Original Articles

# Memory as embodiment: The case of modality and serial short-term memory ☆

CrossMark

Bill Macken [a,*], John C. Taylor [a], Michail D. Kozlov [a], Robert W. Hughes [b], Dylan M. Jones [a]

[a] School of Psychology, Cardiff University, United Kingdom
[b] Department of Psychology, Royal Holloway, University of London, United Kingdom

A B S T R A C T

Classical explanations for *the modality effect*—superior short-term serial recall of auditory compared to visual sequences—typically recur to privileged processing of information derived from auditory sources. Here we critically appraise such accounts, and re-evaluate the nature of the canonical empirical phenomena that have motivated them. Three experiments show that the standard account of modality in memory is untenable, since auditory superiority in recency is often accompanied by visual superiority in mid-list serial positions. We explain this simultaneous auditory and visual superiority by reference to the way in which perceptual objects are formed in the two modalities and how those objects are mapped to speech motor forms to support sequence maintenance and reproduction. Specifically, stronger obligatory object formation operating in the standard auditory form of sequence presentation compared to that for visual sequences leads both to enhanced addressability of information at the object boundaries and reduced addressability for that in the interior. Because standard visual presentation does not lead to such object formation, such sequences do not show the boundary advantage observed for auditory presentation, but neither do they suffer loss of addressability associated with object information, thereby affording more ready mapping of that information into a rehearsal cohort to support recall. We show that a range of factors that impede this perceptual-motor mapping eliminate visual superiority while leaving auditory superiority unaffected. We make a general case for viewing short-term memory as an embodied, perceptual-motor process.

## 1. Introduction

The cognitive approach to explaining behavior addresses itself, at heart, to processes involved with the generation and transformation of representations cleft both from the perceptual processes whereby the represented objects and events are transduced and the motor processes wherein their ultimate effects shape the actions of the organism. Constraints arising from perceptual and motor processes are typically cast as subsidiary to those operating at the core of the cognitive system. A classic instance of this relates to the role of modality of presentation in short-term memory performance, the investigation of which dates back to the origins of

the cognitive approach to short-term memory (e.g., Conrad & Hull, 1968; Crowder & Morton, 1969), and continues to form part of the empirical canon to which theorizing in short-term memory addresses itself (e.g., Burgess & Hitch, 2006; Grossberg & Pearson, 2008; Henson, 1998; Page & Norris, 1998).

The received narrative of the role of presentation modality in serial short-term memory is succinctly captured thus: "For short-term memory, auditory presentation is consistently superior to visual presentation, with the difference restricted to recently presented items." (Penney, 1975, p. 68). Similarly, the view is encapsulated from the outset of investigation of this *modality effect* in the 'idealized' serial position functions depicted by Crowder and Morton (1969, p. 366) in which visual and auditory serial position functions are identical for early and mid portions, with audition emerging superior towards the end. Nearly 50 years of theorizing about the basis of this effect has followed the classical cognitive scheme described above, involving the separation of perceptual processing from the core, modality-independent cognitive system supporting short-term serial memory (see e.g., Hurlstone, Hitch, & Baddeley, 2014). Here we revisit the role of modality in

short-term memory and find not only that its empirical character has been misconstrued, but that, despite the way in which it has been incorporated into mainstream cognitive theorizing, it poses a fundamental challenge to that way of explaining performance.

### 1.1. Classical approaches to the Modality Effect (ME)

The modality-specific aspect of performance in serial recall is most usually framed in terms of processes or representational forms which confer an advantage for auditory verbal over visual verbal information. An early approach invoked a bespoke limited-capacity store dedicated to the exclusive retention of acoustic input, the Precategorical Acoustic Store (PAS; Conrad & Hull, 1968; Crowder, 1970; Crowder & Morton, 1969). In this view, superior recall of auditory items stems from the fact that the PAS holds input in a more durable form than does a store containing precategorical representations of visual stimuli. At the point of recall, representations of pre-recency items within such precategorical memory stores (both visual and acoustic) will have decayed or have been overwritten by later items. However, recent auditory items enjoy a recall advantage over visual items due to the greater durability of PAS compared to the precategorical visual store. While specific aspects of the PAS account have fallen out of favor in the ensuing decades, the notion of a dedicated auditory input store still figures in many contemporary accounts of the role of modality (see e.g., Hurlstone et al., 2014).

Other theories eschew modality-specific stores but still invoke constructs in which auditory input is afforded a special status, enjoying either greater positional (Henson, 1998) or temporal (Glenberg & Swanson, 1986) resolution, or requiring less attention-dependent maintenance processes (Penney, 1989). Still other approaches assume that representations of auditory items are inherently richer than those of visual items with memory items represented as a mixture of modality-dependent (physical) features and modality-independent features (e.g., Nairne, 1990; Neath, 2000). Auditory items are assumed to have more modality-dependent features than visually presented items making them less prone to interference.

Despite the differences in these classical approaches, a major stumbling block for all of them, and the empirical starting point for our re-appraisal, is the rarely remarked-upon observation that the superiority of recall of auditory items at recency can be accompanied simultaneously by visual superiority at pre-recency: an *inverse modality effect* (IME; e.g., Beaman, 2002). The idea of intrinsically superior memory representations or processes for auditory (over visual) stimuli cannot easily explain such an effect. Although only commented upon relatively recently (Beaman, 2002), it transpires that there are many instances—although the picture is not universally consistent—in which evidence for an IME is present when an auditory-alone presentation (i.e., with no concurrent visual presentation) is contrasted with a visual-silent presentation (Baddeley & Larsen, 2007; Frankish, 1989, 2008; Harvey & Beaman, 2007; Jones, Macken, & Nicholls, 2004; Maylor, Vousden, & Brown, 1999; Penney & Blackwood, 1989; Routh, 1971; Tremblay, Parmentier, Guérard, Nicholls, & Jones, 2006). Table 1 gives a list of studies and their outcomes in which we simply note whether or not visual recall was superior to auditory recall in mid list serial position curves, since the relevant papers don't actually statistically test for such effects.

A further anomaly for the classical cognitive approach arises when a survey of the literature reveals that there are in fact relatively few studies in which audition and vision are compared directly. While the term 'modality effect' connotes a contrast between auditory and visual presentation, in many relevant studies, the 'auditory' items are *not* in fact *presented* auditorily. As shown in Table 1, the ostensible 'auditory' condition often comprises visually presented sequences that are read aloud simultaneously by the participant. For clarity, we propose that such conditions are better described as being *visual-vocalized*, rather than auditory. Arguably, the properties of simultaneously read and spoken material derive as much from the fact that they involve articulatory control processes as from their auditory characteristics. It is not unreasonable therefore to question the extent to which such a comparison is a reflection of modality as it is typically construed. Indeed, scrutiny of such studies poses further questions. Although visual-vocalized lists show enhanced recency relative to visual-silent lists (and so appear functionally similar to direct auditory presentation), it is not clear whether this is wholly or even partly due to the auditory properties of the setting. For example, Crowder (1970) compared the serial recall of visual-silent, visual-vocalized and bimodal (auditory and visual) digit sequences. Recall in recency was superior for both the vocalized and the bimodal conditions compared to the visual-silent condition, suggesting that superior recall in recency does indeed derive from the presence of an auditory signal. However, in another study (Crowder, 1986) comparing visual-silent lists with vocalized, whispered or mouthed lists, recency effects of equal magnitude were obtained for all three articulation condition (see also Greene & Crowder, 1984) all of which suggests that it may be misleading to ascribe recency effects in visual-vocalized sequences to the action of the auditory character of the input.

While not providing hard statistical evidence, since relevant tests were not conducted, Table 1 is suggestive of the historical presence of an IME in terms of numerically superior visual versus auditory recall in pre-recency. When a 'true' auditory vs. visual contrast is made, those cases in which IME is lacking (i.e., numerically equivalent performance in mid-list positions) may be due to features of the experimental design that deviate from the usual requirement for forward serial output of the whole sequence. For example, in the serial position curves reported by Drewnowski and Murdock (1980), where no IME was observed, the to-be-remembered sequence-length was varied on a trial-by-trial basis and participants were not asked for a fixed number of responses on each trial. Participants conceivably may have strategized, truncating their responses for the longer sequences. Empirically—and in addition to the absence of an IME in pre-recency—there were no significant recency effects for either auditory *or* visual sequences, as evaluated using a correct-in-position scoring regime (see Fig. 3 of Drewnowski & Murdock, 1980). In other words, no modality effect of any sort was observed using a strict serial recall criterion. Furthermore, in that study, presentation modality was manipulated between participants (see also, Corballis, 1966; Madigan, 1971), raising the possibility that early-list differences are additionally masked by group differences in overall performance.

Thus far, then, we witness foundational instability—both methodological and empirical—underlying the canonical constitution of the modality effect *qua* auditory superiority. More recently, to account for instances of visual superiority, Baddeley and Larsen (2007) proposed that the apparent visual advantage for pre-recency items derives from the opportunistic recruitment of additional visual (and not just phonological) codes, presumably via the visuo-spatial sketchpad (e.g., Baddeley, 2000), to assist with visual list maintenance. This would explain why visual items show an advantage over auditory items, even though the latter are assumed to enjoy direct access to an otherwise modality-independent phonological store (Baddeley, 2003) or to the PAS (Crowder & Morton, 1969). However, this explanation remains problematic or at least underspecified since it is not clear when, and to what extent, visual code recruitment is expected to counteract the supposed advantage for auditory items and why, for example, the effect of such visual codes is apparently restricted to pre-recency.

**Table 1**
Summary of studies investigating the effects of modality, indicating which presentation conditions were compared and indicating when evidence of an IME— numerically superior recall for visual over auditory presentation in medial portion of the serial position curve—was present in the reported data.

| Study | Conditions tested | | | | | | Evidence for IME |
|---|---|---|---|---|---|---|---|
| | Visual/silent | Visual/mouthed | Visual/whisper | Visual/vocal | Visual + auditory | Auditory | |
| Conrad and Hull (1968) | ✓ | | | ✓ | | | |
| Corballis (1966)[a] | ✓ | | | | | ✓ | |
| Crowder (1970) | ✓ | | | ✓ | ✓ | | |
| Crowder (1988) | ✓ | ✓ | ✓ | ✓ | | | |
| Drewnowski and Murdock (1980)[a,b] | ✓ | | | | | ✓ | |
| Frankish (1989) and Frankish (2008) | ✓ | | | | | ✓ | ✓ |
| Gathercole (1986) | ✓ | | | ✓ | | | |
| Greene and Crowder (1984) | ✓ | ✓ | | ✓ | | | |
| Harvey and Beaman (2007) | ✓ | | | | | ✓ | ✓ |
| Jones et al. (2004) | ✓ | | | | | ✓ | ✓ |
| Murray (1968) | ✓ | | | | | | |
| Madigan (1971)[a] | ✓ | | | | | ✓ | |
| Maylor et al. (1999) | ✓ | | | | | ✓ | ✓ |
| Nairne & Walters (1983) | ✓ | ✓ | | ✓ | | | |
| Penney and Blackwood (1989) | ✓ | | | | | ✓ | ✓ |
| Routh (1971) | ✓ | | | | | ✓ | ✓ |
| Tremblay et al. (2006) | ✓ | | | | | ✓ | ✓ |
| Turner et al. (1987) | ✓ | ✓ | | ✓ | | | |
| Turner, Scwartz, Clifton, and Engle (1994) | ✓ | ✓ | | | | | |
| Watkins, Watkins, and Crowder (1974) | ✓ | | | ✓ | | | |

[a] Modality effect only compared between subjects.
[b] Non-standard variant of the correct-in-position scoring method was used

The IME, therefore, poses a fundamental problem for the classical approach to the role of modality in memory: any account that explains the effect of modality by reference to inherent properties of the information derived from that modality—*by virtue of it being so derived*—cannot be applied consistently and coherently throughout the serial position function that has been the focus of so much theorizing about the nature of memory.

### 1.2. An embodied approach

Our approach to the question of modality in short-term memory derives from a radically different approach to short-term memory itself. Bluntly, we regard short-term memory as a perceptual-motor task setting, in the same way that, for example, the goal-directed manual apprehension and manipulation of a solid object may be regarded as a perceptual-motor task setting. In the latter case, the task involves processes that render object-oriented visual perceptual representations that may provide control programs for the manual interaction with the object in order to accomplish the task-specific goals. In the case of short-term serial recall, the object of concern is the sequence of verbal material presented for reproduction, and the motor system adopted for manipulation of this object is the articulatory control system involved in the production of speech. Overall performance in the setting is an outcome of perceptual and motor processes and the interactions between them. From this perspective, modality of presentation comes into play with respect to object formation processes as they operate in visual and auditory presentation, and how the perceptual representations so formed afford, to greater or lesser degrees, facile manipulation of those objects and their constituents in the speech motor system.

Key to this account are both the nature and the consequences of perceptual object formation. While objects are fundamental functional units for both vision and audition, there are important differences between modality in how they are formed. As a generalization, processes of auditory object formation play out with respect to the temporal dimension, where extended acoustic events are grouped together over time, on the basis of gestalt-like properties of similarity and continuity of frequency, timbre, rhythm, and so on (e.g., Bregman, 1990). Visual object formation can also be characterized in terms of gestalt grouping cues, but in this case, spatial extent provides the substrate (see e.g., Scholl, 2001). This distinction is important for our account of modality effects in serial recall. Auditory presentation of a sequence – emanating from a single spatial location at a regular rate, in a spectrally consistent voice – means that there is a strong tendency for that sequence to form a coherent object. On the other hand, the forces of object formation are considerably weaker for the corresponding visual presentation where successive visual events are presented discretely over time in the same spatial location.

This difference in the way in which object formation plays out in the different modalities means that the consequences of object formation impact differently depending on modality. The consequence of importance here is that the fate of the nominal content of an object is determined by its being incorporated into an object; specifically, given the key role that boundary (or contour) processing plays in object formation (e.g., Wagemans et al., 2012), content that resides at or near the boundary of the object is relatively highly resolved, while content in the interior is less so. So, in constituting the boundary of an object, information acquires perceptual salience, making that information readily addressable, whereas the strong binding of individual list items within an object means that they lose salience and their individual identity becomes less addressable. Critically for our account, these consequences stem from object formation, not from the particular modality of presentation, per se.

Such functional consequences of auditory object formation are demonstrated in a range of settings involving the processing of sequences of auditory events, verbal or otherwise. For example, while participants are able to make judgements about rapid sequences of sounds (e.g., a tone, a click, a vowel, a buzz, at a rate of less than 200 ms per sound) in terms of whether successive sequences contain those sounds in the same or a different order, the ability to actually report the order of individual elements, or to identify which if any elements have changed order, only emerges at slower rates of presentation, such as would weaken the tendency for those sounds to cohere into a single object, concomitantly making them more addressable with respect to, for example, verbal labelling (see e.g., Warren, 1999 for an overview). Similarly, the ability to judge whether the order of a pair of tones differing in frequency is the same or different on two presentations

is impeded by the presence of single flanker tones, in a similar frequency range to the target tones, immediately preceding and succeeding those target tones. That this impact is due to the binding of those target tones into an object with the flanker tones, thereby reducing their individual addressability (rather than, for example, some sort of pro- or retroactive masking), is demonstrated by the fact that the addition of further sequences of tones preceding and succeeding the flanker tones at the same frequency actually restores performance, by capturing the flanker tones into a different object, thereby perceptually isolating again the target tones (Bregman & Rudnicky, 1975).

The lineaments of perceptual object formation can similarly be observed in more typical verbal short-term memory settings. For example, in relatively short (i.e., lasting less than 5 or 6 s) verbal sequences, an advantage for auditory over visual verbal presentation is evident especially at initial and terminal boundaries – i.e., at primacy and recency – while in longer sequences, the auditory advantage tends to be restricted to the recency portion (i.e., the classical ME) (see e.g., Jones, Hughes, & Macken, 2006; Macken, Taylor, & Jones, 2014; Maidment & Macken, 2012). That this advantage is due the stronger object formation processes – and therefore increased boundary salience – in the auditory compared to visual presentations is indicated by the way in which redundant prefixes and suffixes impact on the pattern of serial recall. The occurrence of a redundant end of list suffix eliminates or attenuates the ME (e.g., Crowder & Morton, 1969; Nicholls & Jones, 2002), however a redundant prefix also reduces the auditory advantage in those cases when it appears in primacy in shorter sequences (Jones et al., 2006). Furthermore, in both cases, prefixes and suffixes have their effect to the extent that they are perceptually incorporated into the auditory object corresponding to the memory sequence, thereby displacing the initial and terminal items from their privileged boundary positions; the addition of further redundant, task irrelevant auditory material that serves to 'capture' the suffix or the prefix into a separate object (in a manner analogous to the Bregman & Rudnicky, 1975, findings described above), restores serial recall performance for those initial and/or terminal items by, we argue, restoring those items to their boundary position within the auditory object corresponding to the to-be-remembered sequence. In one illustration (see Nicholls & Jones, 2002), the addition of the redundant (i.e., not to be recalled) spoken word 'go' at the end of a random sequence of to-be-recalled digits eliminated the ME. However, the further addition of a concurrent sequence of the spoken word 'go', in such a way that the suffix is incorporated into the object corresponding to that sequence, rather than the memory sequence, restores the auditory advantage in recency, even though the original suffix is still in exactly the same temporal and spectral relation to the end of the memory sequence (see also Maidment & Macken, 2012).

The typical form of auditory presentation in serial recall, then, affords stronger obligatory object formation than does the typical form of sequential visual presentation. The consequences of such object formation are that information at or near boundaries is well-resolved and readily addressable, while information in the interior of the object is less well resolved and the identity of the individual constituents – the list items – becomes less addressable. Such effects are not restricted to the formation of auditory objects, since analogous outcomes are observed in visual object formation where information in the interior of visual objects loses spatial resolution compared to information at or near the boundaries (e.g., Katshu & D'Avossa, 2014; Manassi, Sayim, & Herzog, 2012). On the other hand, because the typical form of presentation for visual-verbal serial recall does not lead to the sequence forming a coherent object, not only does such presentation not lead to the type of boundary salience that is evidenced in auditory recency, neither do items within the sequence lose their individual

addressability which would occur due to being bound into a single object corresponding to the whole sequence. Again, the critical point here is that object formation – rather than modality, per se – determines how the list content is represented.

These outcomes of object formation then enter into the process of subvocal rehearsal that underlies aspects of performance in serial recall. This involves the cumulative assembly of a rehearsal cohort incorporating successive items as the sequence unfolds, a process that we conceive of here as the mapping of the perceptual form onto the motor control processes which allow for the manipulation – i.e., the maintenance and reproduction in whatever form – of the sequence. Importantly, this assembly is subject to the real time constraints involved in subvocal motor processing (see e.g., Taylor, Macken, & Jones, 2015; Warren, 1999) so that assembly and iterative maintenance of a progressively longer motor control programme comes into conflict with the process of perceptual-motor mapping as the sequence unfolds. For this reason, subvocal rehearsal is most effective in supporting serial recall of the earlier parts of the sequence. While such processing normally takes place covertly in a serial recall setting, these inferences are supported by the findings of just such a pattern of behavior when overt rehearsal is required (Tan & Ward, 2000; Ward, 2002).

However, the efficacy of this process of mapping from perceptual to motor form is influenced not only by the real-time constraints on motor processing but also by the extent to which the perceptual form accords with the task-specific requirement to reproduce the sequence in terms of its serial constituents. For example, an auditory sequence that alternates on successive items from one voice to another is less well recalled than a sequence presented in a single voice, and this detriment arises because the alternating presentation leads to the formation of two perceptual objects, one corresponding to each of the voices, neither of which corresponds to the form of sequence required by the task (i.e., each item in its original successive order) (Hughes, Marsh, & Jones, 2009, 2011). The relevance of this here is that because the items bound within an object lose addressability with respect to their individual identity, the incorporation of that item information into the rehearsal cohort is compromised compared to a situation in which the items are not so bound. Specifically, given the argument about differences in object formation for visual and auditory sequences presented above, visual presentation will afford more ready incorporation into the rehearsal cohort of the sort of detailed item-by-item sequential articulatory specification that best supports serial recall performance (e.g., Acheson & MacDonald, 2009; Macken & Jones, 1995; Taylor et al., 2015; Woodward, Macken, & Jones, 2008) and therefore lead to an advantage with visual presentation for those items whose recall is sustained by subvocal rehearsal. As such, the strong tendency for the auditory sequence to form a coherent object leads via the same process to both *enhanced* performance for boundary information and *reduced* performance for interior information compared to visual presentation, while the weaker such tendency for visual presentation has precisely the opposite consequences.

An account based on the interplay of object formation and perceptual-motor mapping processes has the potential, then, to account for both the ME and IME in a coherent way. We also propose that such a framework can account for the impact of the requirement to vocalize and mouth list items on presentation, factors that have been shown to increase recency (e.g., Greene & Crowder, 1984) as well as often reducing performance overall with respect to control conditions (e.g., Arenberg, 1968; Crowder, 1970, 1986; Greene & Crowder, 1984). On the one hand the requirement to articulate list items as they are presented is likely to disrupt the free assembly of a cumulative rehearsal cohort that supports performance under control conditions. Notably, in this respect, our account therefore predicts that the IME is less likely to occur under

the requirement to vocalize or mouth, if that effect resides in the ready incorporation of successive list items into such a rehearsal cohort in the earlier part of list presentation for visual presentation. At the same time, the requirement to incorporate each list item means that the participant has to convert the whole list into a motor object, but this motor object will be critically different from the one corresponding to the rehearsal cohort assembled under normal conditions. Rather than being cumulative and iterative, it will instead form a motor object that is an analogue of the list and therefore it will constitute an object bounded by initial and terminal points corresponding to those of the list, with the concomitant salience and recall advantage conferred for the information at those boundaries. On the other hand, not only does the cumulative nature of unconstrained rehearsal mean that the terminal boundary of an object so formed is constantly updated, given that such rehearsal cohort formation might only be deployed strategically for the early part of the list (see e.g., Grenfell-Essam, Ward, & Tan, 2013; Tan & Ward, 2000; Ward, 2002), the terminal item might never even enter into a motor object, and so would only benefit if perceptual object formation processes were operating on the list, that is, for auditory but not visual presentation. In the experiments that follow, we test this embodied account by manipulating a range of factors that may be expected to impact on the object formation and perceptual-motor mapping processes that we propose give rise to the impact of modality on serial recall.

## 2. Experiment 1

Given the sort of methodological and empirical variability, described above within investigations of the role of modality in memory we begin by establishing within a single experiment the pattern of performance across the serial position curve associated with 'pure' auditory and visual presentation, as well as both silently mouthed and vocalized articulation of each successive list item on its presentation.

### 2.1. Method

#### 2.1.1. Participants

Twenty-two[1] Cardiff University Psychology undergraduates (19 female) aged 18–26 years (Mean: 19.5 years) participated in exchange for course credit. All had normal or corrected-to-normal vision and hearing. Ethical approval for all experiments reported here was received from the Cardiff University, School of Psychology Ethics Committee in accordance with British Psychological Society ethics guidelines.

#### 2.1.2. Materials

The stimuli were random permutations of seven consonants (R, X, H, Y, L, Q, K) presented either visually or auditorily. No items were repeated within a sequence and each sequence was unique. Visual stimuli were presented in 60 point *Arial* font. Auditory stimuli were recorded in a female voice (16-bit, 48 kHz) using a condenser microphone and *Audacity* (v. 1.3.12) audio workstation software (http://audacity.sourceforge.net) and digitally edited to 250 ms in length.

#### 2.1.3. Design

A 4 (presentation mode: auditory, visual-silent, visual-mouthed; visual-vocalized) × 7 (serial position), within-participant design was employed. Presentation mode was blocked

and block-order was randomized across participants. Each block comprised 30 trials, preceded by two practice trials. In the auditory condition, to-be-remembered sequences were presented via headphones. In the visual conditions, stimuli were presented centrally on a computer monitor. In the visual-vocalized condition, participants were instructed to read the items aloud as they were presented. In the visual–mouthed condition, participants mouthed the sequence items silently. In the visual-silent condition, participants were instructed to read the items silently. With the permission of the participants, compliance with instructions relating to the visual-vocalized and visual-mouthed conditions was monitored, respectively, via a sound and video link.

#### 2.1.4. Procedure

The experiment was conducted in a sound-attenuating booth. Prior to commencement, participants were told that their task would be to remember the order in which seven letters were presented. At the beginning of each block, participants were informed as to the modality of the stimuli in the upcoming trials (auditory or visual) and whether or not there was a requirement to vocalize or mouth the stimuli during visual presentation. Each trial began with a blank screen (1 s) followed by sequential presentation of the stimuli (250 ms duration, 750 ms ISI). At the end of each trial, the seven letters were re-presented on screen in a random permutation. Participants were instructed to use the mouse-pointer to click the letters in the order in which they had been presented. As each letter was selected, it disappeared from the array of available letters and was added to the reconstructed sequence. Each item could only be selected once and all items had to be selected before the next trial was initiated. Each trial commenced automatically on completion of the previous trial. The duration of the experiment was approximately 60 min.

### 2.2. Results and discussion

Serial position curves for each presentation-mode are shown in Fig. 1. Several patterns can be identified in these data. Comparison between the visual-silent and auditory conditions reveals the presence both of an ME in recency (auditory superiority) and an IME in mediacy (visual superiority). Further, both the visual-vocalized and visual-mouthed conditions reduce performance in pre-recency (compared to both the auditory and visual-silent conditions) while exhibiting strong recency effects (relative to the visual-silent condition), of a similar magnitude to that seen for auditory presentation.

These impressions were confirmed statistically. A 4 (presentation mode) × 7 (serial position) repeated measures ANOVA revealed significant main effects of serial position, $F(6,126)$ = 41.48, $p < 0.001$, $\eta_p^2 = 0.66$ and presentation mode, $F(3,63)$ = 10.58, $p < 0.001$, $\eta_p^2 = 0.35$, and a significant interaction, $F(18,378) = 11.69$, $p < 0.001$, $\eta_p^2 = 0.36$, indicating different effects of presentation mode across serial position.

In order to examine these different effects, we separately contrasted visual-silent with the other three presentation modes. Paired simple effects comparisons between the auditory and visual-silent modes confirm the impression given by Fig. 1, that while there was superior performance for auditory items in recency, overall, there was no significant difference between auditory and visual-silent performance, $F(1,21) = 2.63$, $p = 0.12$, $\eta_p^2 = 0.11$. A significant presentation mode by serial position interaction, $F(6,126) = 18.99$, $p < 0.001$, $\eta_p^2 = 0.48$, indicates that the ME in recency is offset by the IME in pre-recency with significantly better visual compared to auditory recall at serial position 4, $t$ $(21) = 2.85$, $p = 0.010$, and auditory superiority at serial positions 6 and 7, $t(21) = 2.57$, $p = 0.018$; $t(21) = 7.06$, $p < 0.001$, respectively

---

[1] Sample sizes used in the experiments reported here are typical of those that have robustly revealed the modality effect in the historical literature, and therefore provide a standard for testing the robustness and limits of the inverse modality effect.
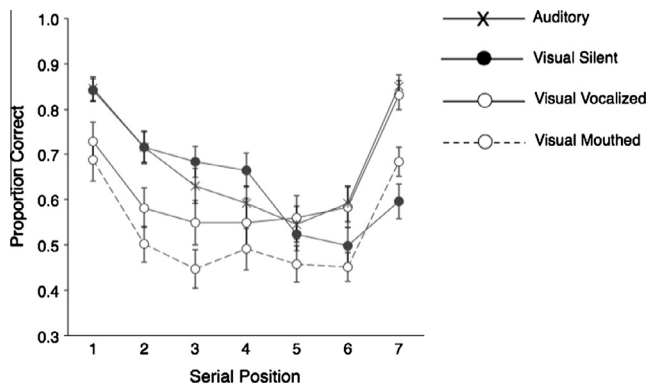
**Fig. 1.** Mean proportion correct scores for the serial recall of seven-item sequences under the presentation modes employed in Experiment 1. Error bars denote *SE*.

(nonsignificant *p* values for serial positions 1, 2, 3, and 5 = 0.910. 0.965, 0.117 and 0.485 respectively).

Comparing visual-vocalized and visual-silent modes (i.e., the effect of vocalization), the main effect of serial position was significant, $F(6, 126) = 26.08$, $p < 0.001$, $\eta_p^2 = 0.55$, with no main effect of presentation mode, $F(1, 21) = 1.07$, $p = 0.21$, $\eta_p^2 = 0.05$, and a significant interaction, $F(6, 126) = 24.04$, $p < 0.001$, $\eta_p^2 = 0.53$. Thus, relative to the visual-silent condition, the large recency effect observed in the visual-vocalized presentation mode (Fig. 1) is again offset in pre-recency. Vocalization therefore impedes performance in pre-recency while enhancing it in recency.

Finally, the comparison between the visual-mouthed and visual-silent conditions (i.e., the effect of articulation, in the absence of vocalization) revealed significant main effects of both presentation mode (silent > vocalized), $F(1, 21) = 12.88$, $p = 0.002$, $\eta_p^2 = 0.38$ and of serial position, $F(6, 126) = 35.56$, $p < 0.001$, $\eta_p^2 = 0.63$, and a significant interaction, $F(6, 126) = 13.86$, $p < 0.001$, $\eta_p^2 = 0.39$. Therefore, even in the absence of auditory feedback, the act of articulation serves to boost recency while impeding recall performance in pre-recency (relative to visual-silent presentation in both cases).

Although performance on visual-mouthed lists appears to be generally inferior to the remaining conditions, Fig. 1 suggests that mouthing the list enhances recency to the same extent as listening to items or vocalizing visual items. To corroborate this, a further ANOVA was carried out, to assess the effect of presentation mode on recency, defined as the difference between performance on the last item in a list and the average performance on the remaining items (Greene & Crowder, 1984). There was a significant main effect of presentation mode, $F(3, 63) = 22.33$, $p < 0.001$, $\eta_p^2 = 0.52$. but apart from visual-silent leading to reduced recency compared to visual-mouthed, visual-vocalized and auditory modes, $t(21) = 4.59$, $p < 0.001$, $t(21) = 5.7$, $p < 0.001$, and $t(21) = 6.01$, $p < 0.001$, respectively, the other presentation modes were undifferentiated in recency, $F(2, 42) = 1.99$, $p = 0.15$, $\eta_p^2 = 0.09$. To summarize, the effect of modality is neutral in primacy, negative (visual silent > auditory) in medial sequence locations and positive (auditory > visual silent) in recency. In contrast, both vocalization and articulation exert negative effects in primacy (visual-vocalized < visual-silent; visual-mouthed < visual-silent) and positive effects in recency (visual-vocalized > visual-silent, visual-mouthed > visual-silent).

The results of Experiment 1, then provide statistical evidence of the existence of the IME, supplementing the more impressionistic evidence gleaned from the historical review presented above. Recall performance at medial sequence locations in visual-silent sequences was higher than all other presentation modes, critically, including auditory. This IME was abolished by both vocalization and silent mouthing of visual sequences which reduced recall in

early and medial parts of the list while at the same time leading to equivalent recency effects to that of auditory presentation. The abolition of the IME in the visual-vocalized and visual-mouthed conditions undermines accounts invoking the recruitment of visual codes (e.g., Baddeley & Larsen, 2007; Beaman, 2002) since such codes should be available in for visual-silent and visual-vocalized presentation. We propose instead that the IME arises due to the more facile perceptual-motor mapping involved in incorporation of visually-presented list items into a rehearsal cohort, under standard visual presentation conditions, due to their being less strongly bound into a coherent perceptual object compared to their auditory counterparts. Disrupting this perceptual-motor mapping eliminates the IME.

The same constraints that impede recall of visual-vocalized and visual-mouthed sequences at pre-recency enhance recall later in the sequence: The emergence of a large recency effect in both these conditions is in line with our expectation that the obligatory articulation through to the end of the sequence serves to emphasize the salience of the sequence endpoint via motor object formation processes analogous to those found with auditory perceptual object formation. Finally, the equivalent effects on recency of silent mouthing and overt vocalization of visual sequences indicates an effect of vocalization that is not due to concomitant auditory input.

Effects of modality, then, are malleable and heterogeneously determined. Overall we have shown that the IME found when comparing standard visual and auditory presentation can be abolished by impeding the assembly of a rehearsal cohort via forced-pace articulation of list items, while at the same time, the auditory advantage over visual presentation in recency can be abolished by requiring articulatory embodiment of the whole sequence in a form corresponding to its presentation order. The equivalent recency effect under mouthed and vocalized articulation points to a motor, rather than perceptual, basis for recency effects for visually presented sequences. We explore this proposal further in Experiment 2.

## 3. Experiment 2

If, as we have claimed, the pattern of modality effects emerges not from inherent coding differences between auditory and visual information, but from the combined and distinct operation of object formation processes operating on the sequence and perceptual-motor mapping processes involved in rehearsal cohort formation, any task that interferes with just one of these processes should abolish only 'modality' dependent recall differences that are attributable to that process. Articulatory suppression has been widely used as a means of impeding subvocal rehearsal (e.g., Baddeley, 1986; Baddeley, Lewis, & Vallar, 1984; Macken & Jones, 1995; Nairne, 1990) and as such represents an ideal candidate for probing the articulatory determinants of modality effects which we are proposing reside in the perceptual-motor mapping process. In Experiment 2, we compared visual and auditory serial recall with and without articulatory suppression. As with overt vocalization and silent mouthing, we predict that articulatory suppression will abolish the IME by impeding rehearsal cohort formation. However, we predict that the ME will be left unaffected by articulatory suppression since, as argued above, this effect is due to obligatory object formation processes operating on the auditory sequences that do not obtain for visual presentation.

### 3.1. Method

#### 3.1.1. Participants

Twenty-four Cardiff University Psychology undergraduates, (20 female) were recruited as described for Experiment 1.

### 3.1.2. Materials

The materials were identical to those used in Experiment 1.

### 3.1.3. Design

A 2 (modality: auditory, visual) × 2 (articulatory suppression: control, suppression) × 7 (serial position), within-participant design was employed. Modality and articulatory suppression conditions were blocked and block order was counterbalanced across participants. Each block comprised 16 trials, preceded by two practice trials. As in Experiment 1, to-be-remembered auditory sequences were presented via headphones while visual stimuli were presented centrally on a computer monitor.

### 3.1.4. Procedure

The procedure was the same as for Experiment 1 except for the following details: At the beginning of each block, participants were informed as to the modality of the stimuli in the upcoming trials (auditory or visual) and whether or not there was a requirement to engage in articulatory suppression. In the control condition, participants were instructed to attend to the stimuli silently. In the suppression condition, participants were to begin whispering the number sequence '8, 9, 10' at a rate of 3 items/s as soon as the instruction screen appeared. A fixation cross was then presented for 1 s followed by sequential presentation of the to-be-remembered stimuli. In the suppression condition, participants were required to continue suppressing until the offset of the last to-be-remembered item. With their permission, participants were monitored by the experimenter via a microphone relayed from the testing booth in order to ensure compliance with the articulatory suppression instruction.

At the end of each trial, the seven letters were re-presented on screen in a random permutation along with an additional 'don't know' response option, in the form of a question mark [?]. Participants were instructed to use the mouse-pointer to click the letters in the order in which they had been presented as described in Experiment 1. The procedure lasted approximately 30 min.

### 3.2. Results and discussion

Serial position curves for the four conditions of Experiment 2 are shown in Fig. 2. Under control conditions, the pattern of performance replicates that of Experiment 1, with an advantage for auditory over visual items in recency (i.e., the ME) contrasting with an opposing advantage for visual items at medial sequence positions (i.e., the IME). Crucially, articulatory suppression abolishes the visual advantage in mediacy, without abolishing the auditory advantage in recency. These impressions were confirmed statistically with a 2 (modality) × 2 (suppression) × 7 (serial position) repeated measures ANOVA. The main effects of modality, articulatory suppression and serial position were all significant: modality (auditory > visual), $F(1,23) = 6.83$, $p = 0.016$, $\eta_p^2 = 0.23$; articulatory suppression (control > suppression), $F(1,23) = 124.3$, $p < 0.001$, $\eta_p^2 = 0.84$; and serial position, $F(6,138) = 58.63$, $p < 0.001$, $\eta_p^2 = 0.72$.

The crucial three-way interaction of modality by articulatory suppression by serial position was significant, $F(6,138) = 2.67$, $p = 0.018$, $\eta_p^2 = 0.10$ confirming that articulatory suppression modulates the effect of modality on recall performance but not uniformly across serial positions. Specifically, the recall advantage for visually presented sequences in pre-recency is abolished by articulatory suppression, whereas the advantage for auditory sequences in recency is unaffected. The interaction between modality and suppression across serial position is illustrated in Fig. 3, where the effect of modality – the difference between auditory and visual presentation – is plotted for each serial position. The effect of modality on recall diverges markedly in medial
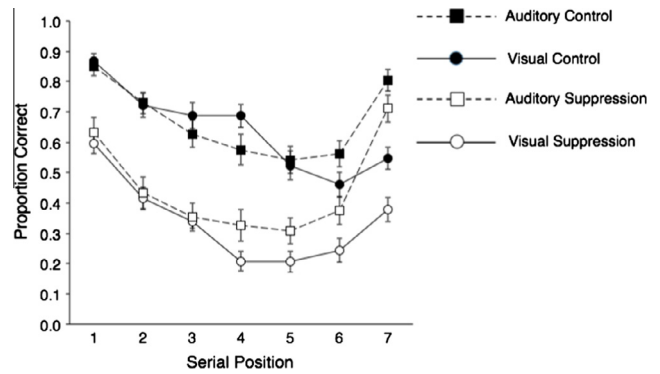
**Fig. 2.** Mean proportion correct scores for the serial recall of seven-item visual and auditory sequences with and without articulatory suppression (Experiment 2). Error bars denote SE.

sequence positions, such that the negative deflection in the control condition (corresponding to the IME) is completely abolished—and indeed reversed—under articulatory suppression. In order to confirm this statistically, paired *t*-tests were undertaken, comparing the effect of modality at each serial position.

Only the comparison at serial position 4 was significant, $t(23) = 3.65$, $p = 0.001$, confirming both the abolition of the IME and the survival of the ME under articulatory suppression.

Articulatory suppression, then, abolishes the IME while leaving the ME unaffected, again supporting the idea that the IME is articulatory in origin and that the mid-list recall advantage found with visually presented sequences arises as a result of the opportunity for greater facility in the perceptual-motor mapping process involved with rehearsal cohort formation for visual compared to auditory lists. Furthermore, the survival of the ME under articulatory suppression indicates that it instead results from the obligatory object formation processes operating on the auditory sequence, (e.g., Jones et al., 2004; Maidment & Macken, 2012; Nicholls & Jones, 2002) bolstering the claim made in Experiment 1 that modality effects are functionally as well as mechanistically heterogeneous.

## 4. Experiment 3

If articulatory suppression eliminates the IME, as we have claimed, by impeding the real time assembly of a rehearsal cohort during list presentation, then the same effect should be achieved by introducing constraints on the time available to implement such rehearsal cohort assembly. In Experiment 3, we test this by manipulating presentation rate, both halving and doubling it relative to the 750 ms used in Experiments 1 and 2.

### 4.1. Method

#### 4.1.1. Participants

Thirty-six Cardiff University Psychology undergraduates, (32 female) were recruited as escribed in Experiment 1.

#### 4.1.2. Materials

The materials were identical to those used in Experiment 1. Only ISI (offset to onset) was changed in order to manipulate rate.

#### 4.1.3. Design

A 2 (modality: auditory, visual) × 3 (Rate: 375 ms; 750 ms; 1500 ms, onset-onset) × 7 (serial position), within-participant design was employed. Conditions were blocked and block-order was counterbalanced across participants. All participants per-
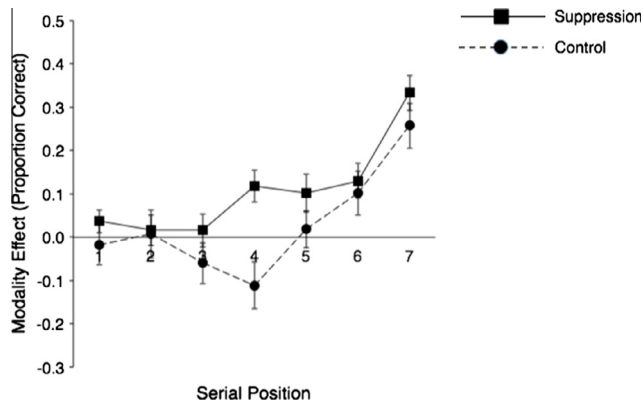
**Fig. 3.** The effect of modality (auditory minus visual presentation) on serial recall with and without articulatory suppression (Experiment 2). Error bars denote *SE.*

formed six blocks. Each block comprised 16 trials, preceded by two practice trials.

### 4.1.4. Procedure

Apart from the rate manipulation, the procedure was the same as for the control conditions of Experiment 2 and lasted approximately 45 min.

### 4.2. Results and discussion

Serial position curves for the six conditions of Experiment 3 are shown in Fig. 4. Two patterns are apparent: First, the IME is affected by stimulus presentation rate. The performance advantage for visual stimuli in pre-recency at a rate of 750 ms/item (Fig. 4b. cf. Experiments 1 and 2) is abolished for the faster rate of 375 ms/item (Fig. 4a), and when the presentation rate is halved to 1500 ms/item, the IME is still present although it appears attenuated in its extent (Fig. 4c). Secondly, the ME in recency is unaffected by the rate manipulation. Data were initially subjected to a 3 (rate) × 2 (modality) × 7 (serial position), repeated measures ANOVA. The main effect of serial position was significant, $F(1, 33) = 94.3$, $p < 0.001$, $\eta_p^2 = 0.74$. The main effects of rate and modality were not significant, $F(2, 66) = 0.93$, $p = 0.40$, $\eta_p^2 = 0.03$ and $F(6, 198) = 0.01$, $p = 0.91$, $\eta_p^2 < 0.01$, respectively. The two-way interactions with serial position were significant: rate × serial position, $F(12, 396) = 3.40$, $p < 0.001$, $\eta_p^2 = 0.09$ and modality × serial position, $F(6, 198) = 20.4$, $p < 0.001$, $\eta_p^2 = 0.38$. However, the two-way interaction between modality and rate and the three-way interaction were not significant $F(2, 66) = 0.88$, $p = 0.42$, $\eta_p^2 = 0.03$ and $F(12, 396) = 0.35$, $p = 0.98$, $\eta_p^2 = 0.01$, respectively.

To test our specific predictions, pairwise comparisons were conducted across the serial position curve. There was no effect of modality at serial position 1 for any of the rate conditions (all $p$s > 0.05). Conversely, the usual auditory superiority was observed at serial position 7 in all conditions: 375 ms, $t(33) = 4.94$, $p < 0.001$; at 750 ms, $t(33) = 2.95$, $p = 0.06$; and at 1500 ms, $t(33) = 4.29$, $p < 0.001$. As such, the ME is immune to the rate manipulation. However, in medial positions, paired $t$-tests comparing auditory to visual recall performance at serial positions 2–6 reveals the influence of rate of presentation on the IME. At 375 ms/item, there was no effect of modality on performance in these positions (all $p$s > 0.05). However, at 750 ms/item (as in Experiments 1 and 2), visual presentation was superior to auditory at serial position 3 and 4, $t(33) = 2.12$, $p = 0.042$ and $t(33) = 2.39$, $p = 0.023$, respectively, while the remaining comparisons were not significant ($p$s > 0.05). Finally, for the slowest 1500 ms/item condition, the IME was evident only at serial position 3, $t(33) = 3.62$, $p = 0.001$.

Unlike the ME, then, the IME depends on the rate at which the sequence is presented, specifically, rates such as are likely to impede the opportunity for the real-time assembly of the list into a rehearsal cohort eliminate the advantage for visual presentation. These findings converge with those of Experiments 1 and 2 in pointing to the heterogeneity of modality effects in serial recall. Specifically, the IME arises in situations where sub-vocal rehearsal is afforded by relatively slow presentation rates. When the rate is doubled (relative to Experiments 1 and 2), the IME is abolished. Such an effect, along with the other manipulations designed to impede goal-oriented articulatory processes implemented in Experiments 1 and 2, implicates as the basis for the IME those time-limited articulatory control processes utilized to convert the perceptual form of the presented sequence into a rehearsal cohort to enable maintenance of the material. That slower rates (than those used in Experiments 1 and 2) also appear to attenuate the extent of the IME suggests that, even given the obligatory nature of auditory object formation, extending the time over which such objects are formed may begin to afford opportunity for successful incorporation of the sort of detailed articulatory coding that sustains serial recall and which is already more readily available for visual presentation at equivalent rates.

## 5. General discussion

Our findings may be summarized as follows. In Experiment 1, comparison between visual-silent and auditory sequences confirmed the existence of two contrasting modality effects in serial recall, the classical effect in recency (i.e., the ME) and the opposite effect in the mid-list (i.e., the IME). Two concurrent articulatory tasks—vocalizing and mouthing the visual stimuli on presentation—each abolished the IME, while simultaneously improving recency to the same magnitude as that seen with auditory presentation. In Experiment 2, the IME was eliminated by articulatory suppression, while the ME was left unaffected. In Experiment 3, the IME was abolished when presentation rate was doubled (relative to that employed in Experiments 1 and 2), while the ME was unaffected. This finding is consistent with our interpretation of Experiments 1 and 2: the IME arises due to the different constraints involved in assembling sequences of visual and auditory origin into a subvocal rehearsal cohort. On this basis, it seems clear that the pre-recency superiority of visual over auditory serial recall has an articulatory (or rather, sub-vocal articulatory) basis; the IME is abolished when the task is undertaken along with articulatory suppression and when the rate of presentation is so rapid as to impede the timely assembly of the visual list content into an integrated articulatory programme in order to subsume subvocal rehearsal. That impeding such assembly via the requirement for both item-by-item vocalization and mouthing of the sequence on presentation also abolishes the advantage of visual presentation lends further weight to this conclusion.

The immunity of auditory recency to all these factors is consistent with the view that it arises due to obligatory perceptual object formation, rather than deliberate perceptual-motor mapping processes (e.g., Jones et al., 2004, 2006; Maidment & Macken, 2012; Nicholls & Jones, 2002). On the face of it, the preservation of auditory recency in Experiment 3 even at the longest ISI might seem an anomalous with this view; since timing is a key factor affording the formation of coherent, bounded auditory objects, it might be expected, therefore, that the slower presentation rate would weaken that coherence, thereby necessarily reducing the boundedness of the object therefore diminishing the salience of the information at that boundary.

However, in this respect, it is important to recollect that timing is just one of several cues that affect perceptual organization. The
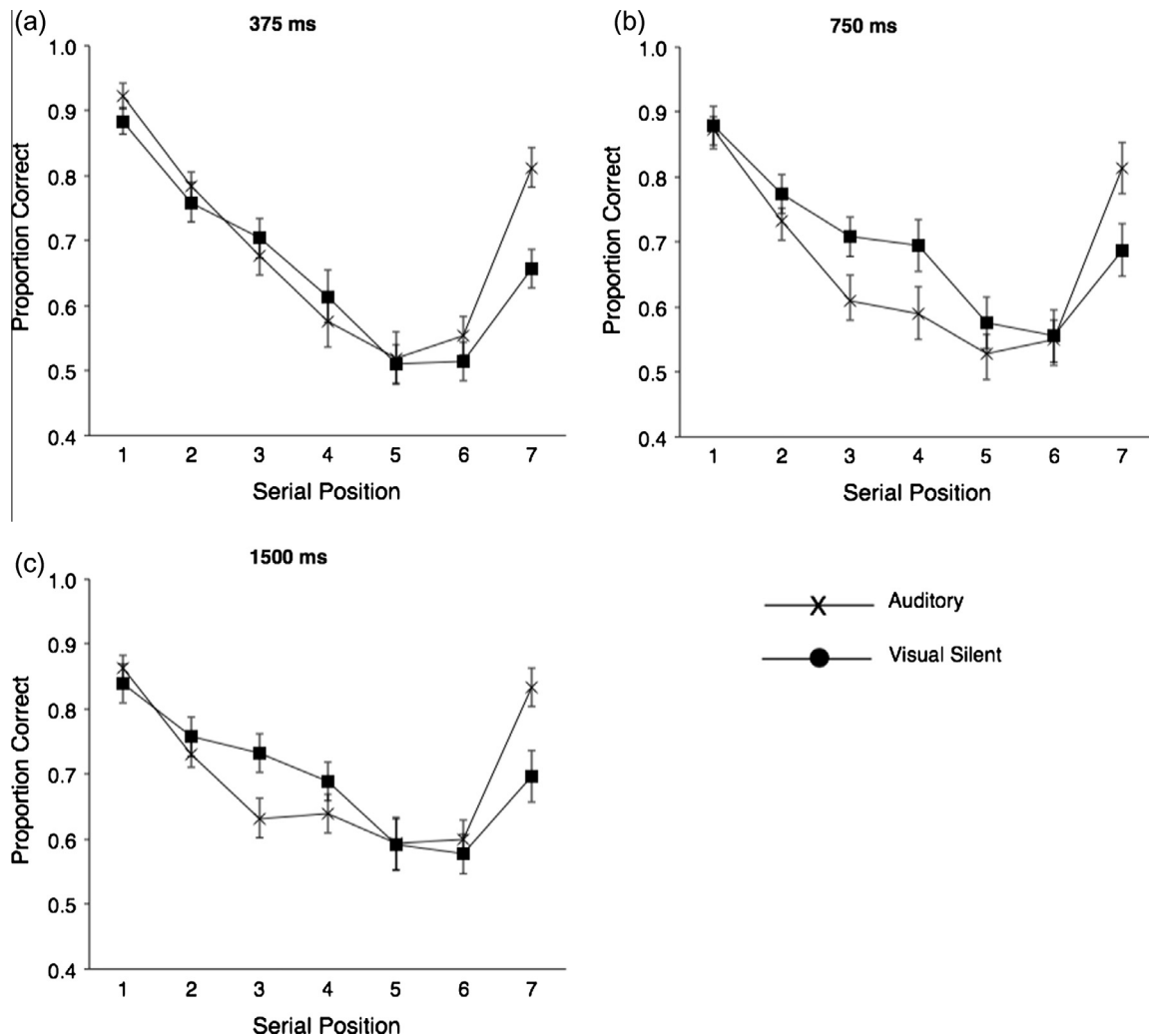
**Fig. 4.** Mean proportion correct scores for the serial recall of seven-item visual and auditory sequences at three presentation rates employed in Experiment 3. Error bars denote SE.

reciprocal relationship between rate and physical similarity is well established (see e.g., Bregman, 1990) wherein segregation into separate streams of alternating high and low tone bursts requires faster rates when the tones are closer in frequency than when they are further apart. Along with frequency similarity, a range of other acoustic factors, such as spectral similarity, temporal regularity, and so on, all combine to determine perceptual outcomes in any auditory environment (see e.g., Bregman, 1990), and in the current setting, the auditory sequences are acoustically coherent along many such dimensions such that perceptual objects may be formed even at the slower rates. Indeed, there is evidence that the auditory system may form objects over timescales considerably larger than those examined in typical psychoacoustic and perceptual settings. For example, the deviant, or oddball, effect whereby an unexpected auditory event captures attention is object-based, in that the occurrence of a novel auditory event is not in itself sufficient to lead to attentional capture, but rather, the deviance is computed with respect to the structure of the auditory objects within the environment (e.g., Hughes, Vachon, & Jones, 2005, 2007; Sussman, 2005). So, it is when a perceptual object, or stream, deviates from a trajectory defined by its past behavior that attention is captured. For current concerns, a key aspect of this effect is that it can be elicited not only by deviations on a relatively local timescale (e.g., a change of voice on one item in an otherwise

homogeneous auditory sequence in a serial recall trial: Hughes et al., 2007) but also by deviations with respect to structure defined by regularities in acoustic change over many successive trials spanning minutes, rather than seconds (Vachon, Hughes, & Jones, 2012).

Although the precise role played by auditory object formation is well established for recency (and indeed, with shorter sequences, for primacy: Jones et al., 2006; Macken, Taylor, & Jones, 2015; Maidment & Macken, 2012), the role we have described for it in pre-recency, that is, as a force impeding segmentation and rehearsal cohort formation, requires further elaboration. The preservation of auditory recency even under slow presentation rates implicates processes associated with auditory object formation and the concomitant boundary salience, but since these slower rates also attenuate the extent of IME then object formation cannot in and of itself necessarily lead to reduced addressability of within-object information. It seems plausible, therefore, to suggest that factors associated with both object formation and with scale (in this case, temporal scale) interact to fully determine the addressability and facility of perceptual-motor conversion of list content. Analogously, if we consider a visuo-manual interaction with objects, then elements bound into a small (spatial) scale object will not afford ready individual manual manipulation (while the object boundaries will remain addressable—for example, graspable—in

this sense), while at an increased scale, even those elements within a coherent object will become more manipulable in themselves. However, if the elements are not bound into an object in the first place, then their availability for manipulation will be afforded at both the smaller and larger scales. Our speculation as to the basis of the precise interaction between rate, modality and serial position demonstrated here, then, is that if the temporal scale over which an auditory object is formed is sufficiently great as to allow for temporally constrained articulatory control processes to 'manipulate' elements of that object—that is, to convert them into articulatory gestures—then the detrimental effect of object formation will be ameliorated for those within-object elements, while the benefits accruing to object boundaries remain.

The idea that the enhanced recency for articulated visual lists depends on accompanying acoustic input is ruled out, since the enhancement is equivalent for vocalized and silently mouthed conditions. This finding highlights another broad issue, namely that quantitatively or functionally similar patterns of behavior—in this case, enhanced recency for auditory, mouthed and vocalized sequences—is not always evidence of common underlying mechanisms. In this respect, performance differences between visually and auditorily presented sequences have much in common with those found between auditory and lip-read material (Maidment, Macken, & Jones, 2013). Firstly, lip-read recency, while superficially identical to auditory recency under control conditions, is eliminated by articulatory suppression, suggesting that like the mid-list advantage for visual sequences examined here, it has an articulatory rather than perceptual basis. Secondly, a heterogeneity of process is indicated by evidence that cross-modal interactions between auditory and lip-read sequences and auditory and lip-read suffixes differ in their mode of action depending on whether or not the suffix is bound to its auditory complement: Functionally identical performance effects emerge via fundamentally different mechanisms, dependent on the detailed context of sequence presentation. An analogous picture is apparent in the current data. We found no evidence that recency in visual-vocalized sequences was dependent on the presence of an auditory complement. Conversely, auditory recency was unaffected by concurrent articulatory demands.

If different manifestations of supposedly identical effects are shown to derive from fundamentally different mechanisms, then this has considerable theoretical implications for the understanding of short-term memory and its place within a broader cognitive architecture. In particular, these data suggest that it is appropriate to ask precisely what is meant by 'modality'. It is implicitly assumed by both decay-based (e.g., Baddeley, 2003; Crowder & Morton, 1969; Henson, 1998) and interference-based (e.g., Nairne, 1990; Neath, 2000) accounts of short-term memory that modality is an intrinsic stimulus feature and, as such, modality effects arise because such *a priori* features define the route by which verbal material gains access to a temporary storage system or give stimuli more or less protection from interference. Our data suggest instead that what underlies modality effects in particular, and recall performance more generally, is the way in which each presentation-mode affords perceptual object formation and the impact of that on mapping to a motor form that serves to maintain and reproduce the material. Modality effects thus represent the outcome of an interaction between and within the physical properties of the stimulus (visual, auditory, articulatory) and processes governing perceptual-motor integration, as well as the particular demands of a given task as they unfold within the task setting (Macken et al., 2015).

The modality effect is, therefore, one of a series of canonical effects in verbal short-term memory that have been reconstrued within an embodied framework; others include the word-length effect (e.g., Baddeley, Thomson, & Buchanan, 1975), the phonolog-ical similarity effect (Baddeley, 1968; Conrad & Hull, 1964), the lexicality effect (Gathercole, Pickering, Hall, & Peaker, 2001; Hulme, Maughan, & Brown, 1991; Roodenrys, Hulme, & Brown, 1993), the irrelevant speech effect (Macken, Phelps, & Jones, 2009; Macken et al., 2015), the talker variability effect (Hughes et al., 2009), as well as articulatory suppression and the suffix effect (e.g., Baddeley, Lewis & Vallar, 1985; Jones et al., 2004; Maidment & Macken, 2012) The classical conceptualization of each has been the subject of renewed scrutiny from an embodied, perceptual-motor perspective suggesting that short-term memory performance can be accounted for without the classical cognitive gesture of cleaving the perceptual and motor domains from the core cognitive system.

For example, the word-length effect—the better recall of lists of long compared to short words—arises not from the decay of item representations in temporary storage (e.g., Baddeley, 1986) but rather from the confounding effect of longer words being typically of increased articulatory complexity (Service, 1998). The phonological similarity effect—the poorer recall of similar (e.g., "b, g, c...") compared to dissimilar items (e.g., "r, j, q...")—emerges from the combined and distinct influence of auditory-perceptual and articulatory processes (Jones et al., 2004, 2006; Maidment & Macken, 2012). Finally, and of particular relevance to the characterization of modality effects, lexicality effects—superior serial recall as a function of the items' lexical status (or word-likeness)—operate via distinct mechanisms depending both on the presentation mode of the stimulus and the specific demands of the task (Macken et al., 2014). Thus, the absence of a lexicality effect in auditory serial recognition contrasts with a robust lexicality effect in both auditory and visual serial recall. It was long argued that this difference was due to the reduced burden on item memory in recognition, compared to recall, and therefore the obviation of processes whereby long-term linguistic/phonological knowledge could be utilized to bolster the integrity of volatile short-term representations (e.g., Gathercole et al., 2001; Jefferies, Frankish, & Ralph, 2006). On this basis, since no overt recall is demanded by serial recognition, no lexicality effect is predicted and for auditory sequences at least, none is found. However, a robust lexicality effect is found for visual serial recognition and, crucially, it is abolished by articulatory suppression (Macken et al., 2014). The presence or absence of lexicality effects then, like modality effects, is not simply determined by the properties of the stimulus (in this case its lexical status), but instead will have an articulatory basis where performance is measured using a task that affords or requires sub-vocal rehearsal (such as is the case in visual serial recognition) or an auditory perceptual basis when measured by a task that facilitates auditory pattern matching, such as is the case in auditory serial recognition. We propose, therefore, that the multiplicity of modality effects reported here and their immunity or otherwise in the face of a range of manipulations presents further evidence for the centrality of domain-general perceptual and motor processes in short-term memory (Hughes et al., 2009, 2011; Jones et al., 2004, 2006; Maidment et al., 2013).

It was always the case that cognitive psychology, as a paradigm, would have to address questions of modality, not only because, necessarily, part of the problem for an information processing system is how the information to be processed is transduced in the first place, but also because, to be a viable way of investigating human behavior it would have to account for effects of the modality in which nominally equivalent information was presented. The classical cognitive solution to this has been to partition the transduction process from the central processing of the derived representations, and to ascribe particular inherent advantages or characteristics to the various modalities. A pattern of behavior in which one modality always sustains a particular qualitative or quantitative relationship to another is amenable to such an

approach, and cognitive psychology is replete with instances that ascribe various kinds of superiority to the auditory over the visual. What we have shown here is that the empirical basis for this is unsound, and so the question of modality and its effect on short-term memory performance raises general questions about how to construe such performance, and about the viability of a cognitive approach in general.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2016.06.013.

## References

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin, 135*, 50–68.
Arenberg, D. (1968). Input modality in short-term retention of old and young adults. *Journal of Gerontology, 23*, 462–465.
Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *The Quarterly Journal of Experimental Psychology, 20*, 249–263.
Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417–432.
Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829–839.
Baddeley, A. D., & Larsen, J. D. (2007). The phonological loop unmasked? A comment on the evidence for a "perceptual-gestural" alternative. *The Quarterly Journal of Experimental Psychology, 60*, 497–504.
Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the phonological loop. *Quarterly Journal of Experimental Psychology, Section A – Human Experimental Psychology, 36*, 233–252.
Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575–589.
Beaman, C. P. (2002). Inverting the modality effect in serial recall. *Quarterly Journal of Experimental Psychology, 55*, 371–389.
Bregman, A. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, Mass.: MIT Press.
Bregman, A. S., & Rudnicky, A. I. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology, Human Perception and Performance, 1*, 263–267.
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*, 627–652.
Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*, 429–432.
Conrad, R., & Hull, A. J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science, 10*, 135–136.
Corballis, M. C. (1966). Rehearsal and decay in immediate recall of visual recall of visually and aurally presented items. *Canadian Journal of Psychology, 20*, 43–50.
Crowder, R. G. (1970). The role of one's own voice in immediate memory. *Cognitive Psychology, 1*, 157–178.
Crowder, R. G. (1986). Auditory and temporal factors in the modality effect. *Journal of Experimental Psychology: Learning Memory and Cognition, 12*, 268–278.
Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception and Psychophysics, 5*, 365–373.
Drewnowski, A., & Murdock, B. A. (1980). The role of auditory features in memory span for words. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 319–322.
Frankish, C. (1989). Perceptual organization and precategorical acoustic storage. *Journal of Experimental Psychology: Learning Memory and Cognition, 15*, 469–479.
Frankish, C. (2008). Precategorical acoustic storage and the perception of speech. *Journal of Memory and Language, 58*, 815–836.
Gathercole, S. E. (1986). The modality effect and articulation. *Quarterly Journal of Experimental Psychology, 38*, 461–474.
Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences in serial recognition and serial recall. *Quarterly Journal of Experimental Psychology, 54A*, 1–30.
Glenberg, A. M., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12*, 3–15.
Greene, R. L., & Crowder, R. G. (1984). Modality and suffix effects in the absence of auditory stimulation. *Journal of Verbal Learning and Verbal Behavior, 23*, 371–382.
Grenfell-Essam, R., Ward, G., & Tan, L. (2013). The role of rehearsal on the output order of immediate free recall of short and long lists. *Journal of Experimental Psychology: Learning Memory and Cognition, 39*, 317–347.

Grossberg, S., & Pearson, L. R. (2008). Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: Toward a unified theory of how the cerebral cortex works. *Psychological Review, 115*, 677–732.
Harvey, A. J., & Beaman, C. P. (2007). Input and output modality effects in immediate serial recall. *Memory, 15*, 693–700.
Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36*, 73–137.
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*, 1411–1425.
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*, 1435–1447.
Hughes, R. W., Vachon, F., & Jones, D. M. (2005). Auditory attentional capture during serial recall: Violations at encoding of an algorithm-based neural model? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 736–749.
Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 33*, 1050–1061.
Hulme, C., Maughan, S., & Brown, G. D. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*, 685–701.
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*, 339–373.
Jefferies, E., Frankish, C. R., & Ralph, M. A. L. (2006). Lexical and semantic influences on item and order memory in immediate serial recognition: Evidence from a novel task. *Quarterly Journal of Experimental Psychology, 59*, 949–964.
Jones, D. M., Hughes, R. W., & Macken, W. J. (2006). Perceptual organization masquerading as phonological storage: Further support for a perceptual-gestural view of short-term memory. *Journal of Memory and Language, 54*, 265–281.
Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 656–674.
Katshu, M., & D'Avossa, G. (2014). Fine-grained, local maps and coarse, global representations support human spatial working memory. *PLoS One, 9*(9), e107969.
Macken, W. J., & Jones, D. M. (1995). Functional characteristics of the inner voice and the inner ear: Single or double agency? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 436–448.
Macken, W. J., Phelps, F. G., & Jones, D. M. (2009). What causes auditory distraction? *Psychonomic Bulletin & Review, 16*, 139–144.
Macken, B., Taylor, J., & Jones, D. M. (2014). Language and short-term memory: The role of perceptual-motor affordance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1257–1270.
Macken, B., Taylor, J., & Jones, D. (2015). Limitless capacity: A dynamic object-oriented approach to short-term memory. *Frontiers in Psychology, 6*.
Madigan, S. A. (1971). Modality and recall order interactions in short-term memory for serial order. *Journal of Experimental Psychology, 87*, 294–296.
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*, 989–997.
Maidment, D. W., Macken, B., & Jones, D. M. (2013). Modalities of memory: Is reading lips like hearing voices? *Cognition, 129*, 471–493.
Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision, 12*, 1–14.
Maylor, E. A., Vousden, J. I., & Brown, G. D. A. (1999). Adult age differences in short-term memory for serial order: Data and a model. *Psychology and Aging, 14*, 572–594.
Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology, 78*, 679–684.
Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*, 251–269.
Nairne, J. S., & Walters, V. L. (1983). Silent mouthing produces modality- and suffix-like effects. *Journal of Verbal Learning and Verbal Behavior, 22*, 475–483.
Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin and Review, 7*, 403–423.
Nicholls, A. P., & Jones, D. M. (2002). Capturing the suffix: Cognitive streaming in immediate serial recall. *Journal of Experimental Psychology: Learning Memory and Cognition, 28*, 12–28.
Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of serial recall. *Psychological Review, 105*, 761–781.
Penney, C. G. (1975). Modality effects in short-term verbal memory. *Psychological Bulletin, 82*, 68–84.
Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory and Cognition, 17*, 398–422.
Penney, C. G., & Blackwood, P. A. (1989). Recall mode and recency in immediate serial recall: Computer users beware! *Bulletin of the Psychonomic Society, 27*, 545–547.
Roodenrys, S., Hulme, C., & Brown, G. (1993). The development of short-term memory span: Separable effects of speech rate and long-term memory. *Journal of Experimental Child Psychology, 56*, 431–442.

Routh, D. A. (1971). Independence of modality effect and amount of silent rehearsal in immediate serial recall. *Journal of Verbal Learning and Verbal Behavior, 10*, 213–218.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80*, 1–46.

Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology, 51A*, 283–304.

Sussman, E. S. (2005). Integration and segregation in auditory scene analysis. *Journal of the Acoustical Society of America, 117*, 1285–1298.

Tan, L. S., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning Memory and Cognition, 26*, 1589–1625.

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

Tremblay, S., Parmentier, F. B. R., Guérard, K., Nicholls, A. P., & Jones, D. M. (2006). A spatial modality effect in serial memory. *Journal of Experimental Psychology: Learning, Memory & Cognition, 32*, 1208–1215.

Turner, M. L., LaPointe, L. B., Cantor, J., Reeves, C. H., Griffeth, R. H., & Engle, R. W. (1987). Recenecy and suffix effects found with auditory presentation and with mouthed visual presentation: They're not the same thing. *Journal of Memory and Language, 26*, 138–164.

Turner, M. L., Scwartz, M. K., Clifton, G. E., & Engle, R. W. (1994). Effects of vocabulary size and acoustic simialrity on serial recall of mouthed stimuli. *Journal of General Psychology, 121*, 361–376.

Vachon, F., Hughes, R. W., & Jones, D. M. (2012). Broken expectations: Violation of expectancies, not novelty, captures auditory attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 38*, 164–177.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin, 138*, 1172–1217.

Ward, G. (2002). A recency-based account of the list-length effect in free recall. *Memory & Cognition, 30*, 885–892.

Warren, R. M. (1999). *Auditory perception: A new synthesis*. Elmsford, NY: Pergamon Press.

Watkins, M. J., Watkins, O. C., & Crowder, R. G. (1974). The modality effect in free and serial recall as a function of phonological similarity. *Journal of Verbal Learning and Verbal Behavior, 13*, 430–447.

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