Modality effects in free recall: A retrieved-context account

Pazdera, J. K. and Kahana, M. J.

University of Pennsylvania

Author Note

Support for this work was provided by the National Institutes of Health (MH55687). The authors would like thank Rivka Cohen for assistance in designing and programming our simulations, Tung Phan for helping to design our intrusion analysis for Experiment 1, as well as Bob Greene, Alice Healy, and Geoff Ward for providing insightful feedback on an earlier draft of this manuscript. Correspondence concerning this article should be addressed to Michael J. Kahana. E-mail: kahana@psych.upenn.edu

Abstract

The modality effect refers to the robust finding that memory performance differs for items presented aurally, as compared with visually. Whereas auditory presentation leads to stronger recency performance in immediate recall, visual presentation often produces better primacy performance. To investigate these differences, we conducted two large-scale immediate free recall experiments via Amazon Mechanical Turk. In both experiments, participants studied visual and auditory word lists of varying lengths and rates of presentation. We observed a recency advantage and primacy disadvantage during auditory trials, relative to visual trials, reflecting common modality and inverse modality effects in free recall. Participants were more likely to initiate recall from recency items on auditory trials than on visual trials, though the modality effect persisted regardless of recall start location. Meanwhile, an analysis of intrusion errors revealed that participants were more likely to erroneously recall words from one list prior on visual trials than they were on auditory trials. We discuss our findings within the framework of retrieved-context theories of episodic memory, fitting the Context Maintenance and Retrieval model to our data in order to develop a retrieved-context account of the modality effect. Through our simulations, we demonstrate that both the modality and inverse modality effects can be explained as products of faster context drift and stronger context-to-item association formation during auditory presentation, relative to visual. Finally, we demonstrate that our model predicts the persistence of the modality effect across recall start positions, as well as the novel effects of modality we observed on prior-list intrusion behavior.

Keywords: memory, free recall, modality effect, context, model

Modality effects in free recall: A retrieved-context account

Dissociations between recall of recent and remote memories have provided some of the most compelling evidence that differential retrieval processes underlie short-term and long-term memory. The modality effect is one such phenomenon, characterized by the greater accessibility of recently presented material when it was studied aurally as compared with visually. In contrast, retrieval of more remote memoranda demonstrates no such auditory advantage, and in some cases even shows an auditory disadvantage, referred to in recent literature as the inverse modality effect. This clear demarcation between the effects of presentation modality on recent and remote information has often been cited as evidence that recent items hold a unique status in memory, governed by a distinct set of rules.

The present paper proposes an alternative account for the modality effect, based on the retrieved-context theory of Howard and Kahana (2002). Before introducing our model, we first describe several key findings concerning modality and inverse modality effects in free recall, then summarize the major prior theoretical accounts. Finally, we report data from two experiments designed to provide new insights into how modality of presentation influences the dynamics of recall, including how subjects initiate recall and make transitions among items. We then fit our retrieved-context model to key aspects of the reported data both to evaluate the model and to help understand the underlying processes that may differ between the encoding of visually and aurally studied material.

The Modality and Inverse Modality Effects

The modality effect is ubiquitous, appearing across multiple memory tasks including free recall (Craik, 1969, 1970; Murdock, 1968; Murdock & Walker, 1969; Murray, 1966), serial recall (Conrad & Hull, 1968; Corballis, 1966; Laughery & Pinkus, 1966; Morton & Holloway, 1970), and paired-associate learning (Murdock, 1966). The modality effect appears not only during auditory presentation, but also during vocalized reading (e.g., Conrad & Hull, 1968; Crowder, 1971; Murray, 1966), silent mouthing (e.g., Greene &

Crowder, 1984, 1986; Nairne & Crowder, 1982; Nairne & Walters, 1983), lip reading (Campbell & Dodd, 1980; Greene & Crowder, 1984), sign language (Shand, 1980; see also Shand & Klima, 1981), and finger spelling (Krakow & Hanson, 1985). More recently, the existence of an inverse modality effect has also been established, in which silent visual presentation produces better recall than auditory presentation for early or mid-serial items. This finding was first noted by Craik (1969), but despite its relatively frequent appearance in studies of the modality effect (see Grenfell-Essam, Ward, & Tan, 2017 for a review), the inverse modality effect has not been directly addressed in the literature until quite recently (Beaman, 2002; Grenfell-Essam et al., 2017; Macken, Taylor, Kozlov, Hughes, & Jones, 2016). Researchers have investigated numerous potential influences on the modality effect, including word length (M. J. Watkins, 1972; M. J. Watkins & Watkins, 1973), word frequency (Nilsson, 1975; O. C. Watkins & Watkins, 1977), acoustic similarity (Crowder, 1971; Greene, 1989; M. J. Watkins, Watkins, & Crowder, 1974), list suffixes (e.g., Huang & Glenberg, 1986; Morton & Holloway, 1970; Spoehr & Corin, 1978), distractor tasks (e.g., Gardiner & Gregg, 1979; Gathercole, Gregg, & Gardiner, 1983; Glenberg, 1984; Routh, 1976), presentation rate (e.g., Corballis, 1966; Laughery & Pinkus, 1966; Murdock & Walker, 1969), and list length (Grenfell-Essam et al., 2017; Roberts, 1972).

Faster rates of stimulus presentation typically increase the magnitude of the modality effect. Murdock and Walker (1969), Murray (1966), and Roberts (1972) all manipulated presentation rate in free recall, and observed larger modality effects at faster rates of presentation. In serial recall, Corballis (1966) and Laughery and Pinkus (1966) have also shown increased auditory advantages under conditions of fast or accelerating presentation. Meanwhile, Macken et al. (2016) manipulated presentation rate without observing such an effect; however, they found evidence to suggest that faster presentation may attenuate inverse modality effects. The influence of list length on the modality effect has received somewhat less attention in the literature. A recent study by Grenfell-Essam et al. (2017) tested seven list lengths between 2 and 12, finding that the modality effect extended from one to three items as list length increased across this range. In contrast, an experiment by Roberts (1972) found no difference in the modality effect across lists lengths of 10 through 40. Together, these studies suggest that the extent of the modality effect may expand with list length, though only to a certain upper limit. Additionally, both studies observed prerecency visual advantages at longer list lengths, providing evidence of an interaction with the inverse modality effect, as well.

Theoretical Accounts of the Modality Effect

Precategorical acoustic storage

Early accounts of the modality effect widely attributed the auditory recency advantage to the existence of separate sensory registers for visual and auditory information. Researchers suggested that the auditory store must hold more information (Murdock, 1967; Murdock & Walker, 1969) or persist longer (Craik, 1969; Crowder & Morton, 1969; Murdock, 1966) than the corresponding visual store. The first formal model of the modality effect, Precategorical Acoustic Storage (PAS; Crowder & Morton, 1969), proposed that when a stimulus is initially perceived, it enters a modality-specific sensory buffer prior to being categorized through linguistic and semantic processing. Information held in precategorical storage can be utilized during recall, but decays over time and can be replaced by subsequent stimuli. In the PAS view, the auditory store retains information longer than the visual store, resulting in greater accessibility of recent auditory information. PAS has long remained influential, though discoveries such as the long-term modality effect (Gardiner & Gregg, 1979; Glenberg, 1984), the inverse modality effect (Beaman, 2002; Grenfell-Essam et al., 2017; Macken et al., 2016), and visual suffix effects (Campbell & Dodd, 1980; Greene & Crowder, 1984) have prompted subsequent revisions (Crowder, 1978, 1983; Penney, 1989) and continue to challenge the model today.

Temporal distinctiveness theory

Temporal distinctiveness theory (Glenberg & Swanson, 1986; see also Gardiner, 1983; Gardiner & Gregg, 1979; Glenberg & Fernandez, 1988) arose as a response to the discovery of long-term recency and modality effects in the continuous-distractor free recall paradigm (Bjork & Whitten, 1974; Gardiner & Gregg, 1979; Glenberg, 1984). The theory proposes that memory search processes use the temporal coding of items as a retrieval cue, through the delineation of temporal search sets. Inspired by O. C. Watkins and Watkins (1975) and the Search of Associative Memory (SAM) model (Raaijmakers & Shiffrin, 1980, 1981), temporal distinctiveness theory incorporates a cue-overload assumption, such that the ability to recall any individual item is inversely proportional to the total number of items in the search set. Therefore, the more items whose temporal contexts fall within the search window, the less likely it is that each item will be recalled. To explain the modality effect, Glenberg and Swanson (1986) hypothesized that auditory items have more precise temporal coding than visual items, resulting in reduced confusability between the temporal contexts of auditory items. As such, the narrow search set used to recall end-of-list items will overlap with fewer auditory items than visual items, producing improved recall for the final few items of auditory lists. This effect washes out as the search set increases in scope, explaining the lack of auditory advantages among prerecency items. The idea that auditory presentation improves the precision of temporal coding has been criticized by Greene and Crowder (1988), however, due to their finding that mid-serial auditory stimuli show no better encoding of serial order information than do mid-serial visual stimuli.

The feature model

Nairne (1990) developed a more general framework for explaining modality effects, in which the features of currently active short-term memory traces act as retrieval cues during the search of long-term memory. These features include both modality-specific (e.g., physical traits) and modality-independent (e.g., semantic information) characteristics of an item. Memory traces degrade in short-term memory as their features are overwritten by subsequent stimuli and recalled items with overlapping features. Nairne (1990) explained the modality effect by proposing that auditory stimuli carry a richer set of modality-specific features than do visual stimuli. Consequently, traces of recent auditory stimuli persist longer in short-term memory and are more resistant to output interference, thereby improving their accessibility during recall.

Output interference accounts

The notion that resistance to output interference underlies the modality effect has been popularized by findings that the magnitude of the effect changes as a function of recall start position (Cowan, Saults, & Brown, 2004; Cowan, Saults, Elliott, & Moreno, 2002; Craik, 1969; Grenfell-Essam et al., 2017; Harvey & Beaman, 2007). Studies by Cowan et al. (2002), Craik (1969), and Harvey and Beaman (2007) all found larger modality effects when participants were instructed to initiate recall from earlier, rather than later serial positions. Grenfell-Essam et al. (2017) found a similar pattern when comparing trials where participants naturally initiated recall from the first list item against those where they started from one of the final four list items. Cowan et al. (2004) challenged the output interference hypothesis by demonstrating that ceiling effects may simply be constricting the modality effect under conditions of low interference. However, Harvey and Beaman (2007) have pushed back against this conclusion with findings that the magnitude of the modality effect depends more on output position during written recall than during spoken recall.

Rehearsal and the inverse modality effect

At this stage, there exists no formal model of the inverse modality effect, and in fact most traditional theories of the modality effect fail to explain how any visual advantage might arise. At present, the most popular account of the inverse modality effect is that it originates from an improved ability or increased tendency to rehearse visual items, relative to auditory items (Grenfell-Essam et al., 2017; Macken et al., 2016). Macken et al. (2016)

found that the visual advantage for primacy and mid-serial items could be reduced or eliminated by disrupting rehearsal via articulatory suppression, vocalized reading, silent mouthing, or fast item presentation. These results suggest that the locus of the inverse modality effect lies in a rehearsal advantage during silent reading. Grenfell-Essam et al. (2017) also supported this hypothesis, and further suggested that it may be more difficult to incorporate new aurally presented words into an expanding rehearsal set.

Towards a Retrieved-Context Account of the Modality Effect

Retrieved-context theories of episodic memory (Howard & Kahana, 2002; Lohnas, Polyn, & Kahana, 2015; Polyn, Norman, & Kahana, 2009) incorporate mechanisms similar to those of existing single-process models of the modality effect. In a retrieved-context model, memory traces form associations with the features of the contexts in which they are encoded. Memory search uses the current state of context as a retrieval cue, and recalling a memory reinstates the context associated with that memory. Context can include temporal, semantic, and physical features, subsuming mechanisms of both temporal distinctiveness theory (Glenberg & Swanson, 1986) and the Nairne (1990) feature model. In the present study we examined whether a recent implementation of retrieved-context theory, the Context Maintenance and Retrieval (CMR2) model (Lohnas et al., 2015; Polyn et al., 2009), could account for modality and inverse modality effects by allowing a limited selection of parameters to vary across presentation modality conditions. To help evaluate the model's predictions, as well as to more precisely understand the effects of modality on retrieval dynamics, we conducted two large-scale experiments in which participants completed an immediate free recall task containing lists of variable modality, length, and rate of presentation. We first report data from these two experiments, then present our analysis of the CMR model. Finally, we interpret our modeling results in terms of the psychological processes hypothesized to underlie the modality and inverse modality effects.

Experiment 1

In the process of developing a retrieved-context theory of the modality effect, we first sought to collect a set of behavioral data that would allow us to precisely describe the effects of modality on a variety of recall dynamics. Experiment 1 investigated the interactions between modality, list length, and presentation rate in an immediate free recall task. Participants completed 16 trials, with lists of 12 or 24 words presented either visually or aurally at one of two rates. Each participant completed two trials under each of the eight combinations of these conditions. In order to collect as extensive a dataset as possible, we deployed our experiment online via Amazon's crowdsourcing platform, Mechanical Turk (MTurk).

Methods

Participants. We recruited participants for the present study through MTurk, an online platform for large-scale data collection from diverse populations (Mason & Suri, 2012). In order to qualify for our study, participants were required to possess at least a 95% approval rating on MTurk, as well as to be from the United States. Additionally, the task could only be completed from a laptop or desktop computer, and any individuals attempting to access the experiment from a mobile device or tablet were prevented from doing so. A total of 1100 participants (469 male, 583 female, 48 unreported) completed Experiment 1. Participants were paid $7.50 for completion of the task.

We excluded 11 participants from our analyses due to technical problems that disrupted their sessions. We excluded an additional 201 participants who met at least one of following five criteria, which were designed to filter out individuals who failed to stay on task or who may have cheated: 1) Making zero recalls on multiple trials. 2) Correctly recalling more than 95% of the words in the study. 3) Indicating in a post-experiment questionnaire on memory strategies that they had written or typed notes at any point during the task. 4) Averaging fewer than three correct math problems per trial during their distractor periods. 5) Demonstrating extremely low accuracy on the math distractor task (i.e., participants with a z-score below -2). After applying all exclusion criteria, 888 participants (376 male, 474 female, 38 unreported) contributed to our analyses. Ages ranged from 18 to 71 years, with a mean age of 36.6 years ($SD = 10.4$).

Materials. The word pool for this experiment consisted of 556 English nouns, with each participant assigned a random 288-word subset. The word pool was identical to that used in Experiment 4 of the Penn Electrophysiology of Encoding and Retrieval Study (Kahana, Aggarwal, & Phan, 2018), with the exception that 20 words possessing at least one homophone in the English language were excluded from the pool. Audio recordings of all words in the pool were made by a single voice actress. The duration of audio files ranged from 268 ms to 1054 ms. Visual words, as well as math distractor problems, were presented in white, 60-point Arial font on a black background. Visual words were presented in all capital letters. The task was implemented using the jsPsych library for JavaScript (de Leeuw, 2015), and psiTurk (Gureckis et al., 2016) was used to host and manage the study on MTurk.

Design. The experiment used a fully within-subjects design. There were three independent variables of interest: modality (Visual and Auditory), list length (12 and 24), and presentation rate (Slow and Fast).

To facilitate comparisons between our present experiments and those of the Penn Electrophysiology of Encoding and Retrieval Study (e.g., Healey, Crutchley, & Kahana, 2014; Healey & Kahana, 2016; Kahana et al., 2018; Long, Danoff, & Kahana, 2015), we included two additional extraneous manipulations. These manipulations included varying the positions of semantically related items within each list and varying the duration of the interlist distractor period (12 or 24 seconds). Because these manipulations do not relate to our hypotheses regarding the modality effect, and to limit the complexity of our exposition, we do not report on these manipulations in the present manuscript. For completeness, however, we provide a description of the semantic manipulation in Appendix A.

Procedure. Upon accessing the experiment through MTurk, participants were first presented with a consent form. After reading the form, they were given the choice to either continue the task or exit and return to MTurk. If they chose to continue, they were given a brief audio test in order to ensure that they would be able to hear the words presented in the task. The audio test consisted of listening to recordings of three different words, and typing each of those words into a text box. Participants could replay each word as many times as desired, allowing them to adjust their volume to a comfortable level before beginning the main task. The words involved in the audio test were not part of the word pool for the task, and thus could not appear as stimuli during the main free recall task. After successfully typing all three of the test words, participants viewed the task instructions and entered the main phase of the experiment.

Participants performed an immediate free recall task consisting of two practice trials followed by 16 experimental trials. Each trial consisted of a pre-list arithmetic distractor task, a list of word presentations, and an immediate free recall test. The experiment concluded with a final free recall task, followed by a brief exit survey. Each participant received one trial with each combination of modality, list length, presentation rate, and distractor duration. The 16 trial types were administered in a fully randomized order for each participant.

At the start of each trial, participants were presented with a distractor task for either 12 or 24 seconds. The distractor task consisted of answering math problems of the form $A + B + C = ?$, where A, B, and C were positive, single-digit integers. Problems were displayed onscreen one at a time, and participants were instructed to type the answer to each equation as quickly and accurately as possible. New problems continued to appear until the full 12 or 24 seconds had elapsed, at which point the screen immediately went blank for a duration of 2000 ms. Following this delay, a ten-second countdown appeared on the screen. The countdown was followed by another delay of 1500 ms, during which a fixation cross was displayed. After the delay, the fixation cross was removed and the first word presentation in the trial began.

During visual trials, words were presented onscreen to participants. Presentation rate was modulated such that words were displayed onscreen for 800 ms in the fast-presentation condition and 1600 ms in the slow-presentation condition. A jittered inter-stimulus interval of 800–1200 ms (uniformly distributed) followed each word, during which the screen remained blank. As such, word onsets began 1600–2000 ms apart in the fast-presentation condition and 2400–2800 ms apart in the slow-presentation condition.

During auditory trials, a blank, black screen was displayed while words were presented aurally as a sequence of vocal recordings. Presentation rate was modulated on auditory trials such that a new word would begin playing every 1600-2000 ms on fast trials, while a new word would begin playing every 2400-2800 ms on slow trials. In order to avoid altering the qualities of the words themselves, the audio clips were not modified between presentation rate conditions. As the longest audio recording was 1054 ms, it was never possible for one word to begin playing before the previous word had finished.

In order to avoid bias toward any one condition, practice trials contained 18 s distractor periods, were 18 items long, and had 2000-2400 ms delays between word onsets. One practice list was presented visually, the other aurally, with random ordering.

After the inter-stimulus interval of the final word in each list, an additional jittered delay of 1200–1400 ms (uniformly distributed) occurred, after which a tone sounded and a row of asterisks appeared onscreen for 500 ms to indicate the start of the recall period. The asterisks were then replaced by an empty text box, and participants were given 60 s to type in as many words as they could recall from the most recent list, in any order. Participants submitted each word via a key press. Recalls were spell-checked and scored automatically using an algorithm identical to that used by Healey (2018) in analyzing a similar typed recall task (see Appendix B for a description). Submitting a word cleared the text box and allowed the participant to begin entering their next recall. As words were cleared from the screen after submission, participants could not look back at words they had already recalled. Once the time limit had elapsed, the recall period ended and a black screen was displayed for 2000 ms. Following this delay, the participant was instructed to press any key to begin the next trial. Pressing a key initiated the distractor task for the next trial.

Following the recall period of the final trial, instructions were displayed on the screen for a final free recall test. The instructions were followed by a blank screen for 5000 ms, after which a tone sounded and a row of asterisks appeared onscreen for 500 ms, indicating the start of the final free recall test. During final free recall, participants were given 5 minutes to recall as many words as possible from all trials, in any order. Participants entered their recalls into a text box in a manner identical to all other recall periods. After final free recall ended, an exit survey appeared and participants were given instructions on how to submit their completed assignment on MTurk for compensation.

Results

We have organized our results into several sections, based on the type of analysis conducted. We first report the effects of our manipulations on performance for primacy and recency items, respectively. We next consider differences in recall initiation behavior by condition, as well as differences in memory performance by recall start position. Then, we investigate temporal clustering tendencies, followed by the effect of modality on the quantity and recency of prior-list intrusions. Finally, we present a preliminary analysis of differences by modality in final free recall performance.

Primacy and recency effects. Figure 1 illustrates the serial position curves (SPCs) for all eight combinations of modality, list length, and presentation rate. The data show both modality and inverse modality effects under all list lengths and presentation rates, with slow presentation resulting in a particularly large visual advantage over several primacy items. To analyze differences in recall for primacy items across conditions, we performed a 2 (Modality: Visual and Auditory) × 2 (List Length: 12 and 24) × 2 (Presentation Rate: Slow and Fast) repeated measures ANOVA. For this analysis, we averaged the recall probabilities of the first three list items and used the resulting value in each condition as our measure of primacy performance. We found significant main effects of modality, $F(1, 887) = 132, MS_e = 6.66, p < .001, \eta_p^2 = .130$, list length, $F(1, 887) = 400, MS_e = 18.7, p < .001, \eta_p^2 = .311$, and presentation rate, $F(1, 887) = 114, MS_e = 4.88, p < .001, \eta_p^2 = .114$. We also found a significant interaction effect between modality and presentation rate, $F(1, 887) = 7.67, MS_e = 0.35, p = .006, \eta_p^2 = .009$. The interaction effects between modality and list length, $F(1, 887) < 1$, and list length and presentation rate, $F(1, 887) = 1.46, MS_e = 0.067, p = .228, \eta_p^2 = .002$, were not significant, nor was the three-way interaction between modality, list length, and presentation rate, $F(1, 887) < 1$. Recall probability for primacy items was greater for visual stimuli than for auditory ones, showing an inverse modality effect strongly resembling those observed by Beaman (2002) and Craik (1969). Primacy performance was also higher on shorter lists than on longer lists. Although increasing the rate of presentation reduced primacy performance on both auditory and visual trials, visual performance was disproportionately affected, resulting in an overall reduction of the visual primacy advantage with faster presentation. We therefore observed an attenuation of the inverse modality effect during faster trials.

We analyzed recall for recency items in each condition by averaging the recall probabilities of the last five list items. We then tested for differences between these recency scores using a 2 (Modality: Visual and Auditory) × 2 (List Length: 12 and 24) × 2 (Presentation Rate: Slow and Fast) repeated measures ANOVA. We found significant main effects of modality, $F(1, 887) = 190, MS_e = 6.84, p < .001, \eta_p^2 = .176$, list length, $F(1, 887) = 423, MS_e = 12.9, p < .001, \eta_p^2 = .323$, and presentation rate, $F(1, 887) = 19.4, MS_e = 0.43, p < .001, \eta_p^2 = .021$. As well, we found significant interaction effects between modality and presentation rate, $F(1, 887) = 6.70, MS_e = 0.15, p = .010, \eta_p^2 = .007$, and list length and presentation rate, $F(1, 887) = 18.9, MS_e = 0.38, p < .001, \eta_p^2 = .021$. The interaction between modality and list length was not significant, $F(1, 887) < 1$, nor was the three-way interaction effect between modality, list length, and presentation rate, $F(1, 887) < 1$. As expected, recent auditory items were better recalled than recent visual items, demonstrating a clear modality effect. Recency performance was also greater on 12-item lists than on 24-item lists. Faster presentation reduced recency performance, but this effect was primarily evident during 12-item trials, while performance on 24-item trials showed little effect of presentation rate. Finally, we observed stronger modality effects under conditions of faster item presentation, which resulted from faster presentation reducing performance on visual trials to a greater extent than on auditory trials.

Recall initiation. Inspired by the investigation of Grenfell-Essam et al. (2017) into recall initiation behaviors and their relation to the modality effect, we conducted an analysis of probability of first recall (PFR) in our study, and examined the impact of the first word recalled on the resultant serial position curves. Based on their PFR analysis, we analyzed the total probability of initiating recall from any of the final four list items. As probability of first recall is inherently confounded with the total number of items in the list, we chose to conduct separate 2 (Modality: Visual and Auditory) × 2 (Presentation Rate: Slow and Fast) repeated measures ANOVAs for 12- and 24-item lists. The ANOVA for 12-item lists showed no main effect of modality, $F(1, 887) = 1.92, MS_e = 0.16, p = .167, \eta_p^2 = .002$, but a significant main effect of presentation rate, $F(1, 887) = 7.17, MS_e = 0.58, p = .008, \eta_p^2 = .008$, and no interaction between modality and presentation rate, $F(1, 887) < 1$. The ANOVA for 24-item lists instead showed a significant main effect of modality, $F(1, 887) = 19.3, MS_e = 1.52, p < .001, \eta_p^2 = .021$, with a trending effect of presentation rate, $F(1, 887) = 3.21, MS_e = 0.30, p = .074, \eta_p^2 = .004$, and no interaction between the two, $F(1, 887) = 1.93, MS_e = 0.16, p = .166, \eta_p^2 = .002$. Thus, 12-item lists showed only an effect of presentation rate, while 24-item lists primarily showed an effect of modality. The effect of presentation rate was such that faster presentation increased the likelihood of initiating recall from the last four list items. The effect of modality was such that participants were more likely to initiate recall from one of the final four items on auditory trials than on visual trials, a difference which was primarily loaded on the next-to-last list item. These data are summarized in Figure 2.

In addition to analyzing the probability of first recall, we also calculated separate SPCs by recall start position for both modalities, with the goal of replicating patterns observed by both Craik (1969) and Grenfell-Essam et al. (2017), whereby modality differences tend to be largest when recall begins from the opposite end of the list. Figure 3 shows the average SPCs for visual and auditory trials, conditional on whether the participant initiated recall from the first item of the list, or from one of the final four items. We found that visual presentation produced a dramatically attenuated recency effect when recall began from the first list item, whereas auditory presentation produced a strong recency effect regardless of start position. Consequently, the modality effect was largest when recall began from the start of the list. Meanwhile, a strong inverse modality effect appeared when recall began from the end of the list. When recall began from the start of the list, an inverse modality effect was still apparent on shorter lists, though longer lists produced nearly identical primacy performance for both modalities.

Temporal contiguity. Another question of interest is whether people differ in the degree to which they temporally cluster visual versus auditory stimuli, as well as whether these differences are modulated by the presentation rate of the stimuli. To analyze temporal clustering in our study, we used the temporal clustering factor introduced by Polyn et al. (2009; see also Sederberg, Miller, Howard, & Kahana, 2010). Specifically, for each transition participants made between two correct recalls, we identified all serial positions to which they could have transitioned (repeats were not considered valid) and calculated the fraction of possible transitions with larger lags than the actual transition. Hence, a participant who always transitioned to the nearest valid item would have a clustering score of 1.0, while a participant who transitioned randomly between items would score close to 0.5. We computed each participant's temporal clustering score separately for each of the eight list types, while excluding the first two transitions of each trial from our calculations. Because list length directly affects which transitions are possible, we chose to analyze the transition data from 12-item and 24-item lists separately. Therefore, we ran a series of two 2 (Modality: Visual and Auditory) × 2 (Presentation Rate: Slow and Fast) repeated measures ANOVAs to test for differences in temporal clustering between conditions. Results for 12-item lists showed a trending effect of modality, $F(1, 795) = 3.47, MS_e = 0.080, p = .063, \eta_p^2 = .002$, and no other main or interaction effects: presentation rate, $F(1, 795) = 1.56, MS_e = 0.035, p = .212, \eta_p^2 = .004$, and modality and presentation rate, $F(1, 795) = 1.39, MS_e = 0.035, p = .239, \eta_p^2 = .002$. Results for 24-item lists also showed no significant main or interaction effects: modality, $F(1, 848) < 1$, presentation rate, $F(1, 848) = 1.23, MS_e = 0.016, p = .267, \eta_p^2 = .001$, and modality and presentation rate, $F(1, 848) < 1$. The tendency to temporally cluster items was generally unaffected by our manipulations of modality and presentation rate. A trending effect of modality did emerge, such that participants clustered auditory items to a slightly greater extent than visual items; however, this difference was isolated to short lists and should not be considered reliable.

Figure 4 summarizes these data and additionally shows the lag-conditional response probability (lag-CRP) for all list types. The lag-CRP is calculated as the probability of making a transition of a specific lag from one correct recall to another, given that such a transition was possible (Kahana, 1996). In both the lag-CRP and temporal clustering scores, the first two recall transitions have been excluded from each trial. The lag-CRPs for longer lists are markedly similar between modalities and presentation rates. Shorter lists exhibited greater variability between conditions; most notably, participants were particularly likely to make -1 transitions on slow auditory lists, and were particularly unlikely to make +1 transitions on fast visual trials.

Prior-list intrusions (PLIs) and PLI recency. Prior-list intrusions occur when a participant recalls a word that was presented during any trial before the current one. For the analysis of intrusion behavior in our study, we chose to focus solely on the effects of modality. As modality differed between trials, PLIs may have been influenced not only by their original presentation modality, but also by the modality of the trial during which the intrusion was committed. For instance, a word presented visually on one trial might later intrude on a subsequent auditory trial. As such, our PLI analyses utilized a 2 (Modality at Encoding: Visual and Auditory) × 2 (Modality at Retrieval: Visual and Auditory) within-subject design, where modality at encoding refers to the original presentation modality of the intruding word, and modality at retrieval refers to the modality of the trial where the intrusion took place. As part of our investigation of modality effects in free recall, we chose to analyze both the average number of PLIs participants made and the recency of those PLIs (cf. Zaromb et al., 2006) under each of the four modality combinations. For convenience, we will refer to the four types of PLIs by the abbreviations V-V, V-A, A-V, and A-A, where the first letter refers to the intrusion's modality at encoding and the second letter indicates the intrusion's modality at retrieval.

We conducted a 2 (Modality at Encoding: Visual and Auditory) × 2 (Modality at Retrieval: Visual and Auditory) repeated measures ANOVA to determine whether modality affected the average number of PLIs that participants made per trial. Note that not all four types of PLIs were possible on every trial. For example, if Trials 1 and 2 were visual and Trial 3 was auditory, Trial 2 could only contain PLIs of type V-V and Trial 3 could only contain PLIs of type V-A. As such, we divided each participant's total number of PLIs of each type by the number of trials on which it was possible for them to make that type of PLI. This method provides the expected number of PLIs per trial, given that such a PLI was possible. Neither the main effect of modality at encoding, nor the effect of modality at retrieval were significant, both $F(1, 887) < 1$. However, the two-way interaction between modality at encoding and retrieval was highly significant,

$F(1, 887) = 15.7, MS_e = 0.26, p < .001, \eta_p^2 = .017$. Participants were equally likely to intrusively recall visual and auditory words, and made equal numbers of prior-list intrusions on visual and auditory trials. However, participants made significantly more same-modality PLIs (V-V and A-A) than cross-modality PLIs (V-A and A-V). These results are summarized in the left panel of Figure 5.

To analyze list recency effects among PLIs, we calculated separate PLI recency curves for each of the four types of PLI. We again faced the issue of participants having variable numbers of opportunities to make each type of PLI at each lag. To account for this discrepancy, we calculated PLI recency as follows, with this procedure performed separately for each of the four PLI modality types: First, count how many PLIs of that modality type the participant made from one list back, two lists back, etc. up to the maximum possible 15 lists back. Then, for each lag, count the number of trials on which it was actually possible to make a PLI of that lag and modality type. Next, divide the PLI totals by the trial counts. This gives the expected number of PLIs from each lag, given that such a PLI is possible. Finally, divide these 15 expected values by their sum, thereby converting them to the proportion of PLIs expected to originate from each number of lists back. The resulting values make up our PLI recency curves. The right panel of Figure 5 illustrates PLI recency for each modality for list lags up to five. Note that we excluded each participant's first two trials from our PLI recency calculations.

To test for differences in PLI recency across the four modality types, we compared the proportion of PLIs expected to originate from a list lag of 1 in each condition. As participants generally made few PLIs, only a small subset of participants ($N = 60$) actually made all four types of PLIs over the course of their session. Due to this issue of missing data, we tested our PLI recency data using a linear mixed model in place of a repeated measures ANOVA. Our model incorporated modality at encoding, modality at retrieval, and the interaction between them as fixed effects, with a random intercept for each participant. Random effects of modality at encoding and retrieval were not included, as

PLI data were too sparse to model the effects of modality within individual subjects. We performed hypothesis testing using a Type III ANOVA with degrees of freedom approximated by the Satterthwaite method (Fai & Cornelius, 1996), as implemented in the lmerTest package for R (Kuznetsova, Brockhoff, & Christensen, 2017). Our model showed no effect of modality at encoding, $F(1, 1087) < 1$, a significant effect of modality at retrieval, $F(1, 1124) = 10.8, MS_e = 1.79, p = .001$, and no interaction between modality at encoding and modality at retrieval, $F(1, 1108) < 1$. On visual trials, PLIs were more likely to originate from one list back than they were on auditory trials, with the original modality of the intruding word having no influence on recency. Given that participants did not make fewer intrusions on auditory trials overall, however, it appears that while they made fewer lag-1 PLIs, they also made an increased number of remote PLIs.

Final free recall. Our primary interest with regards to final free recall (FFR) performance was whether modality would impact long-term recall probability. As FFR performance can be confounded by some words having been retrieved during previous recall periods, we analyzed FFR performance only for words that were not previously recalled by the participant. Specifically, we calculated for each participant the number of visual and auditory words that they did not recall either as a correct recall or as a PLI at any point prior to FFR, then identified the proportion of these words that they successfully recalled during FFR. We tested for modality differences in probability of FFR using a two-tailed paired t-test, comparing FFR performance for words that were presented visually to those that were presented aurally. We found a small advantage for retrieving previously unrecalled auditory words ($M = .0354, SD = .0349$) over previously unrecalled visual words ($M = .0329, SD = .0401$) during FFR, though this effect failed to reach significance, $t(887) = 1.68, p = .093$.

Discussion

Experiment 1 investigated the effects of modality, list length, and presentation rate on performance in immediate free recall. We observed modality and inverse modality effects across all combinations of list length and presentation rate, characterized by auditory presentation producing reduced primacy and improved recency performance, compared to visual presentation (Figure 1). The modality effects in Experiment 1 closely match those of Murdock and Walker (1969), with an auditory advantage that extends over at least four items and which grows in magnitude with faster item presentation. Meanwhile, our inverse modality effect findings align well with the data from Beaman (2002) and Craik (1969), who found similar visual advantages for primacy items, although mid-serial advantages have also been reported (e.g., Grenfell-Essam et al., 2017; Macken et al., 2016). Secondary to our modality effect findings, we observed that memory for both primacy and recency items was better on shorter lists than on longer lists, consistent with a recent study by Grenfell-Essam et al. (2017). This improved recall probability on shorter lists may be attributed to shorter lists having smaller search sets than longer lists (O. C. Watkins & Watkins, 1975). We also observed better primacy and recency performance on slower lists than on faster lists, although for recency items this effect was stronger on shorter lists than on longer lists. This result contradicts findings by Laughery and Pinkus (1966) and Roberts (1972), both of whom observed a larger impact of presentation rate on memory performance in longer lists.

In Experiment 1, modality and presentation rate interacted in such a way that visual performance degraded more quickly than auditory performance at faster presentation rates. Similar patterns can be observed in data presented by Laughery and Pinkus (1966) and Murdock and Walker (1969), as well. This heightened sensitivity of visual performance to presentation rate led to stronger modality effects and weaker inverse modality effects when items were presented more quickly. These results are consistent with prior literature, which has typically shown increased modality effects at faster presentation rates (Corballis, 1966;

Laughery & Pinkus, 1966; Murdock & Walker, 1969; Murray, 1966; Roberts, 1972). They also support recent findings that the inverse modality effect can be reduced or eliminated by faster item presentation (Macken et al., 2016). The hypothesis that people are more inclined or able to rehearse visual items, as compared to auditory ones (Grenfell-Essam et al., 2017; Macken et al., 2016), predicts the greater sensitivity of visual performance to presentation rate. Faster presentation reduces the time that participants have to rehearse list items, and should therefore reduce any performance advantages conferred by rehearsal. If people rely on rehearsal to a greater extent during visual trials, we would expect visual performance to be more susceptible than auditory performance to changes in presentation rate. This rehearsal account does not, however, explain why presentation rate also had a stronger impact on visual performance across recency items, where rehearsal would be expected to play a weaker role. Laughery and Pinkus (1966) attributed this effect to visual items requiring a longer encoding process than auditory items, leading to degraded encoding at fast presentation rates. One alternative possibility is that the modality × presentation rate interaction in our study was confounded by visual stimuli having reduced onscreen duration in the fast condition, while auditory recordings remained the same duration regardless of presentation rate. Similar confounds appear in Laughery and Pinkus (1966), and potentially affect Murdock and Walker (1969) and Roberts (1972), as well. However, Murray (1966) explicitly controlled for this confound and still observed that faster presentation reduced overall visual performance more than auditory performance. The true reason for the interaction between modality and presentation rate remains unclear; and, as will be seen, becomes further complicated by the lack of such a finding in Experiment 2.

In our analyses, we considered the possibility that the modality and inverse modality effects in our study might have merely originated from an increased tendency to initiate recall from the final items of auditory lists, relative to visual lists. In Experiment 1, we found that participants were indeed more likely to initiate recall from end-of-list items on long auditory lists than on long visual lists, particularly from the next-to-last item in the list (Figure 2). This result partially conflicts with those of Grenfell-Essam et al. (2017), who found no effect of modality on PFR in lists of any length between 2 and 12. Although we also found no effect of modality in 12-item lists, we did see an effect in 24-item lists. One possibility is that the effect of modality on recall initiation behavior only emerges at longer list lengths. The small size of the effect may also make it difficult to detect. Our results differ from those of Murdock and Walker (1969), as well, who observed differences by modality in the shape of the PFR curve. Whereas visual presentation produced a steep PFR curve similar to those observed for both modalities in our study, Murdock and Walker (1969) found that auditory presentation instead produced a PFR curve that plateaued across the final four list items. We agree with the assessment of Grenfell-Essam et al. (2017) that the unique PFR findings of Murdock and Walker (1969) may be attributable to participants adopting different recall strategies when list length is constant and well-learned. Although we found that modality did influence recall initiation, we also showed that the modality effect persisted regardless of recall start location (Figure 3), a result which past research overwhelmingly supports (Cowan et al., 2002; Craik, 1969; Grenfell-Essam et al., 2017; Harvey & Beaman, 2007). Therefore, modality differences in recall initiation are insufficient to explain away the modality and inverse modality effects we observed. Our results also cannot be explained by an output interference account alone (e.g., Cowan et al., 2002), as differences in output interference could not have affected the first item recalled.

We supplemented these more traditional modality effect analyses with a series of additional tests, which included investigations of temporal contiguity, prior-list intrusions, and final free recall performance. Our study found at most a marginal effect of modality on the amount of temporal contiguity in participants' recalls, with clustering on short auditory trials trending higher than clustering on short visual trials (Figure 4). Moreover, we found no effect of presentation rate on contiguity. Prior literature is somewhat limited with regards to these two factors. Healey, Long, and Kahana (in press) found stronger temporal clustering and greater forward asymmetry during auditory trials than during visual trials in the data from Murdock and Walker (1969), with no effect of presentation rate; in contrast, Kahana (1996) observed greater temporal clustering at slower presentation rates in data from Murdock (1962) and (Murdock & Metcalfe, 1978), perhaps attributable to the longer presentation intervals used in these two studies. Given our currently limited understanding of the influences of modality and presentation rate on temporal contiguity, future modality effect research should continue to explore not only recall performance, but the item-to-item transitions that participants make. Understanding if and how modality impacts the organization and associations between items will help to constrain present and future theories of the modality effect in free recall.

Our prior-list intrusion results (Figure 5) consisted of four main findings. First, visual and auditory words were equally likely to intrude on subsequent lists. Second, visual and auditory trials produced equal numbers of PLIs. Third, words were more likely to intrude on trials of the same modality than on trials of a different modality. Finally, PLIs tended to originate from more recent lists on visual trials than on auditory trials. Our first three results replicate those previously observed in a paired-associate learning task by Murdock (1966), who similarly found a greater tendency to make same-modality PLIs than cross-modality PLIs, with no difference in the total number of PLIs by modality. Although Roberts (1972) found seemingly contradictory results, in which auditory presentation produced more intrusions than visual presentation, this discrepancy may be attributable to his analysis including extra-list intrusions (recalled items that were never actually presented). The tendency to make same-modality intrusions over cross-modality intrusions suggests that trials of the same modality are in some manner more confusable than trials of different modalities. This effect could be driven by the encoding of modality-specific information, as captured in the Nairne (1990) feature model. Alternatively, it may result from the coding of source features, whereby words are labeled or associated with the type of task in which they appeared (e.g., Polyn et al., 2009). Our fourth finding is a novel discovery which suggests that visual trials are more confusable than auditory trials with their immediately preceding lists. Yet, it also appears that visual trials produce fewer PLIs from more remote lists than do auditory trials, resulting in no overall modality difference in the total number of PLIs. Current theories of the modality effect struggle to account for this finding. Temporal distinctiveness theory (Glenberg & Swanson, 1986) is perhaps the only account that makes a concrete prediction regarding modality effects on PLIs. A temporal distinctiveness account would suggest that the sparse temporal coding of visual words should result in a greater likelihood of those words intruding in subsequent trials' search sets. As a result, visual words should intrude more often than auditory words, and possibly exhibit greater recency. However, we found that the modality of the intruding word affected neither the total number of PLIs, nor their recency; rather, it was the modality of the present trial that determined the recency of intrusions. One explanation is that increased PLI recency on visual trials might arise as a correlate of the inverse modality effect. Under a retrieved-context framework, reinstating the context of early-list items might increase the probability of recalling items from the preceding trial, due to contextual similarities between the previous and current list. Hence, increased recall of primacy items on visual trials might induce more lag-1 PLIs. We expand upon this hypothesis later, in developing our retrieved-context account of the modality effect.

Our final analysis focused on the long-term recallability of visual and auditory items in a final free recall task. We found only a weak trend towards an auditory advantage for retrieval of previously unrecalled items during final free recall; however, floor effects may have limited the power of our analysis. We therefore believe that additional analyses of modality effects in final free recall should be considered in future research, though such an investigation lies beyond the scope of the present work.

Experiment 2

Although within-subjects experimental designs generally offer greater statistical power than between-subjects designs, they also allow participants to adjust their strategies to jointly optimize performance across experimental conditions. Due to the cost of switching strategies between lists, participants in a within-subjects experiment may adopt strategies that tend to reduce differences between modality conditions. Prior work has established the influence of experimental design on a number of widely studied memory phenomena (e.g., McDaniel & Bugg, 2008; Serra & Nairne, 1993; Zaromb & Roediger, 2009). Given that our initial study uncovered several new aspects of the modality effect, particularly with respect to probability of first recall and the recency of prior-list intrusions, we sought to replicate these results in a between-subjects design. Therefore, in Experiment 2, half of our participants completed all trials visually, while the other half completed all trials aurally. All other manipulations were identical to those in Experiment 1.

Methods

Participants. All participants were recruited through Amazon's Mechanical Turk. Qualification criteria were identical to those used in Experiment 1. 2000 participants (657 male, 948 female, 395 not reported) completed Experiment 2, and were paid $7.50 for completion of the task.

We excluded 7 participants from our analyses due to technical problems that disrupted their sessions. We excluded an additional 524 participants who met at least one of the five criteria defined in Experiment 1. After exclusion, 1469 participants (459 male, 640 female, 370 not reported) contributed to our analyses. 735 of these participants received the visual presentation condition, while the remaining 734 received the auditory condition. Ages ranged from 18 to 71 years, with a mean age of 36.0 years ($SD = 11.0$).

Materials. All materials were identical to those in Experiment 1.

Design. The design matched that of Experiment 1, with the exception that modality was manipulated between-subjects.

Procedure. The procedure for Experiment 2 was identical to that of Experiment 1, with the exception that each participant was randomly assigned to one of two modality conditions (visual or auditory). Instead of eight visual and eight auditory lists, each participant received the same presentation modality for all 16 experimental trials, as well as for both practice trials.

Results

As in Experiment 1, we have organized our results into sections, based on the type of analysis conducted. We again begin by reporting serial position effects, followed by effects on recall start position and the resulting recall performance. We then present our temporal clustering analysis, followed by our prior-list intrusion findings, and end with some preliminary final free recall results.

To foreshadow our findings, the between-subject modality manipulation preserved most of the key findings obtained in Experiment 1. We again observed prominent modality and inverse modality effects, though they no longer exhibited their interactions with presentation rate. We also identified further support for an increased tendency to initiate recall from the last several items of auditory lists, relative to visual lists. Modality again showed little impact on temporal clustering or the number of intrusions committed, while auditory trials continued to show reduced PLI recency. Finally, we were able to substantiate the previously observed trend towards an auditory advantage during final free recall.

Primacy and recency effects. Figure 6 shows the serial position curves for all eight combinations of modality, list length, and presentation rate in Experiment 2. The data show visually similar modality and inverse modality effects under all combinations of list length and presentation rate. We analyzed differences in recall for primacy items across conditions using a 2 (Modality: Visual and Auditory) × 2 (List Length: 12 and 24) × 2 (Presentation Rate: Slow and Fast) mixed-factorial ANOVA, with modality as a between-subject factor and list length and presentation rate as within-subject factors. Matching Experiment 1, we averaged the recall probabilities of the first three list items and used the resulting value as our measure of recall for primacy items. Results showed significant main effects of modality, $F(1, 1467) = 11.9, MS_e = 1.97, p = .001, \eta_p^2 = .008$, list length, $F(1, 1467) = 568, MS_e = 14.4, p < .001, \eta_p^2 = .279$, and presentation rate, $F(1, 1467) = 153, MS_e = 3.50, p < .001, \eta_p^2 = .094$. All two-way and three-way interaction effects were non-significant: modality and list length, $F(1, 1467) < 1$, modality and presentation rate, $F(1, 1467) = 1.56, MS_e = 0.36, p = .212, \eta_p^2 = .001$, list length and presentation rate, $F(1, 1467) < 1$, and modality, list length, and presentation rate $F(1, 1467) <$. Again, we found an inverse modality effect similar to that of Beaman (2002) and Craik (1969), such that primacy items were better remembered when words were presented visually than when they were presented aurally. Also consistent with Experiment 1, we found greater primacy performance on 12-item lists than on 24-item lists, as well as greater performance with slow presentation than with fast. In contrast with Experiment 1, the inverse modality effect was not attenuated by faster presentation. Visual and auditory primacy performance were instead affected equally by presentation rate in Experiment 2.

As in Experiment 1, we calculated a recency performance score for each participant and each condition by averaging across the recall probabilities of the last five list items. To test for differences in recall for recency items, we conducted a 2 (Modality: Visual and Auditory) × 2 (List Length: 12 and 24) × 2 (Presentation Rate: Slow and Fast) mixed-factorial ANOVA on these scores, with modality as a between-subject factor, and list length and presentation rate as within-subject factors. We found significant main effects of modality, $F(1, 1467) = 66.5, MS_e = 6.55, p < .001, \eta_p^2 = .043$, list length, $F(1, 1467) = 697, MS_e = 12.9, p < .001, \eta_p^2 = .322$, and presentation rate, $F(1, 1467) = 52.3, MS_e = 0.58, p < .001, \eta_p^2 = .034$. The two-way interaction effect between list length and presentation rate was also significant, $F(1, 1467) = 18.5, MS_e = 0.21, p < .001, \eta_p^2 = .012$. The interaction effects between modality and list length, $F(1, 1467) = 1.90, MS_e = 0.035, p = .168, \eta_p^2 = .001$, and modality and presentation rate, $F(1, 1467) = 0.041, MS_e < .001, p = .839, \eta_p^2 < .001$, were not significant, nor was the three-way interaction effect between modality, list length, and presentation rate, $F(1, 1467) < 1$. Consistent with Experiment 1, recall of recency items was greater for auditory stimuli than for visual stimuli, greater on 12-item lists than on 24-item lists, and greater with a slower presentation rate than with a faster one. Also consistent with Experiment 1, we observed an interaction between list length and presentation rate, such that presentation rate had a stronger impact on recency performance in 12-item lists than in 24-item lists. The interaction we observed in Experiment 1 between the modality effect and presentation rate did not reappear in Experiment 2. Whereas visual performance decreased more than auditory performance with faster presentation in Experiment 1, performance decreased equally in both modalities during Experiment 2.

Recall initiation. Based on work by Grenfell-Essam et al. (2017), we again analyzed the tendency of participants to initiate recall from any one of the final four list items, with separate analyses for 12- and 24-item lists. Results are summarized in Figure 7. For Experiment 2 we conducted separate 2 (Modality: Visual and Auditory) $\times$ 2 (Presentation Rate: Slow and Fast) mixed-factorial ANOVAs for 12- and 24-item lists, with modality as a between-subject factor and presentation rate as a within-subject factor. The ANOVA for 12-item lists showed a significant main effect of modality, $F(1, 1467) = 5.28, MS_e = 1.03, p = .022, \eta_p^2 = .004$, with no effect of presentation rate, $F(1, 1467) < 1$, and no interaction between modality and presentation rate, $F(1, 1467) = 2.86, MS_e = 0.11, p = .091, \eta_p^2 = .002$. The ANOVA for 24-item lists also showed a significant main effect of modality, $F(1, 1467) = 7.53, MS_e = 1.44, p = .006, \eta_p^2 = .005$, with no effect of presentation rate,

$F(1, 1467) = 1.06, MS_e = 0.043, p = .303, \eta_p^2 = .001$, and no interaction between the two, $F(1, 1467) = 1.36, MS_e = 0.055, p = .243, \eta_p^2 = .001$. Across both list lengths, auditory presentation increased the likelihood of initiating recall from the last four list items relative to visual presentation, similar to what we observed on long lists in Experiment 1. As is evident from the PFR curves in Figure 7, this effect was again strongest for the next-to-last list item. The effect of presentation rate that we observed on short lists in Experiment 1 was not replicated in Experiment 2.

Figure 8 shows the resultant serial position curves when participants initiated recall from the first list item, versus any one of the final four list items. Consistent with Experiment 1, the recency effect was dramatically attenuated on visual trials where participants initiated recall from the start of the list, whereas auditory trials exhibited a strong recency effect regardless of start position. This pattern again resulted in a substantial modality effect on trials where recall began from the first item of the list, and a relatively small modality effect on trials where recall began from the end of the list. Inverse modality effects predominantly arose when participants initiated recall from the end of the list, with no such effects appearing when recall initiated from the start of the list. Unlike those in Experiment 1, participants in Experiment 2 exhibited a persistent auditory advantage across all mid-serial and recency positions on trials where they initiated recall from the start of the list.

Temporal contiguity. We again used the temporal clustering factor of Polyn et al. (2009) to analyze clustering behavior in Experiment 2. As in Experiment 1, we computed each participant's temporal clustering score separately for each list type, while excluding the first two transitions of each trial. Figure 9 shows the resulting temporal clustering scores, as well as lag-conditional response probabilities (Kahana, 1996) for all list types. Given that list length directly influences which transitions are possible for the participant to make, we again ran separate ANOVAs for 12-item and 24-item lists. Therefore, to test for differences in temporal clustering between conditions, we ran a series of two 2 (Modality: Visual and Auditory) $\times$ 2 (Presentation Rate: Slow and Fast) mixed-factorial ANOVAs, with modality as a between-subject factor and presentation rate as a within-subject factor. Results for 12-item lists showed a trending effect of modality, $F(1, 1441) = 3.39, MS_e = 0.088, p = .066, \eta_p^2 = .002$, a significant main effect of presentation rate, $F(1, 1441) = 5.16, MS_e = 0.064, p = .023, \eta_p^2 = .004$, and no interaction between modality and presentation rate, $F(1, 1441) < 1$. Results for 24-item lists also showed a non-significant trend for modality, $F(1, 1452) = 2.82, MS_e = 0.061, p = .093, \eta_p^2 = .002$, a significant main effect of presentation rate, $F(1, 1452) = 4.19, MS_e = 0.027, p = .041, \eta_p^2 = .003$, and no interaction between the two, $F(1, 1452) < 1$. For both list lengths, the auditory modality showed a trend towards increased temporal clustering. Furthermore, a faster presentation rate resulted in slightly reduced temporal clustering relative to slower presentation, an effect which we did not observe in Experiment 1. The lag-CRPs generally show few differences between modalities and presentation rates, with slow, 12-item auditory trials again exhibiting a notably higher probability of -1 transitions than other 12-item trials.

Prior-list intrusions and PLI recency. Similar to Experiment 1, we chose to focus our prior-list intrusion analyses solely on the effects of modality. The experimental design was simplified in Experiment 2 by the fact that the presentation modality of an intruding word always matched the modality of the trial during which the intrusion was committed. As such, there were only two modality conditions (Visual and Auditory) for PLIs in Experiment 2, replacing the four conditions present in Experiment 1.

We compared the average number of PLIs made by participants in the visual condition to the average number of PLIs made by participants in the auditory condition using a two-tailed independent t-test, and found no significant difference, $t(1467) = 1.34, p = .181$. Participants completing the visual recall task tended to make the same number of PLIs as those completing the auditory task. The left panel of Figure 10 shows the average number of PLIs per trial made by participants in both versions of the task.

In an identical manner to Experiment 1, we calculated for each participant the proportion of prior-list intrusions that originated from varying numbers of trials back. The right panel of Figure 10 shows the resulting PLI recency curves from one to five lists back for participants in either modality condition. To test for modality differences in the recency of PLIs, we used a two-tailed independent t-test to compare the proportion of PLIs expected to originate from one list back. Results showed that intrusions committed by participants in the visual task were significantly more likely than those of participants in the auditory task to originate from one list prior, $t(1029) = 3.64, p < .001$. However, given that participants in both tasks committed the same number of PLIs in total, it appears that this increased tendency to make recent PLIs was again accompanied by a reduction in remote PLIs.

Final free recall. As in Experiment 1, we investigated whether modality impacted the number of previously unrecalled words that participants remembered during final free recall. For Experiment 2, we conducted a two-tailed independent t-test to compare the FFR performance of participants who studied words visually to those who studied words aurally. We found that participants retrieved a significantly greater proportion of previously unrecalled auditory words ($M = .0374, SD = .0337$) than previously unrecalled visual words ($M = .0326, SD = .0318$), $t(1467) = 2.78, p = .005$. This finding helps to substantiate the trending difference we observed in Experiment 1.

Discussion

Experiment 2 sought to replicate the results of our first experiment in a between-subjects paradigm, in which participants completed all of their trials in one modality. We again observed modality and inverse modality effects across both presentation rates and list lengths, characterized by participants in the auditory task showing better recall for end-of-list items than participants in the visual task, but worse recall for primacy items (Figure 6). Consistent with Experiment 1, overall recall probability also increased with a slower presentation rate and shorter list length. We again found that presentation rate affected recency performance to a greater extent on short lists than on long lists. However, whereas participants in Experiment 1 showed larger modality effects and attenuated inverse modality effects on faster trials, we observed no modulation of these effects with presentation rate in Experiment 2. The disappearance of this interaction suggests that the effect may have depended on the specific recall strategies favored by participants in Experiment 1, when modality alternated randomly between trials.

In this experiment we found a consistent effect of modality on participants' tendency to initiate recall with end-of-list items, regardless of list length and presentation rate. Auditory presentation increased the tendency to initiate recall from the end of the list, as compared with visual presentation (Figure 7). In Experiment 1, we also observed an auditory advantage for initiating recall from recent list items, but this effect was significant only in the long-list condition. Despite the lack of past evidence for a modality effect on PFR (Grenfell-Essam et al., 2017), our data suggest that a small, but reliable effect does exist. However, we found that this increased probability of initiating recall from end-of-list items on auditory trials was insufficient to explain the modality effects in our study. Both of our experiments demonstrated that the modality effect persisted irrespective of recall start position (Figures 3 and 8). Our observations support the results of prior studies, which have largely reached similar conclusions regarding the persistence of the modality effect across start positions (Cowan et al., 2002; Craik, 1969; Grenfell-Essam et al., 2017; Harvey & Beaman, 2007). We also observed in both studies that the inverse modality effect was stronger when recall started from the end of the list, than when it started from the beginning. In Experiment 2 we additionally found that, when initiating recall from the beginning of the list, participants in the auditory condition exhibited better memory than those in the visual condition for all mid-serial and end-of-list words. We observed no such generalized auditory advantage in Experiment 1, further suggesting that there may have been differences in recall strategies or task engagement between our two experiments. We were unable to replicate our secondary finding from Experiment 1, whereby fast presentation increased the tendency to initiate recall from end-of-list items on short lists.

Similar to Experiment 1, our temporal contiguity analysis showed modality having a marginal effect on the degree to which recalls clustered by presentation order, with auditory clustering trending higher than visual clustering (Figure 9). Although this was a weak effect, recent findings by Healey et al. (in press) support its validity and warrant further investigation of potential modality effects on temporal contiguity. In addition, we found slightly greater temporal clustering on slow trials, relative to fast trials. Though we did not observe this effect of presentation rate on contiguity in Experiment 1, our result is consistent with data from Murdock (1962) and Murdock and Metcalfe (1978), as analyzed by Kahana (1996). Experiment 2 reaffirms the need for further investigation into recall transition behavior in future modality studies, in order to better inform and constrain our theories of the modality effect.

The results of our prior-list intrusion analyses (Figure 10) were largely consistent between our two experiments. In Experiment 2, we again found that visual and auditory presentation produced equal numbers of prior-list intrusions. Furthermore, we found that PLIs made by participants in the visual task were more likely to originate from one list back than were those of participants in the auditory task. This result supports the similar PLI recency findings of Experiment 1, in which PLIs on visual trials were more likely to originate from one list back than were PLIs on auditory trials. Based on our observations from our first experiment, it is likely that the effect in Experiment 2 was driven by the modality at retrieval, not the original modality of the intruding words.

In Experiment 2, we again included a final free recall component, designed for the investigation of modality effects in long-term retrieval. Results showed that participants who studied all lists aurally retrieved more previously unrecalled words during final free recall than participants who studied all lists visually. Whether this difference stems from a true long-term auditory recall advantage or simply from improved task engagement remains unclear. Although only a preliminary result, this finding invites a deeper analysis of modality effects in final free recall.

A Retrieved-Context Model of Modality Effects

Equipped with the data from our two experiments, we next investigated whether the interlist version of the Context Maintenance and Retrieval model (CMR2; Lohnas et al., 2015) could account for the varied effects of modality on the serial position curve, the probability of first recall, and the tendency to commit prior-list intrusions. In this pursuit, we conducted two sets of simulations. We first assessed CMR2's fit to the average behavioral data from Experiment 2 without regard for the effects of modality. We then tuned the model to specifically simulate the differences in recall probability between modality conditions, and examined its predictions with respect to the variety of effects in our study. We chose to focus our modeling analyses on the between-subject modality effect data (Exp. 2), as it would not require us to define separate parameter sets for different lists within a session. Furthermore, this method avoided the complexities of analyzing recall in a paradigm where an item encoded in one modality can intrude on a trial of a different modality. Given that we observed modality and inverse modality effects for both shorter and longer lists, and for both presentation rate conditions, we simplified our modeling analysis by pooling across presentation rates and only fitting data from the shorter (12-item) lists.

In directing our modeling approach, we asked how we might translate the processes implicated by existing modality effect theories into a retrieved-context framework. We focused on two of these concepts, in particular: the idea that auditory items are more temporally discriminable than visual items (Glenberg & Swanson, 1986), and the idea that auditory stimuli are more richly encoded than visual stimuli (Nairne, 1990). The closest analogue to temporal distinctiveness in a retrieved-context model is the degree to which the contexts of nearby items overlap, a property that is modulated by the rate at which context changes as new items are presented. Faster context drift reduces the overlap between the representations of studied items, as previous contexts decay more quickly. Consequently, faster drift rates produce greater temporal discriminability. In our simulation of the modality effect, we therefore allowed the context drift rate to vary between modalities.

As CMR2 makes the simplifying assumption that each item is represented by a single feature (see below), we did not directly simulate Nairne's (1990) hypothesis that auditory stimuli carry a greater number of modality-specific features than do visual stimuli. Because of this restriction, we instead used context-to-item association strengths to represent richness of encoding. In our model, we translate richer encoding to the formation of stronger associations, which allow items to be more strongly cued by their contexts during retrieval, thereby improving the recallability of those items. Thus, in addition to drift rate, we also allowed the strength of learned context-to-item associations to vary by modality.

Model Overview

The Context Maintenance and Retrieval model is predicated on the idea that items in memory form associations with the contexts in which they appear. Context itself is defined as the presently active set of representations in the brain, which may include physical and temporal information about the present environment, as well internal information about one's recent thoughts, goals, and emotions. In the CMR2 model, items are represented by the distribution of activation across the elements of a feature vector, $\mathbf{f}$, and context is represented by the distribution of activation over a context vector, $\mathbf{c}$. Like Lohnas et al. (2015), we make the simplifying assumption in our simulations that the i^{th} presented item is represented by a single element of the feature vector, $\mathbf{f}_i$. The feature and context layers of the model interact via a pair of associative matrices, M^{FC} and M^{CF}. Each time an item is presented to the model, it triggers a reactivation of any contextual states with which that item has been associated, causing a shift in the present context. Formally, the presentation of the i^{th} item creates a new input to context,

$$\mathbf{c}_i^{\mathbf{IN}} = \frac{M^{FC}\mathbf{f}_i}{||M^{FC}\mathbf{f}_i||}, \tag{1}$$

following which context evolves according to the equation:

$$\mathbf{c}_i = \rho_i \mathbf{c}_{i-1} + \beta \mathbf{c}_i^{\mathbf{IN}}. \tag{2}$$

β is a model parameter that determines the degree to which context changes with each studied item, with larger values of β producing faster decay of previous contextual states. ρ_i is a constant defined such that $||\mathbf{c}_i|| = 1$. Context is therefore modeled as a recency-weighted sum of contextual states. The rate of context drift may differ between encoding and recall events, defined by the β_{enc} and β_{rec} parameters, respectively.

Meanwhile, the features of the presented item form associations with the state of the context vector at the time of presentation, $\mathbf{c}_{i-1}$, updating the associative matrices according to a standard Hebbian learning rule:

$$\begin{aligned} \Delta M^{FC} &= \mathbf{c}_{i-1}\mathbf{f}_i^T \\ \Delta M^{CF} &= \mathbf{f}_{i-1}\mathbf{c}_i^T \end{aligned} \tag{3}$$

Learned context-to-item associations in M^{CF} are further scaled by a primacy factor, ϕ, which represents the decay of attention over the course of a trial. Specifically, the scaling factor for the k^{th} item of a list is calculated as follows:

$$\phi_k = \phi_s e^{-\phi_d(k-1)} + 1, \tag{4}$$

where ϕ_s and ϕ_d are model parameters governing the strength and decay of the primacy gradient, respectively.

The strength of newly learned associations, relative to any pre-experimental associations is determined by the model parameters γ_{FC} and γ_{CF}, which influence updates to the item-to-context (M^{FC}) and context-to-item (M^{CF}) associative matrices, respectively. Greater values of γ produce stronger associations between items and their new contexts, while lower values of γ result in weaker learning and a stronger weighting of preexisting associations. CMR2 represents the semantic associations between items by scaling the initial values of M^{CF}, such that semantically related items can be cued by the same context states. The model's semantic scaling parameter, s, further modifies the strengths of these semantic associations, with higher values of s increasing the degree to which semantic similarity influences recall. In our study, we defined the semantic association between each pair of words as the cosine similarity between the Word2vec vector representations of those words (Mikolov, Chen, Corrado, & Dean, 2013).

At the time of recall, the current context acts as a retrieval cue, producing a vector $\mathbf{f^{IN}} = M^{CF}\mathbf{c}_i$, which serves as the input to a leaky, competitive accumulation process that determines which item the model recalls (Usher & McClelland, 2001). The value of the accumulator at step n is determined by:

$$\begin{aligned} \mathbf{x}_n &= (1 - \tau\kappa - \tau\lambda N)\mathbf{x}_{n-1} + \tau\mathbf{f^{IN}} + \varepsilon, \\ \varepsilon &\sim \mathcal{N}(0, \eta), \\ \mathbf{x}_n &\rightarrow max(\mathbf{x}_n, 0). \end{aligned} \tag{5}$$

Each element of $\mathbf{x}$ corresponds to a single element in $\mathbf{f^{IN}}$. τ is a time constant, κ is a leak parameter, and η is a noise parameter. λ is a parameter that controls lateral inhibition, by scaling the strength of an inhibitory matrix N that connects the accumulator for each item to those of all other items. This process runs iteratively until one of the accumulating elements crosses a threshold, Θ_i (see Equation 6), or until the recall period is over. When an item wins the recall competition, it is considered to have been retrieved from memory. As a method of filtering out undesirable recalls (e.g., prior-list intrusions), the model compares the current context to the context associated with the retrieved item, $u = \mathbf{c}_{t+1}^{\mathbf{IN}} \cdot \mathbf{c}_t$, and overtly recalls the item only if u exceeds a threshold defined by the model parameter c_{thresh}. Regardless of whether the retrieved item was filtered out or recalled, that item is then presented to the model, updating context according to Equation 2 with a drift rate defined by β_{rec}. The updated state of context activates a different set of features on $\mathbf{f^{IN}}$, and the recall competition begins again.

CMR2 incorporates repetition suppression by increasing an item's retrieval threshold to a maximum value immediately after it has been retrieved. The threshold then decreases back towards the baseline value of $\Theta_i = 1$, as a function of the number of subsequently retrieved items, j:

$$\Theta_i = 1 + \omega\alpha^j \quad (6)$$

The model parameter ω controls the degree of repetition suppression, while the parameter α controls the duration of suppression.

Between the end of a recall period and the start of the next list, CMR2 simulates the change in study mode by presenting an additional item to the model in accordance with Equation 2 and a drift rate of β_{post}^{recall}. Unlike item presentations during the encoding period, this post-recall item does not form associations with context and never enters into the recall competition. Hence, the model undergoes a shift in context between trials without encoding any new information.

Methods

Fitting the average data. To identify a set of parameters under which CMR2 fits the average data from our study, we used a particle swarm optimization (PSO) algorithm to search the model's parameter space (Kennedy & Eberhart, 1995). Specifically, we used the constriction factor PSO algorithm of Eberhart and Shi (2000), with 2,000 particles updated and evaluated for 100 iterations (see also Bratton & Kennedy, 2007; Clerc & Kennedy, 2002). We evaluated the parameter set associated with each particle's location at each iteration by using it to simulate 500 sessions (8,000 trials) of Experiment 2. An identical set of 500 sessions was simulated by all particles on all iterations. Simulated sessions were derived from actual sessions, using the exact sets of word lists presented to our participants. The only exception was that we held list length constant at 12 items in our simulated sessions, as noted above. In order to do this, we presented our models with only the first 12 items of each experimental trial, thereby treating 24-item lists as if they were 12 items long. We then compared the simulated recall data from each particle to the average experimental data from all 12-item lists in Experiment 2, pooled across modality and presentation rate conditions. Goodness-of-fit was determined as the root-mean-square deviation (RMSD) between model and data for the following 30 data points: All 12 points of the serial position curve, the probability of first recall for all 12 serial positions, the average number of prior-list intrusions per trial, and the PLI recency score for list lags of 1 through 5. In our RMSD calculations, we double-weighted and triple-weighted several key data points which were particularly critical for the model to fit. The triple-weighted data points were the first and twelfth position of the SPC, as well as the PLI recency score for a list lag of 1. The double-weighted points were the second, third, tenth, and eleventh positions of the SPC, as well as the first and last serial positions of the PFR curve. Based on this goodness-of-fit metric, each particle maintained a record of its own best-fitting parameter set, as well as the best parameter set identified by any particle in the swarm.

After completing 100 iterations of PSO, we identified the best parameter set discovered by each particle. In order to more precisely distinguish between models with similarly good fits, we used each of these 2,000 parameter sets to simulate all 1,469 sessions of Experiment 2 three times (totaling 70,512 trials). We then recalculated the RMSD of each model based on this more precise simulation, and selected the parameter set with the lowest RMSD as our best overall fit to the average data.

Fitting the modality effect. Starting from this more general model of our data, we next endeavored to simulate the effects of modality in our study by systematically varying the rate of context drift, β_{enc}, and the strength of newly formed context-to-item associations, γ_{CF}, while holding all other parameters constant. We conducted a two-dimensional grid search centered on the values of β_{enc} and γ_{CF} from our best overall fit, allowing β_{enc} to vary by up to ± 0.1 and γ_{CF} to vary by up to ± 0.25. Within this search space, we selected parameter values evenly spaced at 80 intervals along each dimension, producing a total of 6,400 unique parameter sets. We used each of these parameter sets to simulate all 1,469 sessions of Experiment 2 five times (totaling 117,520 trials), and calculated each model's average serial position curve across all sessions. We then computed the RMSD between each of these 6,400 serial position curves and the empirical visual and auditory serial position curves from our study. Again, we triple-weighted the first and twelfth positions of the SPC and double-weighted the second, third, tenth, and eleventh positions. For each modality, we selected the combination of β_{enc} and γ_{CF} that produced the lowest RMSD as our best-fitting modality-specific model. We thus identified one parameter set that best simulated visual recall and one parameter set that best simulated auditory recall.

Results

We first review the overall fit of CMR2 to the average data from Experiment 2, collapsing across modalities. We next analyze the model's ability to simulate the classic modality and inverse modality effects as a function of the context drift rate, β_{enc}, and the strength of learned context-to-item associations, γ_{CF}. Finally, we examine the model's predictions regarding the effects of modality on recall initiation, intrusion behavior, and performance by recall start position, to evaluate their alignment with our empirical findings.

The leftmost column of Table 1 lists the set of parameters that best fit the overall average data; also shown are the associated RMSD scores. The behavioral predictions of the best overall model can be found in Figure 11. CMR2 closely simulated the average serial position curve, as well as the quantity and recency of prior-list intrusions. However, the model was less accurate at simulating the probability of first recall, and particularly struggled to account for participants' tendency to initiate recall from the first list item ($PFR_{model} = 4.0\%; PFR_{data} = 18.5\%$).

The heat maps in Figure 12 illustrate the results of the grid search we conducted over

β_{enc} and γ_{CF} values when fitting our modality-specific models, and reflect the goodness-of-fit for each modality at each point in the parameter space. The heat maps show a convex surface with global minima for both modalities lying near the diagonal of the parameter space, indicating that a two-parameter account can better explain the modality effect than can either one-parameter account (models along the dotted lines). The second and third columns of Table 1 show the values of β_{enc} and γ_{CF} that best fit the visual and auditory serial position curves, respectively, as well as the associated RMSD scores. As predicted, performance in the auditory task was best simulated by a faster drift rate at encoding and strengthened context-to-item associations ($\beta_{enc}^{aud} = 0.535, \gamma_{CF}^{aud} = 0.707$), relative to performance in the visual task ($\beta_{enc}^{vis} = 0.489$, $\gamma_{CF}^{vis} = 0.543$). The first row of Figure 13 compares the serial position curves generated by our modality-specific models to those observed empirically in Experiment 2. With an increase in β_{enc} and γ_{CF}, CMR2 demonstrated a recall advantage for recent items and a recall disadvantage for early-list items, encapsulating both the modality and inverse modality effects. These simulated effects were similar in extent and magnitude to those seen in our data, with the modality effect extending over the final four list items and the inverse modality effect extending over the first two items.

By fitting our modality-specific models to serial position effects only, we allowed these models to make their own predictions with respect to the range of additional effects identified in our study. The models' predictions regarding probability of first recall, prior list intrusions, and recall performance by start position are shown in Figures 13 and 14. CMR2 successfully predicted the increased probability of initiating recall from the end of an auditory list, relative to a visual list; however, it overestimated the size of this effect, particularly at the final serial position. Meanwhile, it closely predicted the modality differences we found in PLI recency, with prior-list intrusions made by the visual model tending to originate from more recent trials than those made by the auditory model. Furthermore, this difference correctly occurred without an accompanying change in the total number of prior-list intrusions committed. On trials where recall began from the first list item, the auditory model showed higher recall probability than the visual model at nearly all serial positions, which was an effect we observed in Experiment 2, though not Experiment 1 (see Figures 3 and 8). Meanwhile, on trials where recall began from one of the final four list items, CMR2 correctly predicted an auditory advantage for recent items and a visual advantage for primacy items.

Discussion

Our modeling results support the account that the modality and inverse modality effects in free recall arise as a product of increased context drift rates and stronger context-to-item associations under conditions of auditory presentation, relative to visual presentation. Using the interlist implementation of the Context Maintenance and Retrieval model (Lohnas et al., 2015), we successfully simulated both of these effects in magnitude and extent. We now consider how differences in drift rate and context-to-item associations might come together to produce the effects we observed in our experiments and simulations. First, we explain the mechanisms behind the modality and inverse modality effects. Then, we explain the additional effects of modality we observed on prior-list intrusion behavior, performance by recall start position, and recall initiation. Finally, we discuss why these two key parameters might be expected to differ between modalities.

The primary consequence of faster context drift during encoding is a reduced overlap between the contexts of all items, effectively increasing the distances between them. This results in all items' contexts lying further from the end-of-list context than they would under conditions of slower context drift. This added distance skews recall towards the end of the list, attenuating the primacy effect and reducing the recallability of all prerecency items. The end result is a steeper recency curve that drops to a lower asymptote and rises to a relatively shallow primacy curve. On its own, a faster context drift rate can therefore account for the differences in slope between the visual and auditory serial position curves.

However, it also predicts auditory disadvantages for all early and mid-serial items, whereas our data show the inverse modality effect being isolated only to the first few list items. Critically, enriched context-to-item binding during auditory presentation counterbalances the general performance drop caused by an increased drift rate. Strengthened associations improve retrieval cueing, thereby boosting performance across all serial positions and negating the drop in the asymptote of the serial position curve. The end result is that the increased slope of the SPC across recency items appears as an auditory recency advantage (i.e., the modality effect) and the decreased slope of the SPC across primacy items appears as an auditory primacy disadvantage (i.e., the inverse modality effect). Meanwhile, no modality differences appear among mid-serial items, as the two effects balance out.

Our retrieved-context model of the modality effect also accounts for the reduced recency of prior-list intrusions during auditory trials. The mechanisms underlying this effect are twofold. First, an increased drift rate during auditory presentation increases the distance between the end-of-list context on an auditory trial and the context of the previous trial. Hence, context at the start of recall has reduced similarity to the context of items on the previous list, resulting in a lower baseline probability of those items intruding. The inverse modality effect augments this effect even further. Recalling early-list items reinstates the context that was present at the start of the trial, which is itself relatively similar to the previous trial's context. Therefore, better primacy performance can make the current trial more confusable with the last, which would result in more words from the previous trial intruding. As a consequence of these two influences, lag-1 PLIs occur more often on visual trials than on auditory trials.

That being said, faster context drift during auditory trials should result in the context at recall being less confusable with the contexts of all preceding trials, not just the immediately preceding one. Hence, one might expect fewer PLIs of all lags on auditory trials. Instead we found – and the model predicted – reduced PLI recency on auditory trials with no difference in the total number of PLIs. Therefore, although auditory trials produced fewer recent PLIs than did visual trials, they also produced an increased number of remote PLIs. To explain this behavior, consider which items tend to be cued most strongly by the context at recall. Typically, these will be items from the current list, items from recent lists, and items which are semantically related to items from the current list. When recalled, items from recent lists manifest as PLIs with small lags. Items recalled due to semantic relatedness appear as either extra-list intrusions or prior-list intrusions with lags of arbitrary magnitude. When comparing an auditory list to a visual list, we would therefore expect to see a relative reduction in the cueing of items from recent lists. Consequently, on auditory trials, semantically-cued items from more distant lists may supplant items from one list back in the ranking of most strongly activated items during recall. This, in turn, has the potential to increase the number of remote PLIs retrieved. Therefore, rather than simply producing fewer intrusions than visual trials, auditory trials might produce fewer temporally-cued intrusions but more semantically-cued intrusions.

One final success of our model was its ability to predict differences in the serial position curve as a function of recall start position. Our model predicted the generalized auditory advantage we observed in Experiment 2 when participants initiated recall from the start of the list. It further predicted the occurrence of both modality and inverse modality effects when initiating recall from the final four items of the list. There was, however, one notable discrepancy between model and data, in that the model of visual recall did not exhibit the near-elimination of the recency effect seen during beginning-start visual trials in our study. To understand why the same model parameters might produce a general auditory advantage on a trial with a beginning start and a recency-specific auditory advantage during an end-of-list start, consider the differential effects of drift rate and context-to-item associations in these two situations. When recall begins from the end of the list, the contexts of recent items are reinforced, pulling context even further from the beginning of the list. Given that early-list context is closer to end-of-list context on visual trials than on auditory trials, it would thus be easier on visual trials to transition away from late-list items after starting to recall them. Recall on auditory trials may instead tend to become anchored to the end-of-list context. When recall instead begins from the start of the list, the early-list context is immediately reinstated, which nullifies the visual advantage for transitioning away from late-list items. The stronger context-to-item associations on auditory trials may then dominate the effects of faster context drift and produce auditory memory advantages across all serial positions.

The greatest challenge the model faced was fitting recall initiation behavior. Our particle swarm optimization algorithm struggled to identify a parameter set capable of closely fitting both the serial position curve and the PFR curve. As is evident in Figure 11, the best-fitting model failed to capture the high probability of participants starting recall from the first item of the list. Later, when simulating the modality effect, the model overestimated the impact of presentation mode on the probability of initiating recall from the end of the list, though the predicted effect was in the correct direction. Under an increased drift rate, the reduced similarity between the end-of-list context and the contexts of earlier items decreases their accessibility at the start of the recall period. In our simulations, this effect produced an increased likelihood of initiating recall from the last 1-2 items of auditory trials, as compared with visual trials.

In proposing a new theoretical account of behavioral phenomena, it is important to explain not only how cognitive parameters might differ between task conditions, but why. In explaining why context drift rates and association strengths might differ by presentation mode, we consider the nature of the stimuli themselves. One frequently overlooked confound in the modality effect literature is that auditory presentation differs from visual presentation not only in the sensory mode of the stimuli, but also in their temporal dynamics. Under traditional serial and free recall paradigms, the physical features of visual items remain static for the duration of the presentation interval, while the acoustic features of auditory stimuli change over time. We suggest that the temporal dynamics inherent in auditory stimuli might induce greater shifts in temporal context with each presented item, thereby producing faster drift rates during auditory presentation than during visual presentation. Concurrently, dynamic presentation might provide an enriched set of retrieval cues relative to static presentation, thereby bolstering the degree to which context-to item associations can form. The primary appeal of this temporal dynamics hypothesis lies in its ability to explain the well-documented occurrence of modality effects during various forms of dynamic visual presentation, including silent mouthing, lip reading, sign language, and finger spelling (Campbell & Dodd, 1980; Greene & Crowder, 1984; Krakow & Hanson, 1985; Nairne & Crowder, 1982; Nairne & Walters, 1983; Shand, 1980). As our retrieved-context account does not specifically require the presence of acoustic information, it retains its generalizability to these non-auditory modes.

General Discussion

Through joint empirical and computational work, we sought to develop a retrieved-context account of modality effects in free recall. In doing so, we first collected behavioral data from approximately three thousand participants across two immediate free recall experiments with variable modality, list length, and rate of presentation. These data allowed us to precisely analyze the effects of modality on free recall, not only replicating recent and classic findings, but also identifying new effects on intrusions and their recency. Specifically, prior-list intrusions made by participants during visual trials tended to originate from more recent lists than those made during auditory trials. We then fit the interlist implementation of the Context Maintenance and Retrieval (CMR2) model by Lohnas et al. (2015) to our empirical data, accounting for modality differences in recall dynamics as a product of differential context drift rates (faster drift for auditory items) and context-to-item associative encoding (stronger for auditory items).

Across both within-subject (Exp. 1) and between-subject (Exp. 2) manipulations of modality, we found auditory recency advantages as well as visual primacy advantages, consistent with a wide body of prior work (e.g., Craik, 1969; Grenfell-Essam et al., 2017;

Murdock & Walker, 1969). For participants who completed trials of both modalities (Exp. 1), we found that faster presentation strengthened modality effects and attenuated inverse modality effects, a pattern that replicates previous findings (e.g., Macken et al., 2016; Murdock & Walker, 1969; Murray, 1966; Roberts, 1972). List length, on the other hand, did not influence the modality effect, consistent with the observations of Roberts (1972). However, as we only compared lists of 12 and 24 items, there remains the possibility of list length $\times$ modality interactions at list lengths below 12, as data from Grenfell-Essam et al. (2017) suggests. Although we found that participants were more likely to initiate recall from end-of-list items on auditory trials as compared with visual trials, auditory recency advantages persisted even when recall initiated from the start of the list (see Figures 3 and 8). Furthermore, the visual primacy advantage associated with the inverse modality effect was large when initiating recall from the end of the list, but was attenuated or nonexistent when initiating recall from the beginning of the list. We also observed an auditory final free recall advantage in Experiment 2, such that previously unrecalled auditory words were more likely than previously unrecalled visual words to be retrieved during final free recall. Although trending in the same direction, this effect did not reach significance in Experiment 1. Finally, in both experiments, we observed a novel effect of modality on the recency of prior-list intrusions (Figures 5 and 10), such that visual presentation elicited more intrusions from one list back than did auditory presentation, but fewer intrusions from further lists back. Experiment 1 revealed that this effect was driven by the modality of the current trial rather than the modality of the intruding word itself. Modality did not affect the total number of prior-list intrusions; however, words were more likely to intrude on trials of the same modality than on trials of a different modality, a pattern previously identified by Murdock (1966).

To help elucidate the mechanistic underpinnings of modality effects in free recall, we developed an extension of the CMR2 model (Lohnas et al., 2015) and fit it to the data on recall dynamics obtained in Experiment 2. Based on previous theories by Glenberg and

Swanson (1986) and Nairne (1990), we hypothesized that a faster-drifting temporal context (increased β_{enc}) and strengthened formation of context-to-item associations (increased γ_{CF}) during auditory presentation would account for modality effects in free recall. Through a series of simulations, we showed that a retrieved-context model of memory can successfully account for the classic modality and inverse modality effects as a function of these two parameters. According to our model, the combination of faster context drift and stronger context-to-item associations results in a joint effect in which increased drift rates bias recall towards recency items, while improved retrieval cueing boosts recall across all serial positions. We believe that the strengths of this retrieved-context account of the modality effect are threefold. First, our model can explain both modality and inverse modality effects within a single framework. Existing theories of the modality effect, including PAS (Crowder & Morton, 1969), temporal distinctiveness theory (Glenberg & Swanson, 1986), the feature model (Nairne, 1990), and output interference accounts (e.g., Cowan et al., 2002; Harvey & Beaman, 2007) all fail to address inverse modality effects, instead relying on alternative theories to explain such findings (e.g., a visual advantage for rehearsal). Second, our model correctly predicts several of the auxiliary effects of modality observed both in our study and in prior literature, including the persistence of the modality effect across different recall start positions (e.g., Grenfell-Essam et al., 2017), the attenuation of the inverse modality effect when recall begins from the start of the list (e.g., Craik, 1969), the greater recency of prior-list intrusions on visual trials, and the non-effect of modality on the total number of prior-list intrusions committed. Third, our account requires no assumption that the modality effect depends specifically on acoustic information content. Regardless of modality, our claim is that the temporal dynamics of different types of stimuli variably affect their integration into context. This feature is important for explaining why a variety of visual presentation modes (e.g., sign language, finger spelling, and lip reading) can produce nearly identical modality and inverse modality effects to auditory presentation (Campbell & Dodd, 1980; Krakow & Hanson, 1985; Shand, 1980).

We therefore find a retrieved-context account of the modality effect appealing, both for its generality and for its explanatory power.

Although we have taken first steps towards establishing a retrieved-context account of modality effects in free recall, it cannot be ignored that the modality and inverse modality effects vary in extent and magnitude between different types of recall tasks (Grenfell-Essam et al., 2017); indeed, many studies observe no inverse modality effects at all. It remains to be seen whether our model can account for this variability, though it is conceivable that these differences might arise as a result of drift rates and association strengths balancing differently under different task conditions. For instance, the absence of inverse modality effects at very short list lengths (Grenfell-Essam et al., 2017) may be attributable to context drift rates having less of an impact on primacy performance in short lists. Further modeling is also required to assess the ability of our retrieved-context account to explain the various interactions between the modality effect and list suffixes, distractor tasks, and articulatory suppression, all topics that must be addressed by any comprehensive theory of the modality effect (e.g., Campbell & Dodd, 1980; Crowder, 1971; Gardiner & Gregg, 1979; Gathercole et al., 1983; Greene & Crowder, 1984; Huang & Glenberg, 1986; Macken et al., 2016; Morton & Holloway, 1970; Nairne & Crowder, 1982; Routh, 1976; Spoehr & Corin, 1978). Future modeling work should also consider the addition of source features to our account (Polyn et al., 2009). Our finding that people make more same-modality intrusions than cross-modality intrusions strongly suggests the presence of source coding, as does Murdock and Walker's (1969) discovery that people cluster items by modality during recall.

In explaining why drift rate and context-to-item associations may vary by modality, we suggested that the features of the stimulus changing over the presentation interval may drive larger shifts in context with each learned item, while also supporting richer contextual coding. This hypothesis suggests that adding temporal dynamics to visual item presentation should produce behavior consistent with faster drift rates and stronger context-to-item associations, perhaps mimicking the modality and inverse modality effects even in the absence of acoustic information. Evidence in support of this hypothesis comes from studies of recall for sign language, finger spelling, and lip reading – all of which produce modality and inverse modality effects despite information being presented in a strictly visual fashion (Campbell & Dodd, 1980; Krakow & Hanson, 1985; Shand, 1980). Studying whether modality effects arise with other forms of dynamic visual presentation might offer further validation, and such a possibility should be explored.

Students of memory have long sought to explain modality effects in episodic memory. From differing sensory stores to discrepancies in temporal and featural coding, many theories have tried to explain why recall of recent information might vary by mode of presentation. Researchers have often treated the modality effect as evidence of separate short-term and long-term memory processes. Instead, our modeling work has demonstrated that modality effects can be well-accounted for, even in a single-process system.

References

Beaman, C. P. (2002). Inverting the modality effect in serial recall. *Quarterly Journal of Experimental Psychology*, *55A*(2), 371-389. doi: 10.1080/02724980143000307

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, *6*(2), 173–189. doi: 10.1016/0010-0285(74)90009-7

Bratton, D., & Kennedy, J. (2007). Defining a standard for particle swarm optimization. In *Swarm Intelligence Symposium, 2007* (p. 120-127). Honolulu, HI: IEEE. doi: 10.1109/SIS.2007.368035

Campbell, R., & Dodd, B. (1980). Hearing by eye. *Quarterly Journal of Experimental Psychology*, *32*(1), 85-99. doi: 10.1080/00335558008248235

Clerc, M., & Kennedy, J. (2002). The particle swarm – explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, *6*(1), 58-73. doi: 10.1109/4235.985692

Conrad, R., & Hull, A. J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science*, *10*(4), 135-136. doi: 10.3758/BF03331446

Corballis, M. C. (1966). Rehearsal and decay in immediate recall of visually and aurally presented items. *Canadian Journal of Psychology*, *20*(1), 43-51. doi: 10.1037/h0082923

Cowan, N., Saults, J., & Brown, G. (2004). On the auditory modality superiority effect in serial recall: Separating input and output factors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 639-644. doi: 10.1037/0278-7393.30.3.639

Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding serial recall. *Journal of Memory and Language*, *46*(1), 153–177. doi: 10.1006/jmla.2001.2805

Craik, F. I. M. (1969). Modality effects in short-term storage. *Journal of Verbal Learning* *and Verbal Behavior*, *8*(5), 658-664. doi: 10.1016/S0022-5371(69)80119-2

Craik, F. I. M. (1970). The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, *9*(2), 658-664. doi: 10.1016/S0022-5371(70)80042-1

Crowder, R. G. (1971). The sound of vowels and consonants in immediate memory. *Journal of Verbal Learning and Verbal Behavior*, *10*(6), 587-596. doi: 10.1016/S0022-5371(71)80063-4

Crowder, R. G. (1978). Mechanisms of auditory backward masking in the stimulus suffix effect. *Psychological Review*, *85*(6), 502-524. doi: 10.1037/0033-295X.85.6.502

Crowder, R. G. (1983). The purity of auditory memory. *Philosophical Transactions of the Royal Society of London, B*, *302*(1110), 251-265. doi: 10.1098/rstb.1983.0053

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics*, *5*(6), 365-373. doi: 10.3758/BF03210660

Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, *7*(3), 171-176. doi: 10.1145/363958.363994

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1-12. doi: 10.3758/s13428-014-0458-y

Eberhart, R. C., & Shi, Y. (2000). Comparing inertia weights and constriction factors in particle swarm optimization. In *Proceedings of the 2000 IEEE Congress on Evolutionary Computation* (Vol. 1, p. 84-88). San Diego, CA. doi: 10.1109/CEC.2000.870279

Fai, H. T., & Cornelius, P. L. (1996). Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *Journal of Statistical Computation and Simulation*, *54*(4), 363-378.

Gardiner, J. M. (1983). On recency and echoic memory. *Philosophical Transactions of the Royal Society of London, B*, *302*(1110), 267-282. doi: 10.1098/rstb.1983.0054

Gardiner, J. M., & Gregg, V. H. (1979). When auditory memory is not overwritten.

Journal of Verbal Learning and Verbal Behavior, *18*(6), 705-719. doi: 10.1016/S0022-5371(79)90411-0

Gathercole, S. E., Gregg, V. H., & Gardiner, J. M. (1983). Influences of delayed distraction on the modality effect in free recall. *British Journal of Psychology*, *74*(2), 223-232. doi: 10.1111/j.2044-8295.1983.tb01858.x

Glenberg, A. M. (1984). A retrieval account of the long-term modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 16-31. doi: 10.1037/0278-7393.10.1.16

Glenberg, A. M., & Fernandez, A. (1988). Evidence for auditory temporal distinctiveness: Modality effects in order and frequency judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 728-739. doi: 10.1037/0278-7393.14.4.728

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*(1), 3-15. doi: 10.1037/0278-7393.12.1.3

Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 371-377. doi: 10.1037/0278-7393.15.3.371

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*(3), 371-382. doi: 10.1016/S0022-5371(84)90259-7

Greene, R. L., & Crowder, R. G. (1986). Recency effects in delayed recall of mouthed stimuli. *Memory & Cognition*, *14*(4), 355-360. doi: 10.3758/BF03202514

Greene, R. L., & Crowder, R. G. (1988). Memory for serial position: Effects of spacing, vocalization, and stimulus suffixes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 740-748. doi: 10.1037/0278-7393.14.4.740

Grenfell-Essam, R., Ward, G., & Tan, L. (2017). Common modality effects in immediate free recall and immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(12), 1909-1933. doi: 10.1037/xlm0000430

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., . . . Chan, P. (2016). psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavior Research Methods*, *48*(3), 829–842. doi: 10.3758/s13428-015-0642-8

Harvey, A. J., & Beaman, C. P. (2007). Input and output modality effects in immediate serial recall. *Memory*, *15*(7), 693-700. doi: 10.1080/09658210701644677

Healey, M. K. (2018). Temporal contiguity in incidentally encoded memories. *Journal of Memory and Language*, *102*, 28-40. doi: 10.1016/j.jml.2018.04.003

Healey, M. K., Crutchley, P., & Kahana, M. J. (2014). Individual differences in memory search and their relation to intelligence. *Journal of Experimental Psychology: General*, *143*(4), 1553–1569. doi: 10.1037/a0036306

Healey, M. K., & Kahana, M. J. (2016). A four–component model of age–related memory change. *Psychological Review*, *123*(1), 23-69. doi: 10.1037/rev0000015

Healey, M. K., Long, N. M., & Kahana, M. J. (in press). Contiguity in episodic memory. *Psychonomic Bulletin & Review*. doi: 10.3758/s13423-018-1537-3

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299. doi: 10.1006/jmps.2001.1388

Huang, S., & Glenberg, A. M. (1986). Echoic and retrieval accounts of the long-term modality effect tested using the suffix procedure. *American Journal of Psychology*, *99*(4), 453-470. doi: 10.2307/1422285

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*(1), 103–109. doi: 10.3758/BF03197276

Kahana, M. J., Aggarwal, E. V., & Phan, T. D. (2018). The variability puzzle in human memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(12), 1857–1863. doi: 10.1037/xlm0000553

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of the 1995 IEEE International Conference on Neural Networks* (Vol. 4, p. 1942-1948). Perth, Australia. doi: 10.1109/ICNN.1995.488968

Krakow, R. A., & Hanson, V. L. (1985). Deaf signers and serial recall in the visual modality: Memory for signs, fingerspelling, and print. *Memory & Cognition*, *13*(3), 265-272. doi: 10.3758/BF03197689

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1-26.

Laughery, K. R., & Pinkus, A. L. (1966). Short-term memory: Effects of acoustic similarity, presentation rate and presentation mode. *Psychonomic Science*, *6*(6), 285-286. doi: 10.3758/BF03328069

Lohnas, L. J., Polyn, S. M., & Kahana, M. (2015). Expanding the scope of memory search: Intralist and interlist effects in free recall. *Psychological Review*, *122*(2), 337-363. doi: 10.1037/a0039036

Long, N. M., Danoff, M. S., & Kahana, M. J. (2015). Recall dynamics reveal the retrieval of emotional context. *Psychonomic Bulletin and Review*, *22*(5), 1328-1333. doi: 10.3758/s13423-014-0791-2

Macken, B., Taylor, J. C., Kozlov, M. D., Hughes, R. W., & Jones, D. M. (2016). Memory as embodiment: The case of modality and serial short-term memory. *Cognition*, *155*, 113-124. doi: 10.1016/j.cognition.2016.06.013

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23. doi: 10.3758/s13428-011-0124-6

McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review*, *15*(2), 237–255. doi: 10.3758/PBR.15.2.237

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781v3*.

Morton, J., & Holloway, C. M. (1970). Absence of a cross-modal "suffix effect" in short-term memory. *Quarterly Journal of Experimental Psychology*, *22*, 167-176.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482-488. doi: 10.1037/h0045106

Murdock, B. B. (1966). Visual and auditory stores in short-term memory. *Quarterly Journal of Experimental Psychology*, *18*(3), 206-211.

Murdock, B. B. (1967). Recent developments in short-term memory. *British Journal of Psychology*, *58*(3-4), 421-433. doi: 10.1111/j.2044-8295.1967.tb01099.x

Murdock, B. B. (1968). Modality effects in short-term memory: Storage or retrieval? *Journal of Experimental Psychology*, *77*(1), 79-86. doi: 10.1037/h0025786

Murdock, B. B., & Metcalfe, J. (1978). Controlled rehearsal in single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, *17*(3), 309-324. doi: 10.1016/S0022-5371(78)90201-3

Murdock, B. B., & Walker, K. D. (1969). Modality effects in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*(5), 665-676. doi: 10.1016/S0022-5371(69)80120-9

Murray, D. J. (1966). Vocalization-at-presentation and immediate recall, with varying recall methods. *Quarterly Journal of Experimental Psychology*, *18*(1), 9-18.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251-269. doi: 10.3758/BF03213879

Nairne, J. S., & Crowder, R. G. (1982). On the locus of the stimulus suffix effect. *Memory & Cognition*, *10*(4), 350-357. doi: 10.3758/BF03202427

Nairne, J. S., & Walters, V. L. (1983). Silent mouthing produces modality- and suffix-like effects. *Journal of Verbal Learning and Verbal Behavior*, *22*(4), 475-483. doi: 10.1016/S0022-5371(83)90300-6

Nilsson, L. G. (1975). Locus of the modality effect in free recall: A reply to Watkins. *Journal of Experiment Psychology: Human Learning and Memory*, *104*(1), 13-17. doi: 10.1037/0278-7393.1.1.13

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, *17*(4), 398-422. doi: 10.3758/BF03202613

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156. doi: 10.1037/a0014420

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 14, p. 207-262). New York: Academic Press. doi: 10.1016/S0079-7421(08)60162-0

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93-134. doi: 10.1037/0033-295X.88.2.93

Roberts, W. A. (1972). Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses. *Journal of Experimental Psychology*, *92*(3), 365-372. doi: 10.1037/h0032278

Routh, D. A. (1976). An "across-the-board" modality effect in immediate serial recall. *Quarterly Journal of Experimental Psychology*, *28*(2), 285-304.

Sederberg, P. B., Miller, J. F., Howard, W. H., & Kahana, M. J. (2010). The temporal contiguity effect predicts episodic memory performance. *Memory & Cognition*, *38*(6), 689–699. doi: 10.3758/MC.38.6.689

Serra, M., & Nairne, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition*, *21*(1), 34–40. doi: 10.3758/BF03211162

Shand, M. A. (1980). *Short-term coding processes in congenitally deaf signers of ASL: Natural language considerations* (Unpublished doctoral dissertation). University of California, San Diego.

Shand, M. A., & Klima, E. S. (1981). Nonauditory suffix effects in congenitally deaf signers of American Sign Language. *Journal of Experimental Psychology: Human*

Learning and Memory, *7*(6), 464-474. doi: 10.1037/0278-7393.7.6.464

Spoehr, K. T., & Corin, W. J. (1978). The stimulus suffix effect as a memory coding phenomenon. *Memory & Cognition*, *6*(6), 583-589. doi: 10.3758/BF03198247

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592.

Watkins, M. J. (1972). Locus of the modality effect in free recall. *Journal of Verbal Learning and Verbal Behavior*, *11*(5), 644-648. doi: 10.1016/S0022-5371(72)80048-3

Watkins, M. J., & Watkins, O. C. (1973). The postcategorical status of the modality effect in serial recall. *Journal of Experimental Psychology*, *99*(2), 226-230. doi: 10.1037/h0034651

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*(4), 430-447. doi: 10.1016/S0022-5371(74)80021-6

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *104*(4), 442-452. doi: 10.1037/0278-7393.1.4.442

Watkins, O. C., & Watkins, M. J. (1977). Serial recall and the modality effect: Effects of word frequency. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(6), 712-718. doi: 10.1037/0278-7393.3.6.712

Zaromb, F. M., Howard, M. W., Dolan, E. D., Sirotin, Y. B., Tully, M., Wingfield, A., & Kahana, M. J. (2006). Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 792–804. doi: 10.1037/0278-7393.32.4.792

Zaromb, F. M., & Roediger, H. L. (2009). The effects of "effort after meaning" on recall: Differences in within- and between-subjects designs. *Memory & Cognition*, *37*(4), 447–463. doi: 10.3758/MC.37.4.447

Best-Fitting Parameters

Param	Avg	Vis	Aud
β_{enc}	0.518	0.489	0.535
β_{rec}	0.366	-	-
γ_{FC}	0.264	-	-
γ_{CF}	0.641	0.543	0.707
ϕ_s	3.31	-	-
ϕ_d	0.529	-	-
κ	0.256	-	-
λ	0.366	-	-
η	0.581	-	-
s	0.591	-	-
β_{post}^{recall}	0.677	-	-
ω	6.468	-	-
α	0.770	-	-
c_{thresh}	0.063	-	-

Goodness-of-Fit (RMSD)

	Avg	Vis	Aud
Full	**0.041**	0.047	0.037
SPC	0.023	**0.021**	**0.024**
PFR	0.065	0.075	0.058
PLI	0.015	0.030	0.011
PLIR	0.008	0.023	0.007

Table 1

Left panel indicates the parameter sets for CMR2 that provided the best fit to the average and modality-specific Experiment 2 data. Right panel shows the weighted RMSD values for each individual behavioral metric, as well as the full combination of all metrics. Bold-faced fitness values indicate which metric was used for selecting the best model in each simulation.

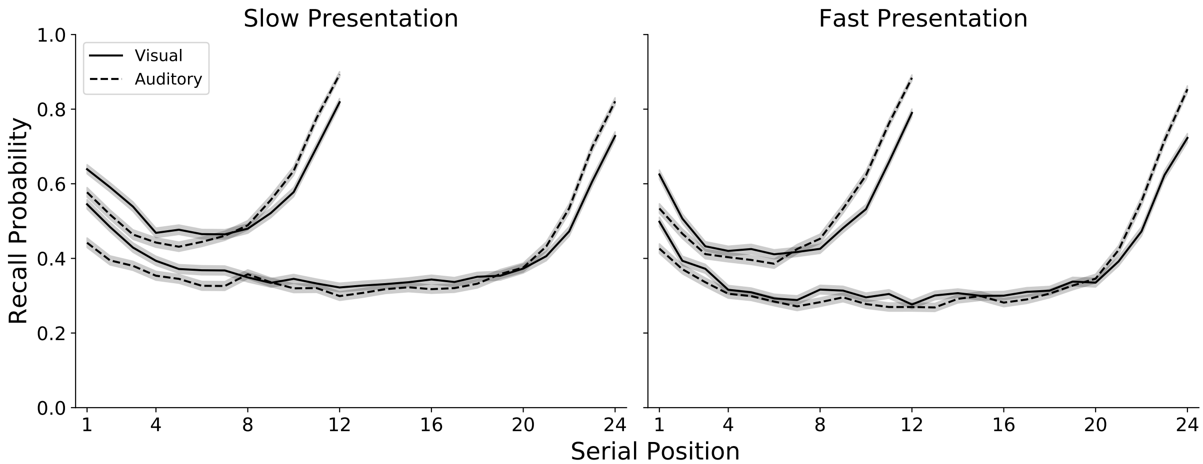

Figure 1. Serial position curves plotted for each combination of modality, list length, and presentation rate in Experiment 1. Shaded regions indicate one standard error of the mean.

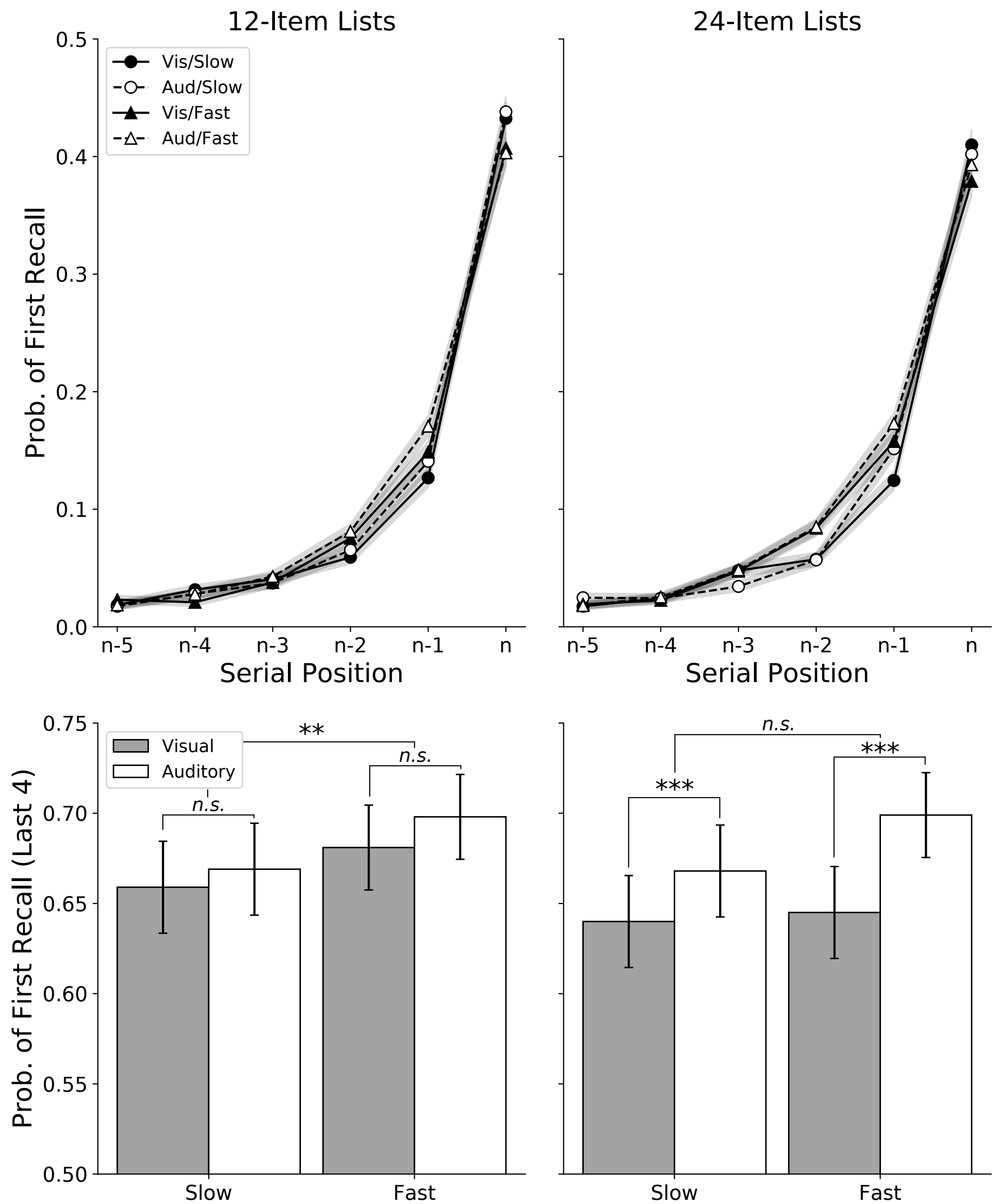

Figure 2. Top row shows the probability of first recall across the final six serial positions of each type of list in Experiment 1, with shaded regions indicating one standard error of the mean. Bottom row shows the total probability of initiating recall from any of the final four list items, with error bars indicating 95% confidence intervals.

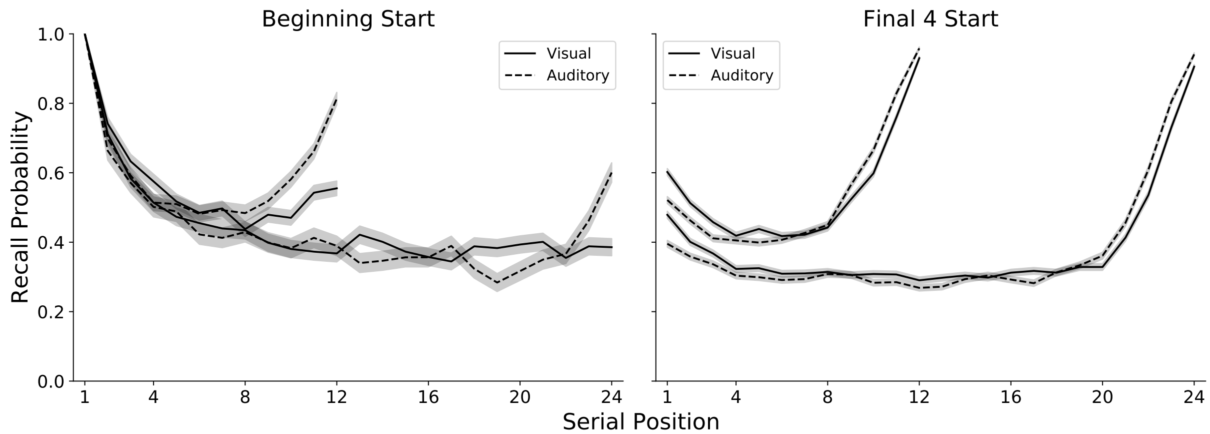

Figure 3. Average serial position curves across all 12- and 24-item trials in Experiment 1 where recall started from the first list item (left panel) versus any one of the last four list items (right panel). Shaded regions indicate one standard error of the mean across trials.

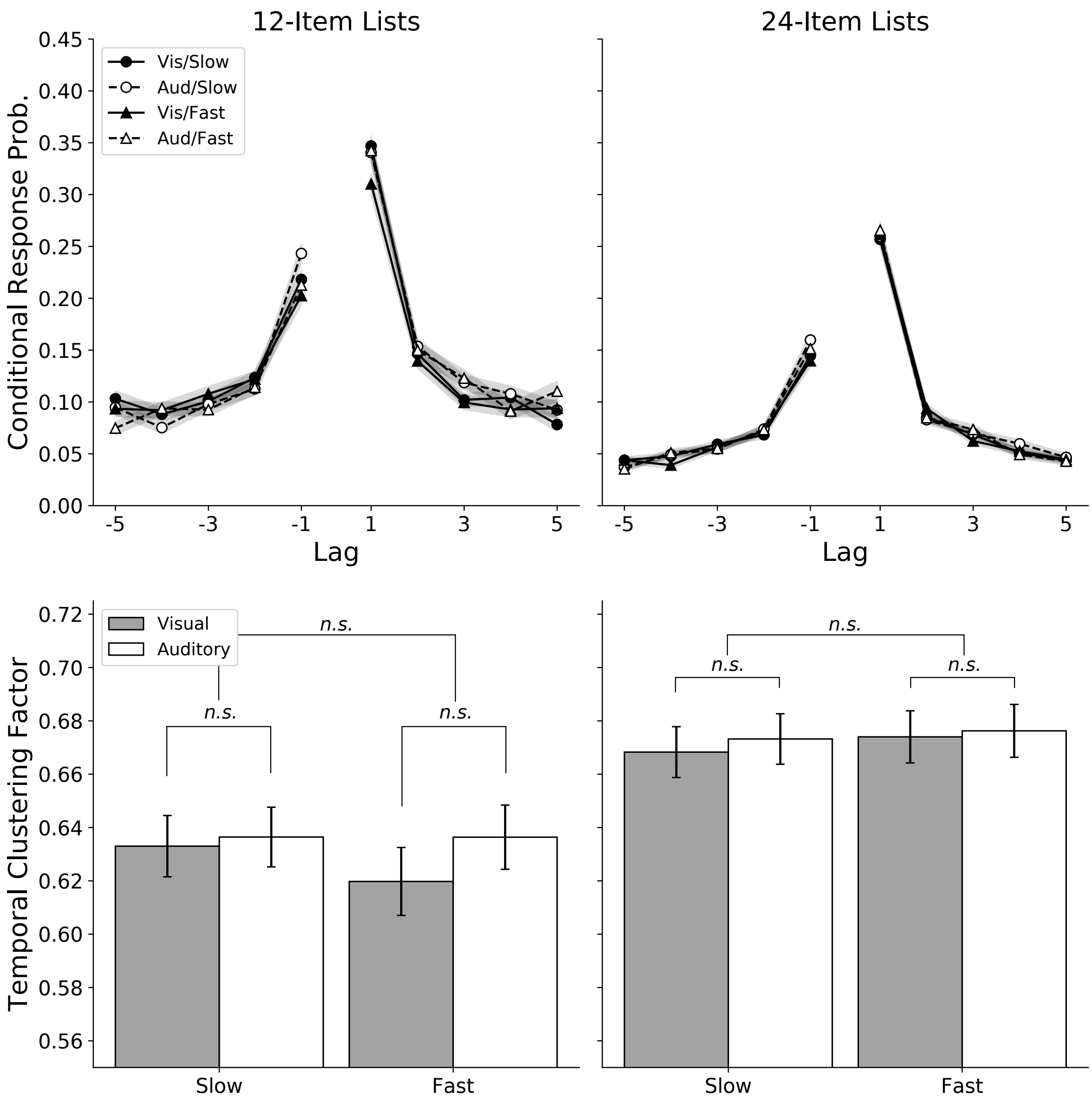

Figure 4. Lag-conditional response probability (lag-CRP) and temporal clustering factor scores for Experiment 1. The first two transitions from each trial have been excluded from these data. Shaded regions indicate one standard error of the mean, while error bars indicate 95% confidence intervals.

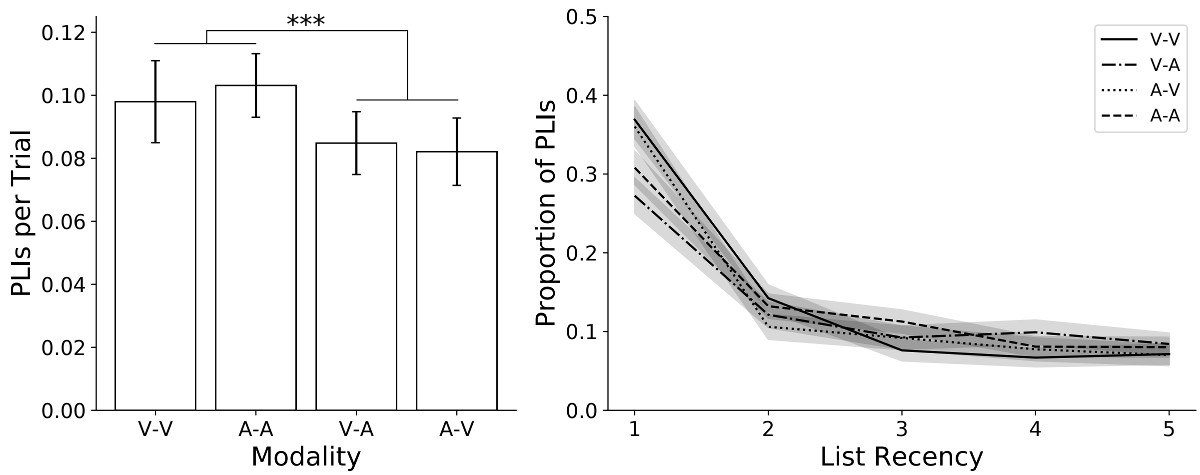

Figure 5. Average prior-list intrusions per trial and PLI recency in Experiment 1, plotted by modality. The first letter in each modality label indicates the original modality of the intruding word, while the second letter indicates the modality of the trial where the intrusion occurred. All scores are conditional on the fact that PLIs of that type were capable of being made on a given trial. Shaded regions indicate one standard error of the mean, while error bars indicate 95% confidence intervals.

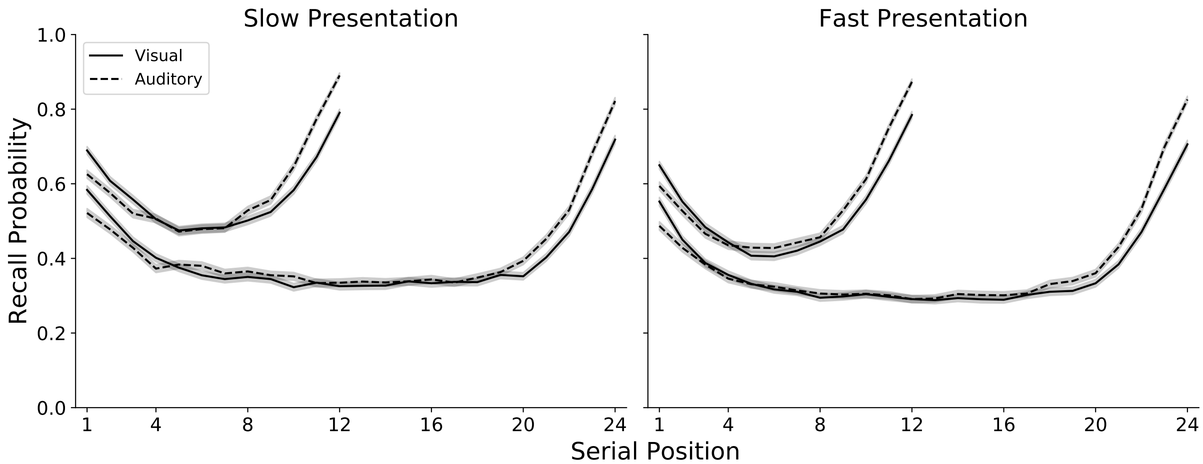

Figure 6. Serial position curves plotted for each combination of modality, list length, and presentation rate in Experiment 2. Shaded regions indicate one standard error of the mean.

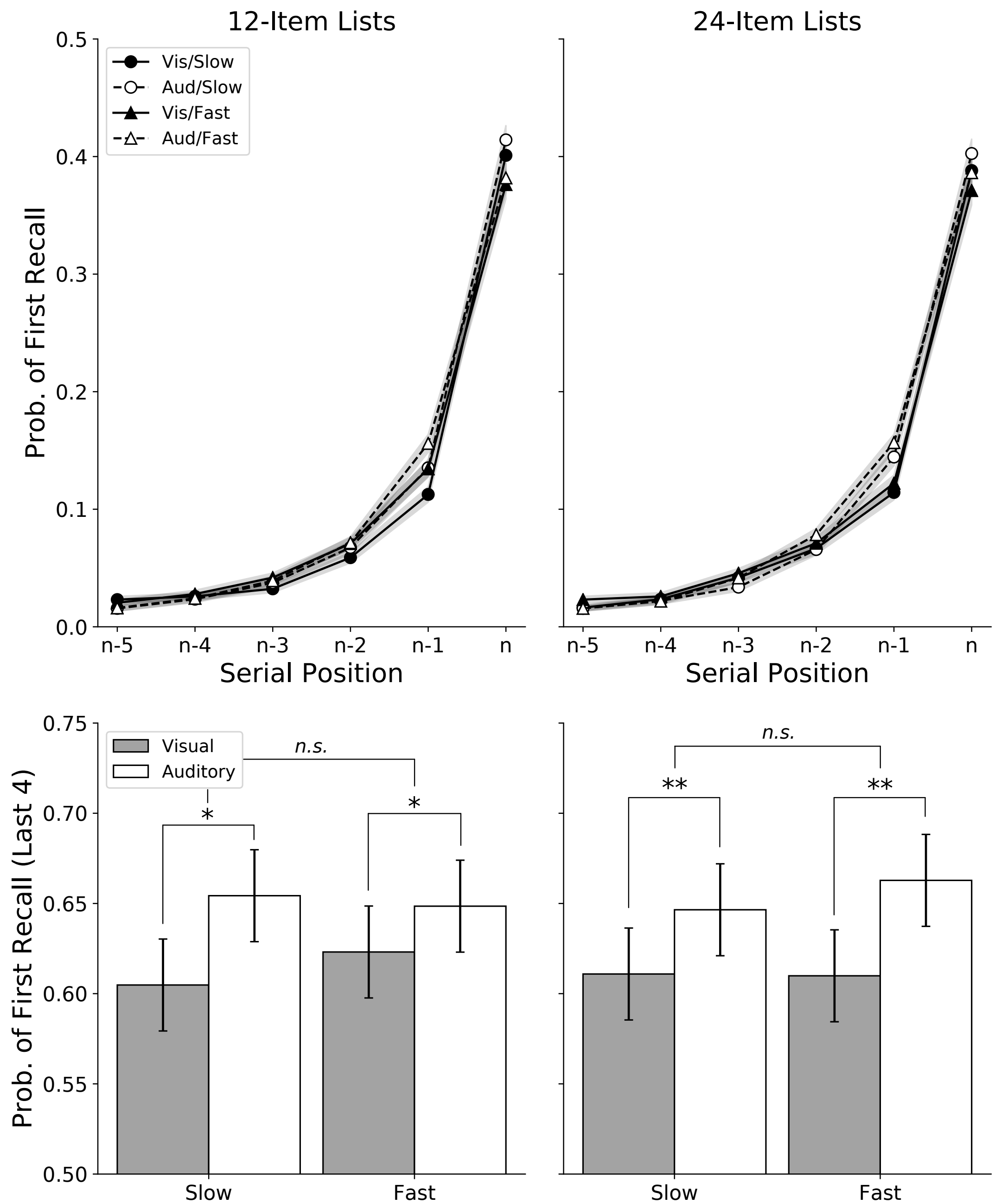

Figure 7. Top row shows the probability of first recall across the final six serial positions of each type of list in Experiment 2, with shaded regions indicating one standard error of the mean. Bottom row shows the total probability of initiating recall from any of the final four list items, with error bars indicating 95% confidence intervals.

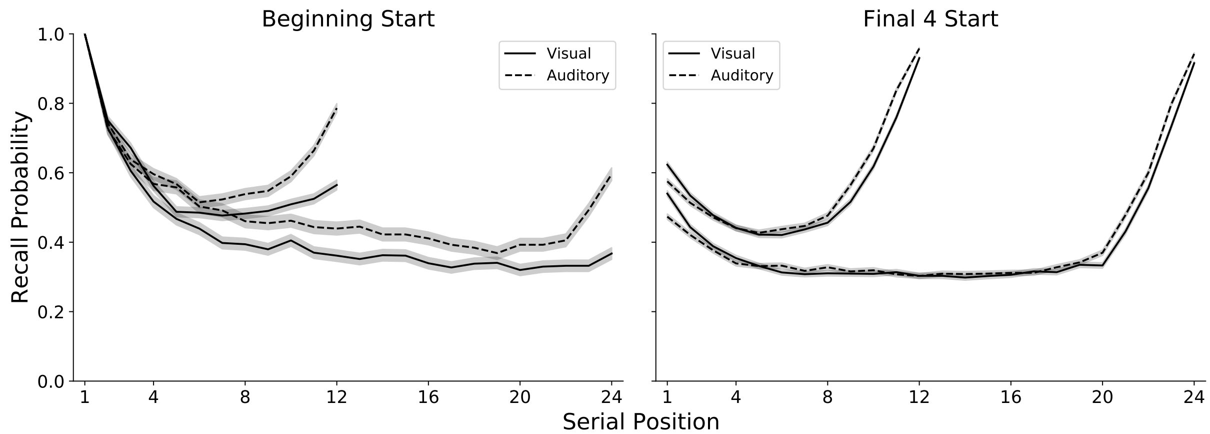

Figure 8. Average serial position curves across all 12- and 24-item trials in Experiment 2 where recall started from the first list item (left panel) versus any one of the last four list items (right panel). Shaded regions indicate one standard error of the mean across trials.

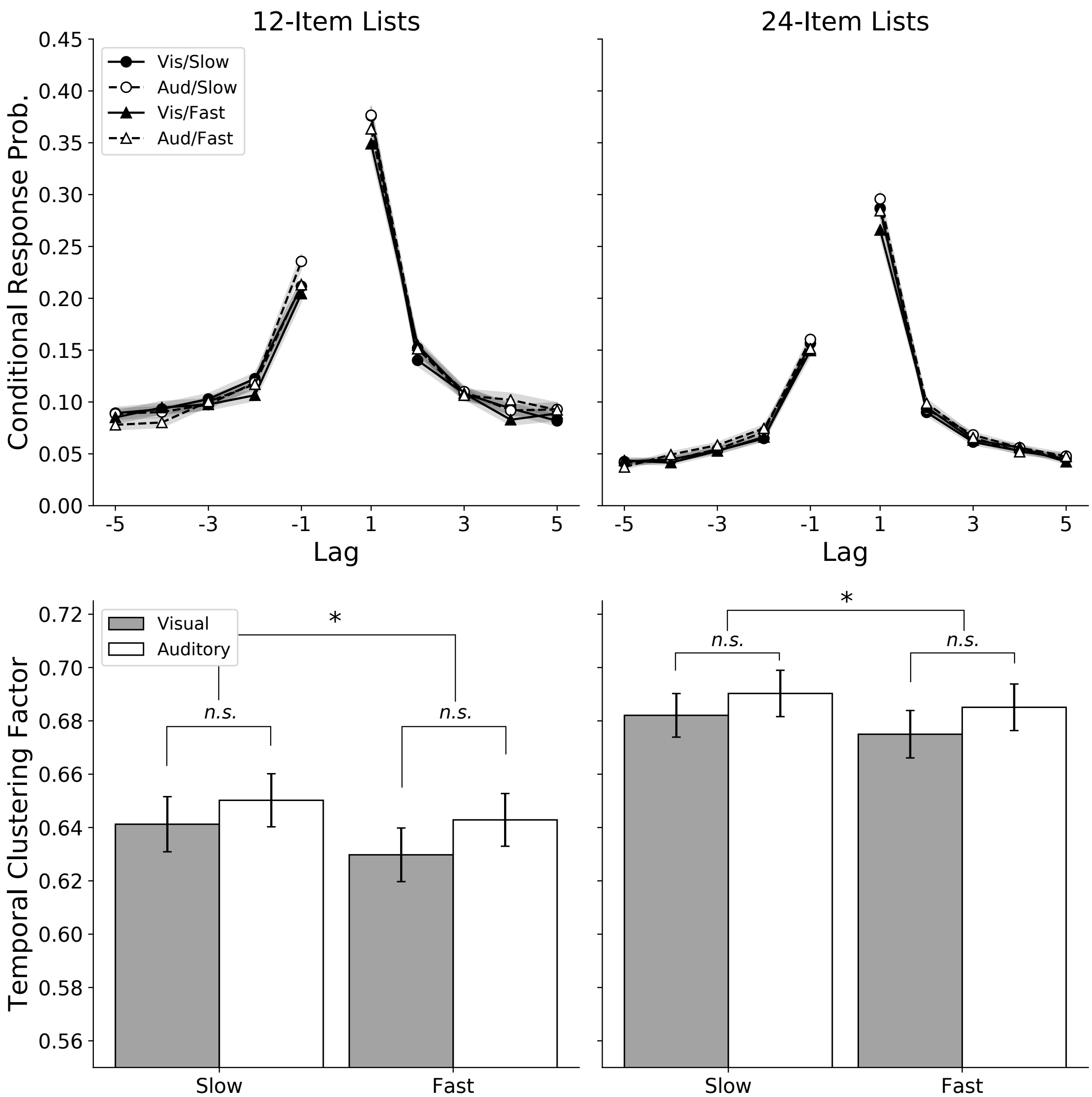

Figure 9. Lag-conditional response probability and temporal clustering factor scores for Experiment 2. The first two transitions from each trial have been excluded from these data. Shaded regions indicate one standard error of the mean and error bars indicate 95% confidence intervals.

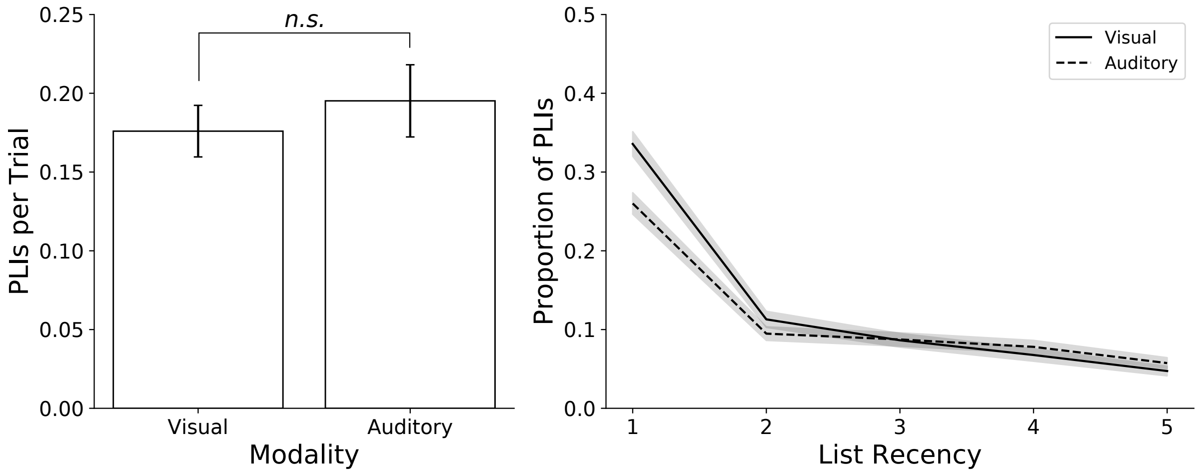

Figure 10. Average prior-list intrusions per trial and PLI recency in Experiment 2, plotted by modality. PLI recency scores are conditional on the fact that PLIs of that lag were capable of being made on a given trial. Shaded regions indicate one standard error of the mean, while error bars indicate 95% confidence intervals.

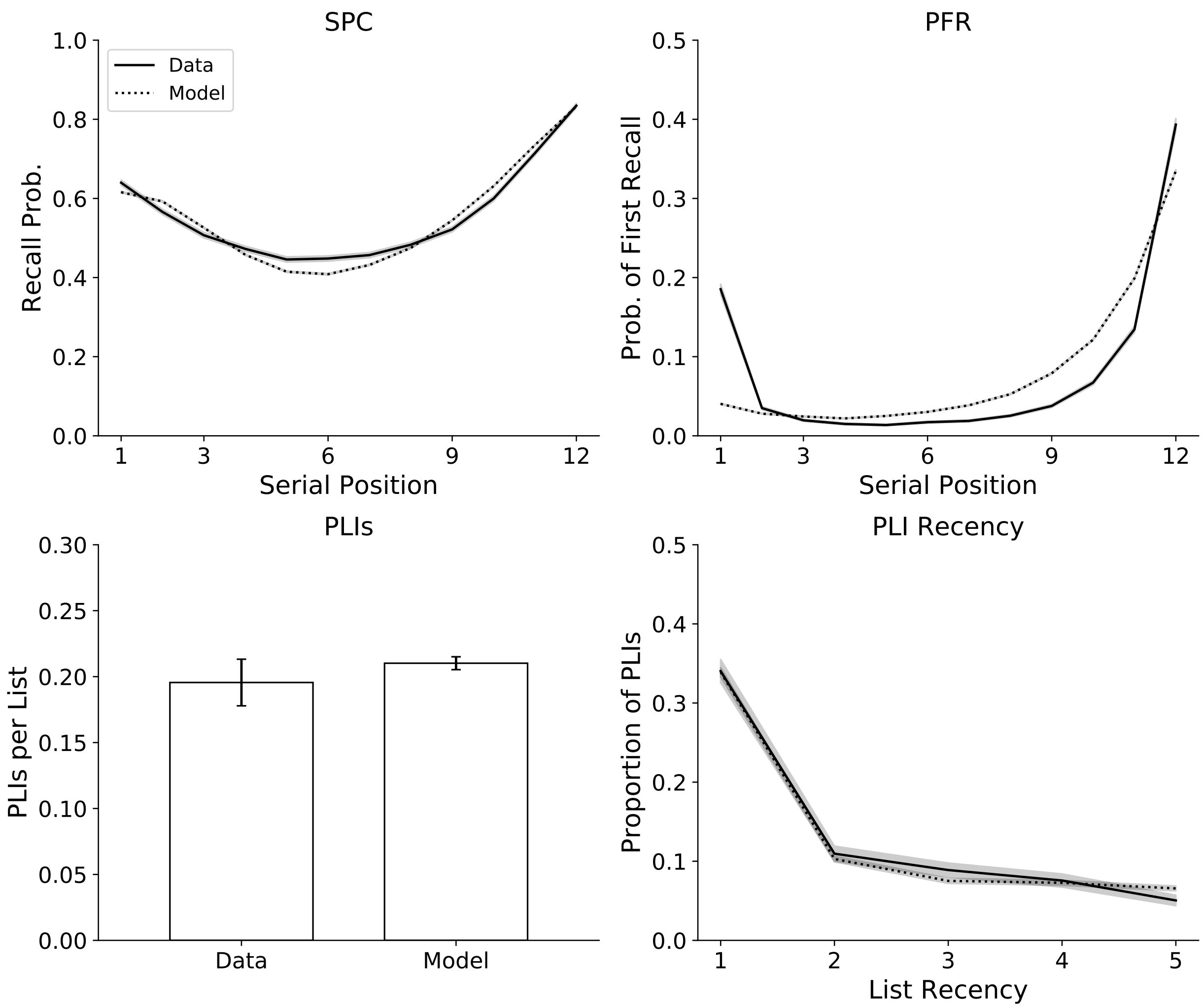

Figure 11. Behavioral predictions from the model that best fit the average data from Experiment 2, pooled across modalities. For comparison, simulated recall performance is plotted alongside the empirical data. Shaded regions indicate one standard error of the mean across sessions, while error bars indicate 95% confidence intervals.

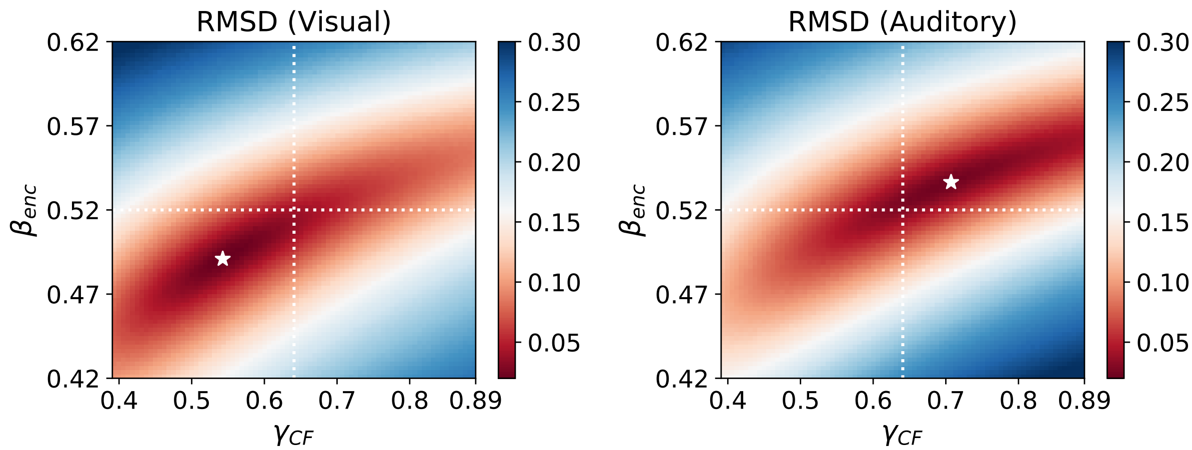

Figure 12. Heat maps illustrate weighted root-mean-square deviation between simulated and empirical serial position curves for both modalities, as a function of β_{enc} and γ_{CF}. Dotted lines indicate the combination of parameters that produced the best fit to the overall average data. The best-fitting parameter combination for each individual modality is indicated by a star.

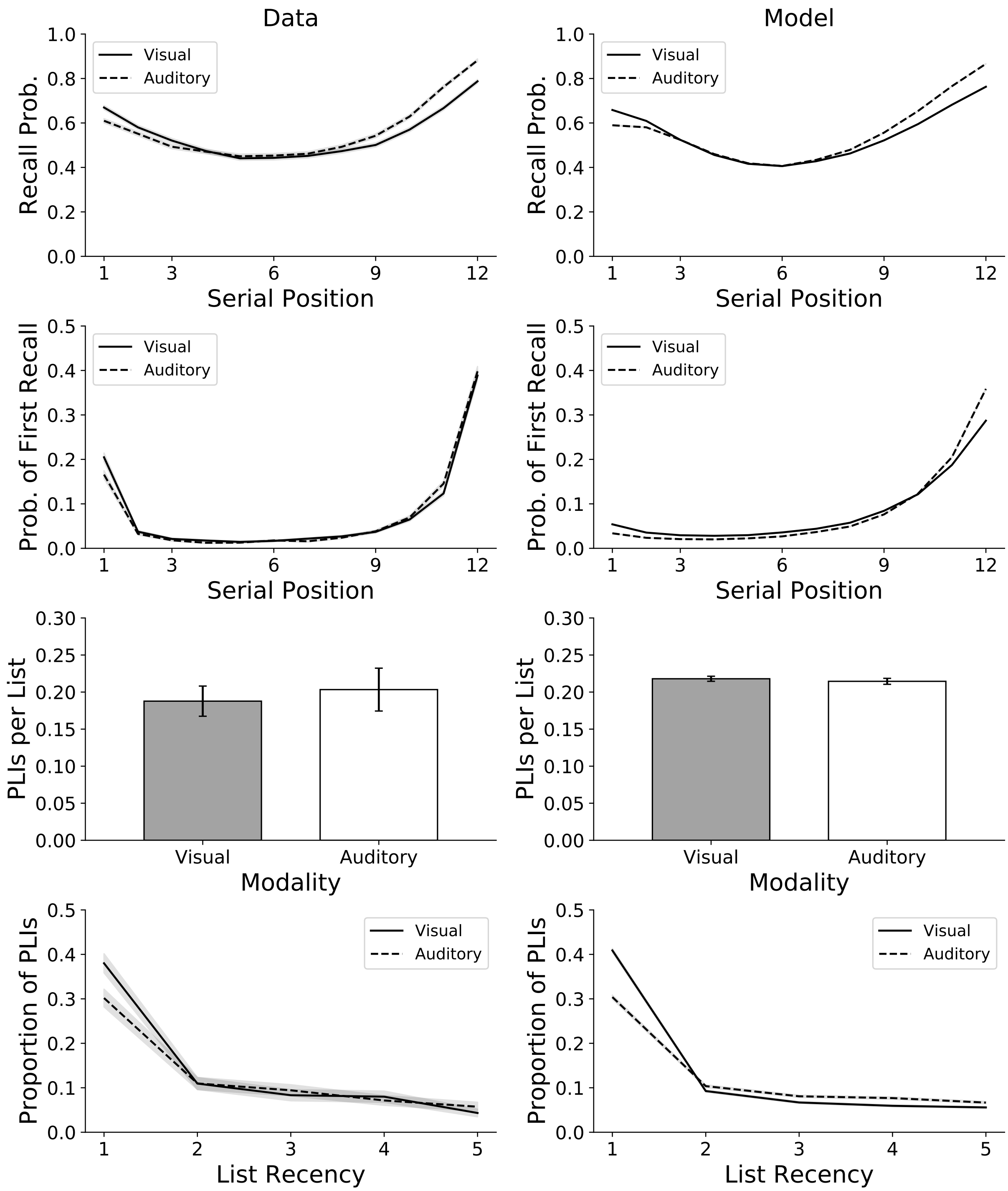

Figure 13. Recall performance by modality as observed in Experiment 2 (left column), and as simulated by CMR2 (right column). Shaded regions indicate one standard error of the mean across sessions, while error bars indicate 95% confidence intervals.

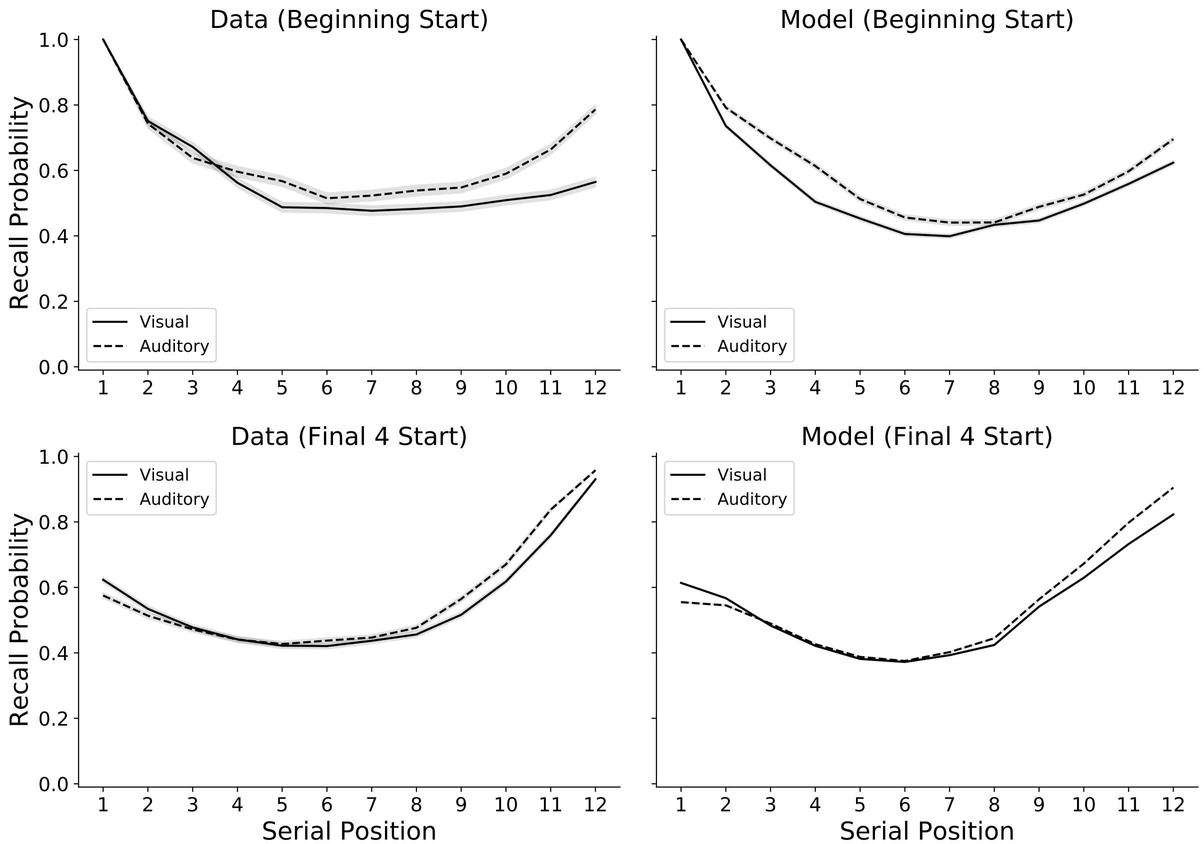

Figure 14. Average SPCs across trials where recall began from either the first list item (top row) or any one of the final four list items (bottom row), as observed in Experiment 2 (left column) and as predicted by CMR2 (right column). Shaded regions indicate one standard error of the mean across trials.

Appendix A

List Construction

Word lists were constructed such that varying degrees of semantic relatedness occurred at both adjacent and distant serial positions. Semantic relatedness between all words in the pool was determined using Google's Word2vec model (Mikolov et al., 2013), by calculating the cosine similarity between the vector representations of all pairs of words. We then used these similarity scores to group word pairs into three linearly spaced similarity bins (low, medium, and high similarity). 12-item lists were constructed from two word pairs from each similarity bin. One word pair from each bin was organized with its words at adjacent serial positions. The other word pair was positioned such that its constituents were separated by at least two other words. 24-item lists were constructed by generating two 12-item lists to the above specifications, then concatenating them. Semantic structure was not controlled on practice lists, which were generated by randomly choosing and ordering words not already selected for use in the experimental trials of that session.

Appendix B

Recall Scoring

Due to the fact that participants typed their responses, misspellings and other typos in the recall data are to be expected. Requiring a perfect match between recalled and presented words would therefore underestimate recall performance, while overestimating the number of extra-list intrusions that participants made. Furthermore, we might expect misspellings to disproportionately affect auditory trials, where participants cannot observe the correct spelling of each word prior to recall. To account for these problems, our scoring process incorporated an automated spellchecking algorithm. Specifically, we scored each recall response as follows:

First, we removed all white space from the response and converted all letters to lower case. Next, we marked the response as a correct recall if it perfectly matched any word from the current trial. If it perfectly matched a word from a previous trial, the response was marked as a prior-list intrusion. If no matches were found, the algorithm searched for any perfectly matching word in Webster's Second International Dictionary (https://libraries.io/npm/web2a). If the response matched any word in the dictionary, it was marked as an extra-list intrusion. If no matching word existed, the response was assumed to be misspelled. Misspelled responses were compared with all words in the dictionary by computing the Damerau-Levenshtein distance (Damerau, 1964) to each, thereby giving a distribution of similarities between the response and all English words. Next, the response was compared to all previously presented words, in order to identify the closest matching list item. If the response was more similar to the closest presented word than it was to 90% of the words in the dictionary, it was corrected to match that presented word, and was scored accordingly as a correct recall or prior-list intrusion; otherwise, it was treated as an extra-list intrusion. Final free recall was scored in the same manner as all other recall periods, with the exception that recalled words were marked as correct if they were presented at any point during the experiment.