digit using *Sound Forge* to create seven 'voice'-variations of the original recording (shifting the original recording of each digit by −10, −8, −6, −4, −2, +2, +4, and +6 semitones). There were therefore eight voice-versions of each digit (the original plus the 7 derivatives). A pilot study was conducted to ensure that the pitch-shifting had not rendered any of the stimuli incomprehensible. An independent sample of 5 participants from the same pool was asked to identify a series of digits spoken via headphones one at a time. All 128 possible stimuli to be used in the experiment proper [i.e., ear(2) × voice (8) × digit(8)] were presented once to each participant. The mean percentage of correct identification was very high generally at 98.44% and the identification rates across the eight voices fell within a narrow range, between 96.25% and 100% correct; a chi-square goodness-of-fit test confirmed that there was no difference in the frequency with which the digits were identified depending on voice, X^2 (7, N = 630) = 0.15, p = 1.

Four auditory conditions were constructed and are depicted schematically in the 'Experiment 4' section of Table 1. The *Single-ear—single-voice* condition was equivalent to the *Single* condition of Experiments 2 and 3: the to-be-remembered list was presented to one ear and in one voice (one of the eight voices selected at random). For half of these lists, all the items were presented to the left ear while for the other half all items were presented to the right ear. The *Alt-ear—single-voice* condition was equivalent to the *Alt* conditions of Experiments 1–3: the list was presented in one voice (one of the eight voices selected at random) in an ear-alternating fashion (whether the first item was presented to the left or right ear was counterbalanced). In the *Single-ear—multi-voices* condition, the list was presented to one ear (half the lists to the left, half to the right) but each item was spoken in a different voice. The order of the eight voices across such a list was determined randomly for each trial in this condition. Finally, in the *Alt-ear—multi-voices* condition, the list was presented in eight voices (again, their order determined randomly for each trial) in an ear-alternating fashion (half the lists starting with a left-ear item, half with a right-ear item).

Design

The experiment involved a mixed design with List-type-Ear (Single, Alt), List-type-Voice (Single, Multi), and Serial Position as within-participant factors and Suppression as a between-participants factor. Regardless of Suppression group, each participant undertook 80 experimental trials divided into two blocks: The Single-voice block comprised 40 trials made up of 20 *Single-ear—single-voice* trials and 20 *Alt-ear—single-voice* trials. The block was preceded by four practice trials, two for each condition. The other, Multi-voices, block comprised 40 trials made up of 20 *Single-ear—multi-voices* trials and 20 *Alt-ear—multi-voices* trials. This block was also preceded by two practice trials per condition. In both blocks, the trials were presented in a pseudorandom fashion 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. One group of participants were free to use subvocal motor-planning during all 80 trials (the no-suppression group). The other group were required to repeatedly vocalize the irrelevant sequence "A, B, C, D, A, B [...]" (in French) as the to-be-remembered list was presented (see below for further details).

Procedure

Participants were tested individually in a dimly lit room. Before the beginning of the experimental session, participants in the with-suppression group were given a training session with the articulatory suppression task. The training session required the participants to repeat aloud (to 'shadow') the sequence "A, B, C, D, A, B [...]" presented to them (to both ears) over headphones. The sequence was presented at the rate of one letter every 350 ms (250 ms on/100 ms off) in a male voice in French ('B', 'C', 'D', is pronounced in the same way as when reciting the English alphabet while 'A' is pronounced as /a/ not /ə/). The participants were required to repeat the letters at the rate at which they were presented to attune them to the rate at which they would be expected to utter the letters during the articulatory suppression in the experiment proper. The training session lasted approximately 2 min. For this group, each trial in the experiment proper began with the presentation of *A*, *B*, *C*, *D* on the screen for 2 s to signal to the participants that they were to begin articulatory suppression. They were instructed to read these letters aloud (but relatively quietly) at the rate they had learned to shadow them during the training session and to continue the suppression during item presentation until recall. The letters disappeared from the screen before the to-be-remembered list commenced. The Experimenter remained present throughout to ensure compliance with the articulatory suppression requirement. In the no-suppression group, the display *A*, *B*, *C*, *D* was replaced with a blank screen for the same duration.

For all auditory conditions and for both groups, after the last item of each list the word *recall* was displayed on the screen. As in all previous experiments, participants were instructed that they should try to recall the to-be-remembered digits in the temporal order in which they were presented by writing them on a response sheet in a strict left-to-right fashion. They were instructed to guess if they were uncertain of a digit's position. Again as in previous experiments, participants had 15 s to recall each list and after this delay a 500 ms tone was presented over the headphones to indicate that the next trial was imminent.

Results[4]

Fig. 4 shows the serial position curves for each of the eight conditions of Experiment 4 [2(List-type-Ear: Single, Alt) × 2(List-type-Voice: Single, Multi) × 2(Suppression)]. Turning first to the data from the no-suppression group (upper two panels), it is evident that the difference between single and ear-alternating conditions is smaller in the Multi-voices condition (upper-right panel) than it is in the Single-voice condition (upper-left panel). In the with-suppression group (lower two panels), the ear-alternation effect was again weak with multiple voices (lower-right panel) but was also relatively weak in the Single-voice condition except at recency (lower-left panel).

The results of an initial mixed 2 × 2 × 2 × 8 ANOVA showed a main effect of List-type-Ear, $F(1,50) = 39.13$, $MSE = .023$, $p < .001$, $\eta_P^2 = .44$, a main effect of Suppression, $F(1,50) = 15.26$, $MSE = .56$, $p < .001$, $\eta_P^2 = .95$, and a main effect of Serial position, $F(7,350) = 135.85$, $MSE = .068$, $p < .001$, $\eta_P^2 = .73$. While the main effect of List-type-Voice was not significant, $F(1,50) = 1.03$, $MSE = .117$, $p = .32$, there was a significant interaction between List-type-Voice and List-type-Ear, $F(1,50) = 13.96$, $MSE = .011$, $p < .001$, $\eta_P^2 = .22$. The only other significant interactions were those between List-type-Ear and serial position, $F(7,350) = 5.51$, $MSE = .008$, $p < .001$, and List-type-Voice and serial position, $F(7,350) = 2.52$, $MSE = .013$, $p < .02$. Importantly, all these effects/null effects and interactions were subsumed within a significant four-way interaction, $F(7,350) = 2.34$, $MSE = .006$, $p = .02$, $\eta_P^2 = .05$, consistent with our impression of the pattern evident across the four panels of Fig. 4.

We turned next to address our specific hypothesis regarding the interplay of ear alternation and voice-variability by analyzing the data from the no-suppression group only (cf. upper two panels of Fig. 4) in a 2 × 2 × 8 repeated-measures ANOVA. Here, the main effect of List-type-Ear was significant, $F(1,25) = 18.66$, $MSE = .03$, $p < .001$, $\eta_P^2 = .43$, while the main effect of List-type-Voice was not, $F(1,25) = .052$, $p > .05$. Of particular interest, List-type-Ear interacted significantly with List-type-Voice, $F(1,25) = 4.30$, $MSE = .01$, $p < .05$, $\eta_P^2 = .15$: The ear-alternation effect was appreciably smaller in the Multi-voices condition than in the Single-voice condition. Simple effects analyses showed that the ear-alternation effect was nevertheless reliable in both the Multi-voice condition, $p = .023$, and the Single-voice condition, $p < .001$. It should also be noted, however, that this List-type-Voice by List-type-Ear interaction was a joint product of a (small and non-significant) detrimental effect of changes in voice (*Multi-voice—single-ear*: $M = 64.7\%$ vs. *Single-voice—single-ear*: $M = 66.8\%$) coupled with a (small and non-significant) reduction of the detrimental effect of ear alternation in the multi-voices condition ($M = 60.9\%$) compared to single-voice condition ($M = 60\%$). Thus, this particular result, if considered in isolation, could be argued to provide stronger evidence against the item-encoding hypothesis than it does evidence in favor of the perceptual-motor account.

Finally, we turn to our second specific hypothesis regarding the effect of articulatory suppression on the ear-alternation effect within that condition (*Single-voice* condition) in which a relatively large ear-alternation effect was produced and hence where it might be possible to observe any attenuation of that effect by suppression. To examine our detailed predictions regarding the interplay of articulatory suppression on the ear-alternation effect for different parts of the list, we derived a 'curve-portion' factor by dividing the serial position curve into a pre-recency portion (first six serial positions) and a recency portion (final two serial positions). A 2 (List-type-Ear) by 2 (Suppression) by 2 (Curve-portion) mixed ANOVA on the data within the Single-voice condition (i.e., those depicted in the upper-left and lower-left panels of Fig. 4) showed a main effect of List-type-Ear, $F(1,50) = 81.04$, $MSE = .004$, $p < .001$, $\eta_P^2 = .62$, a main effect of Suppression, $F(1,50) = 9.52$, $MSE = .086$, $p < .005$, $\eta_P^2 = .16$, and of Curve-portion, $F(1,50) = 39.05$, $MSE = .020$, $p < .001$, $\eta_P^2 = .44$. The two-way interaction between List-type-Ear and Suppression was not significant but there were significant interactions between List-type-Ear and Curve-Portion, $F(1,50) = 13.01$, $MSE = .002$, $p < .002$, $\eta_P^2 = .21$, and, of most interest, between List-type-Ear, Curve-Portion, and Suppression, $F(1,50) = 11.52$, $MSE = .086$, $p < .002$, $\eta_P^2 = .19$. Further scrutiny revealed that whereas there was no interaction between List-type-Ear and Curve-portion in the No-suppression group, $F < 1$, there was indeed such an interaction in the With-suppression group, reflecting the weaker—though still significant ($p < .005$)—effect of ear alternation at pre-recency compared to recency in this group, $F(1,26) = 26.92$, $MSE = .002$, $p < .001$, $\eta_P^2 = .52$. [Note that the same pattern was found if recency was defined

[4] A more comprehensive set of descriptive and inferential statistics for Experiment 4 has been provided as *Supplementary Material*.

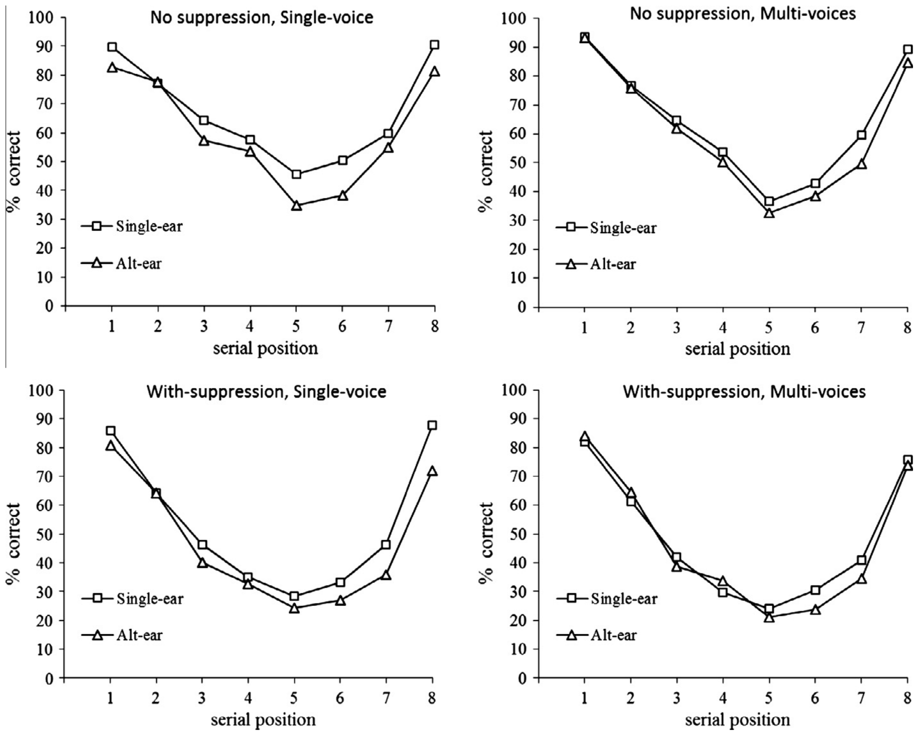

Fig. 4. Percentage correct serial recall at each serial position in the eight conditions of Experiment 4.

instead as the final position only (and pre-recency as the first seven positions) or defined as the last three positions (and pre-recency as the first five).] It is worth noting also that this relative survival of the ear-alternation effect at recency under suppression is not due simply to performance being much better generally at recency (i.e., some sort of scalar effect), as this pattern is not apparent in the With-suppression, multi-voices condition despite a strong recency effect.

Discussion

Experiment 4 produced several useful findings. First, at odds with the item-encoding hypothesis, introducing eight changes of voice to an ear-alternating list failed to exacerbate the impairment of serial recall; indeed the numerical trend was for a reduction of the impact of ear-alternation when voice-changes were added. Given that responses to sounds are slowed the more dimensions upon which they change from one to the next (Mondor et al., 1998), the item-encoding accounts incorrectly predicted that performance should have been particularly poor in the *Alt-ear—multi-voices* condition. Instead, the numerical reduction of the ear alternation effect in the context of concurrent changes in voice is consistent with—but, as acknowledged, does not on its own provide further strong support for—the perceptual-motor account. One, admittedly post hoc, suggestion as to why performance with ear-alternating lists did not benefit more greatly from changes in voice is that the degree of change between the voices was too modest to over-ride the strong cues to segregation provided by an alternation between highly disparate spatial locations.

Further supporting the perceptual-motor account, articulatory suppression attenuated the ear-alternation effect at pre-recency—independently shown to be associated with subvocal motor-planning (e.g., Hitch et al., 2009; Jones et al., 2004, 2006; Macken et al., 2015; Maidment & Macken, 2012)—relative to recency, which has been shown to be immune to articulatory suppression (e.g., Jones et al., 2004). This supports the tenet of the perceptual-motor account that the cost of perceptual variability for parts of the list usually supported by the assembly of items into a motor-plan is attributable to the incompatibility between perceived order and the order in which items need to populate that motor-plan. This result complements the finding that the voice-alternation effect is absent in short-term memory tasks that are not typically associated with a motor sequence-planning strategy (Hughes et al., 2011). Furthermore, to the extent that recall of the list-end boundary of a spoken sequence (i.e., auditory recency) is, in contrast, supported by a more direct use of the passively-derived percept of the sequence (Jones et al., 2004), any modulation of that percept by ear alternation would be expected to exert a detrimental effect even under articulatory suppression. Again, this very detailed prediction of the perceptual-motor account was confirmed in the present experiment.

Finally, it is worth highlighting the fact that there was no main effect of List-type-Voice. That is, voice-variability per se (e.g., within the single-ear condition) was not found to impair serial recall significantly even within the no-suppression group (although there was a numerical trend for such an impairment). At first glance, this aspect of the results seems at odds with the significant disruptive effects of voice-variability reported previously (e.g., Greene, 1991; Hughes et al., 2009, 2011; Martin et al., 1989; Mogensen et al., 2015). However, much of this research (Greene, 1991; Hughes et al., 2009, 2011; Mogensen et al., 2015) involved contrasting a voice-*alternation* condition (specifically, a male and female voice alternating) with a single-voice condition. In the present experiment, in contrast, the voice-variable condition involved presenting each item in a unique voice.[5] Indeed, the weak and non-significant effect in a unique-voice-per-item condition observed here is entirely consistent with the perceptual-motor account of the attenuating impact of voice-variability on the ear-alternation effect: From this standpoint, the precondition for a robust perceptual variability effect is that non-successive items are more coherent than successive items; successive items differing from one another is not sufficient. In a unique-voice-per-item sequence, as the voices are ordered randomly, the coherence between non-successive items would not, on average, be any greater than that between successive items, hence there would be no basis upon which to segregate the items into two streams. This begs the question, however, of why some studies have indeed found voice-variability effects using unique-voice-per-item sequences (Goldinger et al., 1991; Martin et al., 1989; Nygaard et al., 1995). A likely reason for this difference in outcomes is that all studies that have found an effect with unique-voice-per-item sequences have employed a serial recall task in which there was a particularly great burden on item as well as order recall: they used different words for each trial and relatively long lists (e.g., ten items). As Baddeley (2012) recently noted, to study processes specifically involved in short-term serial memory in a relatively process-pure way, different permutations of the same small set of items are used across trials (as was done throughout the present experiments) precisely to minimize the influence of long-term item memory (see also Hughes et al., 2009; present General Discussion). Thus, item-encoding costs might indeed play a role when there is a high load on item-memory and hence, under such conditions, effects are found using unique-voice-per-item sequences. However, these effects, we would argue, are long-term item-memory effects manifesting in the context of a (relatively process-impure) short-term serial recall task.

Another possible reason for the lack of a significant detrimental effect with a unique-item-per-voice sequence in the present experiment is that we used pitch-shifted versions derived from a single voice for our different 'voices' rather than eight real voices as in studies that have found such an effect (e.g., Goldinger et al., 1991; Martin et al., 1989). On the item-encoding hypothesis, it could be argued that an impairment is more likely when voices change not only in terms of frequency-range but also in terms of other features such as timbre, speaking-rate, accent, and emotional tone. However, this 'number-of-changing-features' explanation would again seem to predict, incorrectly, a greater impairment of recall when items change in both voice and location compared to location alone.

General discussion

To summarize the main impacts of the present series of experiments, Experiment 1 provided the first definitive demonstration that the impairment of auditory-verbal serial recall when items alternate between the ears is driven by the localization of the items in auditory space—through the use of IID—not by the fact they are presented to different ears per se: When the sharp changes in IID in an ear-alternating list are reduced by the presentation of noise-bursts in the contralateral ear, the ear-alternation effect is eliminated (Experiment 1). This result discounts an early variant of the item-encoding hypothesis in which the two ears serve as two discrete input-channels: The effect was eliminated despite the fact that the need to switch between any such structural channels would not have been altered by the noise-bursts (see also Deutsch, 1979). The results of Experiment 1 were also in line with the notion that the ear-alternation effect is driven by sequential perceptual organization: In this view, reducing the IID for each successive item promoted their sequential integration despite ear-alternation, thereby changing a perceptual-motor incompatible situation into a relatively perceptual-motor compatible one. More direct evidence for the role of perceptual organization was produced in Experiment 2 in which promoting the segregation of alternating items via pre-exposure to the alternating pattern (cf. Anstis & Saida, 1985; Bregman, 1978) increased the magnitude of the ear-alternation effect, converging with the impact of pre-exposure previously observed in the context of voice alternation (Hughes et al., 2009). The item-encoding hypothesis, however, would have predicted, if anything, a benefit, rather than a cost, of such pre-exposure. Further support for the perceptual-motor account and against the item-encoding hypothesis was found in Experiment 3: Demoting stream segregation while at the same time increasing the burden on item-encoding by making the alternation of the items between the ears irregular and unpredictable markedly attenuated the ear-alternation effect (indeed, the effect of an irregularly alternating sequence was not statistically significant in this experiment). In Experiment 4, we co-varied ear-alternation with voice changes and observed a (non-significant) trend for an improvement in the recall of an ear-alternating list despite an *increase* in perceptual variability between successive items. This is problematic for the item-encoding hypothesis which predicts that recall should be a simple negative function of the degree of perceptual variability. Finally, this experiment also supported the tenet of the perceptual-motor account that the ear-alternation effect is attributable in part to the mismapping

[5] This was quite deliberate because if we had used alternating voices in the context of an ear-alternating sequence, the ear and voice changes would have been perfectly correlated, leading to particularly strong segregation and hence, based on the perceptual-motor account, a larger disruptive effect on serial recall than when items alternated on only one dimension. As the item-encoding hypothesis would predict the same outcome, such an experiment would not have been diagnostic.

between perceptual organization and the assembly of a subvocal motor-plan: When such motor-planning is impeded via an articulatory suppression instruction, the ear-alternation effect is attenuated at pre-recency relative to recency. Moreover, the survival of the effect at recency under articulatory suppression dovetails with other evidence showing that recall of the last few items is supported by a direct use of the perceptual salience of the end-boundary of an auditory sequence rather than via motor recoding (Jones et al., 2004; Maidment & Macken, 2012; Nicholls & Jones, 2002).

The present findings are problematic for theoretical accounts that are conceptually allied to the classical centralist view of short-term memory in which perceptual variability affects item-encoding processes prior to a limited-capacity short-term memory structure or space (e.g., Mogensen et al., 2015; Nygaard et al., 1995; Treisman, 1971). Instead, they support the view that the perceptual organization of events and the compatibility of that organization with subvocal motor-planning is a key determinant of verbal serial short-term memory. In this view, the impact of perceptual alternation brings into relief how the representation of the serial order of a sequence of spoken items is in part a direct by-product of a process whose function it is to organize events into coherent objects, not memory storage per se. One way in which this passive form of organization determines serial recall is by influencing the ease with which items can be uploaded into an actively-assembled motor-plan. When the coherence of successive items in the typical auditory-verbal list is broken by making non-successive items more coherent than successive items, the creation of two perceptual objects each with its own integral serial organization generates a perceptual-motor mismapping problem, increasing the burden on active control and the likelihood of error.

Perceptual-motor incompatibility as a selection-for-action control problem

We argue that within the perceptual-motor framework, the active control process required in the face of perceptual-motor incompatibility may be understood as 'attention-switching' but where both what 'attention' is, and what needs to be switched between and why, is quite distinct conceptually from that typically found in the classical attention-switching account allied to the centralist view. Within the latter account, attention is often construed as a narrow filter designed to protect a limited-capacity processing or short-term storage space from overload (Broadbent, 1958; Lachter, Forster, & Ruthruff, 2004) or as a spotlight that holds a subset of information in working memory in a particularly accessible state (Cowan, 1995; Oberauer, 2002). In this view, attention switching is required when successive items change (e.g., in location or voice) precisely because attention is a narrow protective filter or spotlight that must therefore be shifted in real time if those items are to enter short-term memory (e.g., Page & Norris, 1998) or be made highly accessible (e.g., Cowan, 2001). Only then is item information (e.g., phonological content) linked to a (separate) representation of their order (e.g., Burgess & Hitch, 1999; Henson, 1998; Page & Norris, 1998). Thus, the attention switching on this account is the switching of a pre-short-term memory filter or spotlight between each discrete stimulus as (or before) it occurs and prior to the point at which item order is determined.

The perceptual-motor account, in contrast, is allied closely to a radically different conception in which attention is viewed as a *selection-for-action* (not for-memory) process (Allport, 1993; Hommel, 2010; Houghton & Tipper, 1996; Neumann, 1987, 1996; van der Heijden, 1992). In this view, attention is the process of mapping one of potentially several perceptual objects (derived preattentively in parallel) onto a motor-planning process that is limited simply on account of the fact that the bodily-action being simulated is intrinsically sequential (e.g., speaking, grasping a cup, delivering a right hook; Rosenbaum, 1991). In these terms, then, ear- (or voice-) alternation in serial recall creates a selection-for-action problem: Two distinct temporally-extended perceptual objects are available in parallel, each embodying serial order cues incompatible with the to-be-produced sequence. The selection-for-action process ('attention') must therefore switch between the two objects and each successive act of populating the motor-plan involves having to disembed an item from the object that contains it. Thus, in contrast to the centralist view, the attention-switching on this account is the switching between already ordered sequences for the purposes of (re)assembling the items into a task-appropriate motor-plan. It is interesting to note that Treisman (1971) also considered (but did not test) the possibility that the switching with an ear-alternating sequence is between already-sequenced sensory data: "If retrieval is from a "sensory tape-recording" it appears to retain its dichotic character" (p. 165) and that "on this hypothesis the distinction between perception and memory becomes blurred or meaningless" (pp. 165–166). The present study provides strong empirical support for this hypothesis.

The perceptual-motor mapping (or selection-for-action) account also helps make sense of findings from other early studies of the ear-alternation on serial recall. The view that the alternation effect is due to a two-to-one mapping problem predicts that if individuals had two vocal tracts, the problem would be reduced or eliminated. Moray and Jordan (1966) conducted what may be described as a manual analogue of this hypothetical double vocal-tract experiment: In the context of the split-span dichotic method in which a succession of simultaneous item-pairs are presented to the two ears (cf. Broadbent, 1954)—they provided participants with a means of parallel manual (two-hand) output to match the parallel (two-ear) input: Responses were typed on a 'Palantype' (Stenotype) keyboard which can register the output of two simultaneously pressed keys. With sufficient practice (given the unfamiliar output mode), participants were now, unlike with vocal or written output, able to reproduce the ear-alternating sequence so that it approximated the true temporal order of the items with a high degree of accuracy. This finding is entirely in line with the perceptual-motor mapping account: the difficulty of alternation (with a serial output mode) lies in motor-sequencing the items appropriately, not in encoding or storing item-level information as posited within the item-encoding hypothesis.

The central importance of 'Peripheral' auditory-specific effects in serial recall

A more conceptual potential challenge to our analysis from the standpoint of the classical, centralist, view of short-term memory is that the perceptual alternation phenomena that we have studied here and elsewhere (Hughes et al., 2009, 2011) are specific to *auditory* lists; indeed, the obligatory perceptual integration processes at the heart of our account of these particular phenomena are much more powerful in the case of, if not unique to, auditory sequences (Bregman, 1990). Proponents of the centralist view might therefore point to the fact that most of the benchmark phenomena of serial short-term memory (e.g., the phonological similarity effect; Baddeley, 1966; Conrad, 1964; the distinctive shape of the serial position curve; e.g., Crowder, 1972; the irrelevant sound effect; Colle & Welsh, 1976; Salamé & Baddeley, 1982; the articulatory suppression effect; Murray, 1968; the word-length effect; Baddeley et al., 1975) are observed with visual as well as auditory list-presentation, that is, they transcend sensory modality. Indeed, that this is the case is one of the main bases upon which rests the contention that short-term memory operates upon abstract-phonological, modality-independent, representations (see, e.g., Page & Norris, 1998). From this standpoint, even if the centralist approach was to concede that ear- and voice-alternation effects are underpinned by perceptual organization processes rather than item-encoding costs, they would remain auditory-perceptual effects, peripheral (both structurally and conceptually) to the abstract realm of short-term memory proper. However, on further empirical scrutiny, it turns out that the cross-modal generality of certain features and phenomena of serial recall is a very weak basis for concluding that short-term memory is supported by central modality-independent representations and mechanisms. We turn at this point to examine in detail some of this evidence and the centralist reasoning that it has engendered, as well as its recent re-evaluation, as it has a fundamental bearing not only on the breadth of implication that can be drawn from perceptual alternation effects but also on the veracity of a perceptual-motor account of short-term memory more generally as well as the veracity of defining centralist assumptions.

We turn first to the observation that serial recall performance is (apparently) very similar regardless of whether the list is presented visually or auditorily and that it diverges only at recency, where auditory list recall enjoys a distinct advantage (i.e., the modality effect; Conrad & Hull, 1968; Crowder & Morton, 1969). Such cross-modal generality at pre-recency has been taken classically to point to the modality-independent (e.g., phonological) character of verbal short-term storage. Indeed, the modality effect (or auditory recency) has, like perceptual variability effects, been cast as being peripheral to central short-term memory (e.g., Baddeley, 1986; Hurlstone et al., 2014; Page & Norris, 1998). However, on closer inspection, it turns out that most studies in which recall of visual lists has been compared with truly auditorily-presented lists (as opposed to visual lists read aloud by the participant; e.g., Conrad & Hull, 1968) show that visual list-recall enjoys an advantage at pre-recency (e.g., Frankish, 1989, 2008; Harvey & Beaman, 2007; Macken et al., in preparation; Maylor, Vousden, & Brown, 1999; Sjöblom & Hughes, in preparation). Thus, recall of visual and auditory lists in fact differs throughout the serial position curve, pointing to different processing constraints—and different ones for different parts of the list—related specifically to sensory-modality. In particular, Macken et al. (in preparation) recently showed that the 'inverted modality effect' at pre-recency is—as we have argued here in relation to perceptual alternation effects within audition—explicable by recourse to differences in the ease of perceptual-motor mapping. As argued in the present article, items in a typical auditory-verbal serial recall list (i.e., single voice, single location) are relatively well bound into a coherent stream as the result of obligatory perceptual organization processes. While this is advantageous for perceptual-motor mapping compared to when non-successive items are bound (e.g., under conditions of ear alternation), it nevertheless hampers that mapping process compared to the case with a list of visual items: A sequence of strongly bound auditory items would need to be segmented so as to assemble them into the motor-plan whereas the relative perceptual independence of successively presented visual items (with grouping across space, rather than time, dominating in vision; Bregman, 1990) is more conducive to optimal motor-plan assembly. That is, the looser binding of visual sequences leaves their boundaries less salient than for auditory sequences (see Bregman & Rudnicky, 1975) but at the same time allows for more facile addressing of each constituent individually, independent of the holistic object.

In support of the perceptual-motor account, the inverted modality effect, like the ear-alternation effect, is attenuated at pre-recency by articulatory suppression but the standard modality effect at recency, again like the ear-alternation effect, is immune to articulatory suppression due to the capacity for direct use of the perceptual salience of the list-end boundary. Impeding motor-planning by speeding up the presentation-rate also equates recall of visual and auditory items at pre-recency (Macken et al., in preparation). Thus, differences in recall of auditory and visual lists emerge not only at recency but also at pre-recency due to the different form that perceptual organization takes in the two sensory modalities and its differing implications for motor-plan assembly. That recall performance is not the same across modality at pre-recency removes a key pillar of support for the centralist view while being readily explicable by recourse to perceptual-motor mapping processes. It also means that serial recall phenomena that are specific to a particular modality—such as ear- and voice-alternation effects—are as central to the understanding of serial short-term memory as phenomena that are modality-general.

A perhaps more abiding example of the inference that cross-modal generality in serial recall points to an abstract basis for verbal short-term memory, however, relates to the phonological similarity effect, arguably the most influential phenomenon in the canon of short-term memory. This refers to the finding that items that sound similar (e.g., B G D C [...]) are much more difficult to recall in serial order than items that sound dissimilar (e.g., F H J Q [...]; Baddeley, 1966; Conrad, 1964). That this effect is observed when the lists are presented visually as well as auditorily has long been interpreted as evidence of a common phonological representation as the key unit of verbal short-term memory (e.g., Baddeley, 1986, 2007). However, an alternative interpretation is that what auditory- and visual-list recall have in common is that in each case such recall is supported by the assembly of items (at least at pre-recency) into a subvocal motor-plan; abstract-phonological representations distinct from motoric representations are unnecessary to explain such cross-modal generality. In this view, the 'phonological' similarity effect, regardless of modality of presentation, is an *articulatory*-similarity effect, reflecting speech-errors found occasionally in natural speech but that are exaggerated in the face of the laboratorial tongue-twister that is the phonologically-similar serial recall list (Acheson & MacDonald, 2009; Ellis, 1980; Jones et al., 2004, 2006, 2007; Maidment & Macken, 2012; Page et al., 2007).

The impact on short-term memory theory of the historical dismissal of an articulatory-based interpretation of the phonological similarity effect is difficult to overestimate: it is arguably the sole basis for the fractionation of a passive phonological short-term store from motor-planning processes (Baddeley, 1986, 2007, 2012). But that dismissal of an articulatory account has been shown to have been premature. The critical evidence in question is that the phonological similarity effect is still found even when the articulatory system is blocked via articulatory suppression as long as the items are presented auditorily and hence enjoy automatic access to passive phonological storage (Baddeley, 1986; Baddeley & Larsen, 2007; Baddeley et al., 1984). However, recent research has shown that articulatory motor-planning is indeed a precondition for the phonological similarity effect: For the majority of the list, the effect is eliminated or reduced dramatically by articulatory suppression regardless of modality (Jones et al., 2004; Sjöblom & Hughes, in preparation; see also Murray, 1968). The apparent phonological similarity effect with auditory lists under articulatory suppression is driven by processes operating *at recency* (Jones et al., 2004; Sjöblom & Hughes, in preparation), precisely the portion of the curve where recall is assumed to be supported by processes "peripheral to the working memory system" (Baddeley, 1986, p. 95; see also Page & Norris, 1998). That this residual 'phonological' similarity effect is indeed not phonological but instead underpinned by acoustic-based perceptual organization processes operating at the end-boundary of an auditory sequence is demonstrated by the fact that it is eliminated by the addition of a suffix that is perceptually grouped with the to-be-remembered list but not when that exact same suffix is grouped separately from it (Jones et al., 2004, 2006; Maidment & Macken, 2012). Thus, the 'phonological' similarity effect, while on the face of it pointing to a level of representation that transcends specific perceptual or motor processes, reflects the combined effects of articulatory similarity (at pre-recency) and acoustic similarity (at recency).

A striking convergence of findings is emerging, therefore, based on the study of numerous serial recall phenomena and their various interactions suggesting that serial recall performance is parasitic upon modality-specific perceptual and motor processes and the degree to which the passive organization that yields perceptual objects affords its conversion into a motoric-organization subserving overt sequential output. This list now includes the effect of phonological similarity (Jones et al., 2004, 2006, 2007; Maidment & Macken, 2012), modality and suffix effects (Macken et al., 2015; Maidment, Macken, & Jones, 2013; Nicholls & Jones, 2002), the impact of irrelevant sound (Hanley & Hayes, 2012; Hughes, Tremblay, & Jones, 2005; Jones et al., 2004), the effect of long-term linguistic knowledge (Macken et al., 2014) as well as ear- and voice-alternation effects (present study; Hughes et al., 2009, 2011). Thus, at the very least, this work suggests the need for greater emphasis on perceptual organization and motor-planning processes even within centralist frameworks (for examples of this more integrative approach, see, e.g., Page & Norris, 2003; Page et al., 2007). Indeed, the embedding of serial short-term memory performance within a selection-for-action view—which denies the need to posit limited-capacity processing structures to explain attentional selectivity (Hommel, 2010; Neumann, 1996)—would solve the paradox of a system specialized for short-term retention that is at the same time severely limited in capacity (e.g., Cowan, 2001): Performance-limits—and hence apparent structural limits—arise in short-term memory tasks because the processes that underpin that performance did not evolve for the purpose of remembering over the short-term (or for long-term learning; cf. Baddeley, Gathercole, & Papagno, 1998; Sjöblom & Hughes, in preparation) but for getting the body around safely and effectively; to parse the world into objects and to act upon them appropriately.

A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jml.2016.04.006.

References

Acheson, D. J., & MacDonald, M. C. (2009). Twisting tongues and memories: Explorations of the relationship between language production and verbal working memory. *Journal of Memory and Language, 60*(3), 329–350. http://dx.doi.org/10.1016/j.jml.2008.12.002.

Allport, D. A. (1993). Attention and control: Have we been asking the wrong questions? A critical review of 25 years. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligent, and cognitive neuroscience* (pp. 183–218). Cambridge, MA: MIT Press.

Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency modulated tones. *Journal of Experimental Psychology: Human Perception and Performance, 11*(3), 257–271. http://dx.doi.org/10.1037/0096-1523.11.3.257.

Baddeley, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology, 18*(4), 302–309. http://dx.doi.org/10.1080/14640746608400047.

Baddeley, A. D. (1986). *Working memory*.Oxford: Oxford University Press.
Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49*, 5–28.
Baddeley, A. D. (2007). *Working memory, thought and action*.New York, NY: Clarendon Press/Oxford University Press.
Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29. http://dx.doi.org/10.1146/annurev-psych-120710-100422.
Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173.
Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.). *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York, NY: Academic Press.
Baddeley, A. D., & Larsen, J. D. (2007). The phonological loop unmasked? A comment on the evidence for a "perceptual-gestural" alternative. *Quarterly Journal of Experimental Psychology, 60*(4), 497–504. http://dx.doi.org/10.1080/17470210601147572.
Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology, 36*(2), 233–252. http://dx.doi.org/10.1080/14640748408402157.
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*(6), 575–589. http://dx.doi.org/10.1016/S0022-5371(75)80045-4.
Barrouillet, P., & Camos, V. (2014). *Working memory: Loss and reconstruction*.London, England: Psychology Press.
Beauvois, M. W., & Meddis, R. (1997). Time decay of auditory stream biasing. *Perception & Psychophysics, 59*(1), 81–86. http://dx.doi.org/10.3758/BF03206850.
Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance, 4*(3), 380–387. http://dx.doi.org/10.1037/0096-1523.4.3.380.
Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organisation of sound*.Cambridge, MA: MIT Press.
Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89*(2), 244–249. http://dx.doi.org/10.1037/h0031163.
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. (1954). The role of auditory localization in attention and memory span. *Journal of Experimental Psychology: Human Perception and Performance, 47*(3), 191–196. http://dx.doi.org/10.1037/h0054182.
Broadbent, D. E. (1958). *Perception and communication*.Elmford, NY: Pergamon Press. http://dx.doi.org/10.1037/10037-000.
Broadbent, D. E. (1984). The maltese cross: A new simplistic model for memory. *Behavioral and Brain Sciences, 7*(1), 55–94. http://dx.doi.org/10.1017/S0140525X00026121.
Broadbent, D. E., & Gregory, M. (1964). Stimulus set and response set: The alternation of attention. *Quarterly Journal of Experimental Psychology, 16*(4), 309–317. http://dx.doi.org/10.1080/17470216408416386.
Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*(3), 551–581. http://dx.doi.org/10.1037/0033-295X.106.3.551.
Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55*(4), 627–652. http://dx.doi.org/10.1016/j.jml.2006.08.005.
Cherry, E. C., & Taylor, W. K. (1954). Some further experiments upon the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America, 26*, 554–559. http://dx.doi.org/10.1121/1.1907373.
Chomsky, N., & Halle, M. (1968). *The sound pattern of English*.New York, NY: Harper and Row.
Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(03), 181–204.
Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior, 15*(1), 17–31. http://dx.doi.org/10.1016/S0022-5371(76)90003-7.
Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55*(1), 75–84. http://dx.doi.org/10.1111/j.2044-8295.1964.tb00899.x.
Conrad, R., & Hull, A. J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science, 10*(4), 135–136.
Cowan, N. (1995). *Attention and memory: An integrated framework*.New York, NY: Oxford University Press.
Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87–185. http://dx.doi.org/10.1017/S0140525X01003922.
Cowan, N. (2015). George Miller's magical number of immediate memory in retrospect: Observations on the faltering progression of science. *Psychological Review, 122*(3), 536–541. http://dx.doi.org/10.1037/a0039035.
Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139174909.006.
Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684.
Crowder, R. G. (1972). Visual and auditory memory. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye* (pp. 251–276). Cambridge, MA: MIT Press.
Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics, 5*(6), 365–373. http://dx.doi.org/10.3758/BF03210660.
Culling, J. E., & Summerfield, Q. (1995). Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay. *Journal of the Acoustical Society of America, 98*(2), 785–797. http://dx.doi.org/10.1121/1.413571.
Darwin, C. J., & Hukin, R. W. (1999). Auditory objects of attention: The role of interaural time differences. *Journal of Experimental Psychology: Human Perception and Performance, 25*(3), 617–629. http://dx.doi.org/10.1037/0096-1523.25.3.617.
Deutsch, D. (1979). Binaural integration of melodic patterns. *Perception & Psychophysics, 25*(5), 399–405. http://dx.doi.org/10.3758/BF03199848.
Diehl, R. L., Souther, A. F., & Convis, C. L. (1980). Conditions on rate normalization in speech perception. *Perception & Psychophysics, 27*(5), 435–443. http://dx.doi.org/10.3758/BF03204461.
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*(5), 624–634. http://dx.doi.org/10.1016/S0022-5371(80)90672-6.
Frankish, C. R. (1989). Perceptual organization and precategorical acoustic storage. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*(3), 469–479. http://dx.doi.org/10.1037/0278-7393.15.3.469.
Frankish, C. (2008). Precategorical acoustic storage and the perception of speech. *Journal of Memory and Language, 58*(3), 815–836. http://dx.doi.org/10.1016/j.jml.2007.06.003.
Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on 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*(1), 72–78. http://dx.doi.org/10.3758/BF03198497.
Guérard, K., & Tremblay, S. (2011). When distractors and to-be-remembered items compete for the control of action: A new perspective on serial memory for spatial information. *Journal of Experimental Psychology: Human Perception and Performance, 37*(3), 834–843. http://dx.doi.org/10.1037/a0020561.
Hanley, J. R., & Hayes, A. (2012). The irrelevant sound effect under articulatory suppression: Is it a suffix effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(2), 482–487. http://dx.doi.org/10.1037/a0025600.
Harvey, A. J., & Beaman, C. P. (2007). Input and output modality effects in immediate serial recall. *Memory, 15*(7), 693–700. http://dx.doi.org/10.1080/09658210701644677.
Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36*(2), 73–137. http://dx.doi.org/10.1006/cogp.1998.0685.
Hitch, G. J., Flude, B., & Burgess, N. (2009). Slave to the rhythm: Experimental tests of a model for verbal short-term memory and long-term sequence learning. *Journal of Memory and Language, 61*, 97–111.
Hommel, B. (2010). Grounding attention in action control: The intentional control of selection. In B. Bruya (Ed.), *Effortless attention: A new perspective in the cognitive science of attention and action* (pp. 121–140). Cambridge, MA: MIT Press.
Houghton, G., & Tipper, S. P. (1996). Inhibitory mechanisms of neural and cognitive control: Applications to selective attention and sequential

action. *Brain and Cognition, 30*(1), 20–43. http://dx.doi.org/10.1006/brcg.1996.0003.

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*(2), 316–327. http://dx.doi.org/10.1037/0096-1523.31.2.316.

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, and Cognition, 35*(6), 1411–1425. http://dx.doi.org/10.1037/a0017008.

Hughes, R. W., Marsh, J. E., & Jones, D. M. (2011). Role of serial order in the impact of talker variability in short-term memory: Testing a perceptual organization-based account. *Memory & Cognition, 39*(8), 1435–1447. http://dx.doi.org/10.3758/s13421-011-0116-x.

Hughes, R. W., Tremblay, S., & Jones, D. M. (2005). Disruption by speech of serial short-term memory: The role of changing-state vowels. *Psychonomic Bulletin & Review, 12*(5), 886–890. http://dx.doi.org/10.3758/BF03196781.

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin, 140*(2), 339–373. http://dx.doi.org/10.1037/a0034221.

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., Alford, D., Bridges, A., Tremblay, S., & Macken, W. J. (1999). Organizational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 464–473.

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*(2), 265–281. http://dx.doi.org/10.1016/j.jml.2005.10.006.

Jones, D. M., Hughes, R. W., & Macken, W. J. (2007). The phonological store abandoned. *The Quarterly Journal of Experimental Psychology, 60*(4), 505–511. http://dx.doi.org/10.1080/17470210601147598.

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

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*(3), 656–674. http://dx.doi.org/10.1037/0278-7393.30.3.656.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science, 13*(4), 313–319. http://dx.doi.org/10.1111/1467-9280.00458.

Judd, T. (1979). Comments on Deutsch's musical scale illusion. *Perception & Psychophysics, 26*(1), 85–92.

Kidd, G. R. (1989). Articulatory-rate context effects in phoneme identification. *Journal of Experimental Psychology: Human Perception and Performance, 15*(4), 736–748. http://dx.doi.org/10.1037/0096-1523.15.4.736.

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review, 111*(4), 880–913. http://dx.doi.org/10.1037/0033-295X.111.4.880.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–131). New York, NY: Wiley.

LeCompte, D. C., & Watkins, M. J. (1993). Similarity as an organising principle in short-term memory. *Memory, 1*(1), 3–22.

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

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

Macken, B., Taylor, J., Kozlov, M., Hughes, R. W., & Jones, D. M. (2016). Memory as embodiment: The case of modality and short-term memory (in preparation).

Macken, W. J., Taylor, J. C., & Jones, D. M. (2014). Language and short-term memory: The role of perceptual-motor affordance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1257–1270. http://dx.doi.org/10.1037/a0036845.

Macken, B., Taylor, J. C., & Jones, D. M. (2015). Limitless capacity: A dynamic object-oriented approach to short-term memory. *Frontiers in Psychology, 6*, 293.

Maidment, D. W., & Macken, W. J. (2012). The ineluctable modality of the audible: Perceptual determinants of auditory verbal short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 989–997. http://dx.doi.org/10.1037/a0027884.

Maidment, D. W., Macken, W. J., & Jones, D. M. (2013). Modalities of memory: Is reading lips like hearing voices? *Cognition, 129*(3), 471–493. http://dx.doi.org/10.1016/j.cognition.2013.08.017.

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*(4), 676–684. http://dx.doi.org/10.1037/0278-7393.15.4.676.

Maylor, E. A., Vousden, J. I., & Brown, G. D. (1999). Adult age differences in short-term memory for serial order: Data and a model. *Psychology and Aging, 14*(4), 572–594. http://dx.doi.org/10.1037/0882-7974.14.4.572.

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. 2, pp. 419–492). New York, NY: Wiley.

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

Mogensen, C., Miller, L. M., & Roodenrys, S. (2015). Not so fast! Talker variability in serial recall at standard presentation rates. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 69*(1), 39–53. http://dx.doi.org/10.1037/cep0000044.

Mondor, T. A., & Bregman, A. S. (1994). Allocating attention to frequency regions. *Perception & Psychophysics, 56*(3), 268–276. http://dx.doi.org/10.3758/BF03209761.

Mondor, T. A., & Zatorre, R. J. (1995). Shifting and focusing auditory spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 21*(2), 387–409. http://dx.doi.org/10.1037/0096-1523.21.2.387.

Mondor, T. A., Zatorre, R. J., & Terrio, N. A. (1998). Constraints on the selection of auditory information. *Journal of Experimental Psychology: Human Perception and Performance, 24*(1), 66–79. http://dx.doi.org/10.1037/0096-1523.24.1.66.

Moray, N. (1960). Broadbent's filter theory: Postulate H and the problem of switching time. *Quarterly Journal of Experimental Psychology, 12*, 214–220. http://dx.doi.org/10.1080/17470216008416728.

Moray, N., & Jordan, A. (1966). Practice and compatibility in 2-channel short-term memory. *Psychonomic Science, 4*(12), 427–428.

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology, 78*(4), 679–684. http://dx.doi.org/10.1037/h0026641.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review, 7*(3), 403–423. http://dx.doi.org/10.3758/BF03214356.

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

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

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*(1), 12–28. http://dx.doi.org/10.1037/0278-7393.28.1.12.

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*(7), 989–1001. http://dx.doi.org/10.3758/BF03205458.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 411–421. http://dx.doi.org/10.1037/0278-7393.28.3.411.

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*(1), 49–64. http://dx.doi.org/10.1016/j.jml.2006.09.002.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*(4), 761–781. http://dx.doi.org/10.1037/0033-295X.105.4.761-781.

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

Pashler, H. E. (1998). *The psychology of attention*.Cambridge, MA: MIT Press.

Reisberg, D., Rappaport, I., & O'Shaughnessy, M. (1984). Limits of working memory: The digit digit-span. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(2), 203–221. http://dx.doi.org/10.1037/0278-7393.10.2.203.

Rogers, W. L., & Bregman, A. S. (1993). An experimental evaluation of three theories of auditory stream segregation. *Perception & Psychophysics, 53*(2), 179–189. http://dx.doi.org/10.3758/BF03211728.

Rosenbaum, D. A. (1991). *Human motor control*.San Diego, CA: Academic Press.

Rosenbaum, D. A., Cohen, R. G., Jax, S. A., Weiss, D. J., & van der Wel, R. (2007). The problem of serial order in behavior: Lashley's legacy. *Human Movement Science, 26*(4), 525–554. http://dx.doi.org/10.1016/j.humov.2007.04.001.

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21*(2), 150–164. http://dx.doi.org/10.1016/S0022-5371(82)90521-7.

Schubert, E., & Parker, C. (1956). Additions to Cherry's findings on switching speech between the two ears. *Journal of the Acoustical Society of America, 27*, 792–794. http://dx.doi.org/10.1121/1.1908042.

Sjöblom, A., & Hughes, R. W. (2016). Verbal sequence learning without a phonological store: A perceptual-motor approach (in preparation).

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

Stevens, K. N. (1960). Toward a model for speech recognition. *Journal of the Acoustical Society of America, 32*(1), 47–55. http://dx.doi.org/10.1121/1.1907874.

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

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

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

van der Heijden, A. H. C. (1992). *Selective attention in vision*.London, England: Routledge.

Warren, R. M., & Bashford, J. A. (1976). Auditory contralateral induction: An early stage in binaural processing. *Perception & Psychophysics, 20*(5), 380–386. http://dx.doi.org/10.3758/BF03199419.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–88). London, England: Kegan Paul, Trench, Trubner & Company (Original work published 1923).

Wilson, M., & Fox, G. (2007). Working memory for language is not special: Evidence for an articulatory loop for novel stimuli. *Psychonomic Bulletin & Review, 14*(3), 470–473. http://dx.doi.org/10.3758/BF03194091.

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

Woodward, A. J., Macken, W. J., & Jones, D. M. (2008). Linguistic familiarity in short-term memory: A role for (co-)articulatory fluency? *Journal of Memory and Language, 58*(1), 48–65. http://dx.doi.org/10.1016/j.jml.2007.07.002.