

Modality effects in free recall: A retrieved-context account

Jesse K. Pazdera and Michael J. Kahana
University of Pennsylvania

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 (the inverse modality effect). To investigate and model these differences, we conducted two large-scale web-based immediate free recall experiments. In both experiments, participants studied visual and auditory word lists of varying lengths and rates of presentation. We observed typical modality and inverse modality effects, while also discovering that participants were more likely to initiate recall from recent items on auditory trials than on visual trials. However, modality effects persisted regardless of the first item recalled. Meanwhile, an analysis of intrusion errors revealed that participants were more likely on visual trials than on auditory trials to erroneously recall words from one list prior. Furthermore, words presented in the same modality as the present list intruded more often than those presented in a different modality. We next developed a retrieved-context account of the modality effect by fitting the Context Maintenance and Retrieval model to data across multiple list lengths. Through our simulations, we demonstrate that the modality effect can be explained by faster contextual drift and stronger context-to-item association formation during auditory presentation, relative to visual. Our modeling shows that modality effects can arise without hypothesizing distinct memory stores for recent and remote information. Finally, we propose that modality effects may derive primarily from the temporal dynamics of stimuli, rather than their modality.

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

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. Next, we report data from two experiments designed to provide new insights into how presentation modality influences the dynamics of recall, including how subjects initiate recall and the intrusions they make. 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 cued recall (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,

Support for this work was provided by the National Institutes of Health (MH55687). The authors would like thank Rivka Cohen for assistance in designing our simulations, 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

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. In contrast, 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 than presentation rate in the literature. A study by Grenfell-Essam et al. (2017) tested seven list lengths between two 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 precency visual advantages at longer list lengths only, providing evidence of an interaction between list length and 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 percepts initially enter a modality-specific sensory buffer prior to being categorized through linguistic and semantic processing. Recall processes can access information held in precategorical storage, but this information decays over time and can be overwritten 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, relative to visual. 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 with temporal contexts falling within the search window, the less likely each individual 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 fades 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 subsequent stimuli and recalled items with overlapping features overwrite them. 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 better resist output interference, 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

There currently 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

Retrieved-context theories of episodic memory (Howard & Kahana, 2002; Lohnas, Polyn, & Kahana, 2015; Polyn, Norman, & Kahana, 2009a) 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 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., 2009a), 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.

Experiments

To develop 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. We conducted two experiments investigating the interactions between modality, list length, and presentation rate in an immediate free recall task. To collect an extensive dataset, we deployed our experiments online via Amazon Mechanical Turk (MTurk). In Experiment 1, participants completed 16 trials, with lists of 12 or 24 words presented either visually or aurally at one of

two rates in a fully within-subjects design. 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). Therefore, in Experiment 2, we manipulated modality as a between-subjects variable, such that half of our participants completed all trials visually and half completed all trials aurally. All other manipulations were identical to those in Experiment 1. Due to their similarity, we report on both experiments in a single section, below.

Methods

Participants

This study was approved by the Institutional Review Board at the University of Pennsylvania. We recruited participants for both experiments through MTurk, a web-based crowdsourcing platform (Mason & Suri, 2012). To qualify for our study, we required participants to be located in the United States and possess at least a 95% task approval rating. Additionally, we blocked access to the study from mobile devices and tablets, and we disqualified individuals who participated in Experiment 1 from participating in Experiment 2. Participants received \$7.50 for completion of the task.

In Experiment 1, a total of 888 participants (469 male, 583 female, 48 unreported) contributed to our analyses, with ages ranging from 18 to 71 years ($M = 36.6$, $SD = 10.4$). An additional 212 participants completed the task, but were excluded due to technical issues affecting their sessions ($N = 11$) or because they met at least one of the five exclusion criteria defined in Appendix A ($N = 201$). Most were excluded for using written notes as a memory strategy ($N = 132$), as reported in a post-experiment questionnaire.

In Experiment 2, a total of 1469 participants (459 male, 640 female, 370 not reported) contributed to our analyses, with ages ranging from 18 to 71 years ($M = 36.0$, $SD = 11.0$). We randomly assigned participants to receive either visual or auditory presentation, and of these 1469 individuals, 735 participated in the visual presentation condition and 734 participated in the auditory presentation condition. An additional 531 participants completed the experiment, but were excluded due to technical issues ($N = 7$) or meeting at least one of the five exclusion criteria ($N = 524$). Participants using written notes as a memory strategy again accounted for the majority of exclusions ($N = 430$).

Data Availability

We have made all data, code, and stimuli from the present study publicly available on the Open Science Framework at <https://osf.io/4rz7k/>, as well as on GitHub at <https://github.com/jpazdera/PazdKaha22>.

Materials

The word pool for both experiments consisted of 556 English nouns, with each participant assigned a 288-word subset. The word pool was identical to that used in Experiment 4 of the Penn Electrophysiology of Encoding and Retrieval Study (Aka, Phan, & Kahana, 2021; Kahana, Aggarwal, & Phan, 2018), with the exception that we excluded 20 words possessing at least one homophone in the English language. A single female voice actress produced audio recordings of all words in the pool. The duration of auditory words ranged from 268 ms to 1054 ms. We presented visual words and distractor problems in white, 60-point Arial font on a black background.

Apparatus

We implemented both experiments in JavaScript using the jsPsych library (de Leeuw, 2015). We used PsiTurk (Gureckis et al., 2016) to host and manage the study on MTurk. We performed all analyses using a combination of Python and R.

Design

In both experiments there were three independent variables of interest: modality (Visual and Auditory), list length (12 and 24), and presentation rate (Slow and Fast). Experiment 1 used a fully within-subjects design, whereas Experiment 2 used a mixed design with modality as a between-subjects factor and list length and presentation rate as within-subjects factors.

We additionally controlled the positions of semantically related items in each list, such that every trial contained low-, medium-, and high-similarity words positioned at both adjacent and remote locations, as detailed in Appendix B. Although we do not report on semantic clustering in the present work, we controlled within-list semantics in order to enable future investigations of semantic similarity as a variable in our dataset.

To facilitate comparisons between our present experiments and those of the Penn Electrophysiology of Encoding and Retrieval Study (e.g., Aka et al., 2021; Healey, Crutchley, & Kahana, 2014; Kahana et al., 2018; Long, Danoff, & Kahana, 2015), we included an extraneous manipulation in which we varied the duration of the interlist distractor period (12 or 24 seconds). Because this manipulation does not relate to our hypotheses regarding the modality effect, and to limit the complexity of our exposition, we do not report on this

manipulation in the body of the present manuscript. However, as one might expect the duration of interlist distractors to influence prior-list intrusions, we do report an analysis of these effects in Appendix C.

Procedure

Participants first completed an audio test, in which they listened to three recorded words and were required to correctly type each into a text box. This test helped to prevent "bots" (Aguinis, Villamor, & Ramani, 2020) from accessing the experiment, and ensured that participants would be able to hear the auditory stimuli used in the study.

Participants then performed an immediate free recall task consisting of two practice trials followed by 16 experimental trials. Each trial consisted of an arithmetic distractor task, a series of 12 or 24 word presentations (18 on practice trials), and an immediate free recall test. In Experiment 1, each participant completed one experimental trial with each combination of modality, list length, presentation rate, and distractor duration in a fully randomized order. In Experiment 2, participants instead completed two trials of each list length, presentation rate, and distractor duration, all presented in the same modality.

At the start of each trial, participants completed a distractor task for either 12 or 24 seconds (18 seconds on practice trials). This task consisted of typing the answers to math problems of the form $A + B + C$, where A , B , and C were positive, single-digit integers. As answers were entered, new problems continued to appear until the full duration had elapsed. The distractor period was followed by a 2-second black screen, a 10-second visual countdown, and a fixation cross for 1.5 seconds before the first word presentation began.

During visual trials, participants viewed words onscreen. We modulated presentation rate such that each word appeared for 0.8 seconds in the fast-presentation condition and 1.6 seconds in the slow-presentation condition (1.2 seconds on practice trials). A jittered (uniformly distributed) inter-stimulus interval of 0.8–1.2 seconds followed each word. During auditory trials, participants viewed a blank screen while listening to words as a sequence of vocal recordings. As auditory words varied in duration, presentation rate on auditory trials was based on the interonset intervals of words in the visual condition, such that a new word began playing every 1.6–2.0 seconds on fast trials and every 2.4–2.8 seconds on slow trials (2.0–2.4 seconds on practice trials).

The final word in each list was followed by an additional delay of 1.2–1.4 seconds, after which a tone sounded and a row of asterisks appeared onscreen for 0.5 seconds to indicate the start of the recall period. An empty text box then replaced the asterisks, and participants were given 60 seconds to enter as many words as they could recall from the current trial, in any order. The text box was cleared after the partici-

pant entered each word, preventing them from looking back at previously-recalled words. A 2-second delay followed the end of the recall period, after which the participant was instructed to press a key to begin the next trial.

We automatically spell-checked and scored recalls using an algorithm identical to that used by Healey (2018) in a similar typed-recall task. We considered submitted words to be misspelled if they did not match any previously presented word and did not appear in Webster's Second International Dictionary (<https://libraries.io/npm/web2a>). We corrected misspellings to the most similar previously presented word (based on Damerau-Levenshtein distance; Damerau, 1964) if the misspelling was closer to that word than to 90% of the words in the dictionary.

Following the recall period of the final trial, participants received instructions for an unexpected final free recall test. During final free recall, participants spent five minutes recalling 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 received instructions on how to submit their completed assignment on MTurk for compensation.

Results

We first report the effects of modality, list length, and presentation rate on recall performance as function of serial position, focusing on primacy and recency items. We next consider differences in recall initiation behavior by condition, as well as differences in memory performance by recall start position. Then, we investigate the effects of modality on participants' tendency to commit prior-list intrusions.

Primacy and recency effects

Figure 1 illustrates the serial position curves (SPCs) for all eight combinations of modality, list length, and presentation rate. Data from both experiments show clear modality and inverse modality effects under all conditions, with an auditory advantage for recent items and a visual advantage for early-list items. To test for differences in primacy and recency between conditions, we defined primacy items as the first three words of each list and recent items as the final five words of each list. For each experiment, we fit two mixed effects logistic regression models to the proportion of primacy and recency items, respectively, that were recalled on each trial. Both models included modality (Visual and Auditory), list length (12 and 24), presentation rate (Slower and Faster), and all two- and three-way interactions between them as fixed effects, with a random intercept for each participant. We then analyzed the fixed effects via a series of Type III Wald chi-squared tests, the results of which are summarized in Table 1.

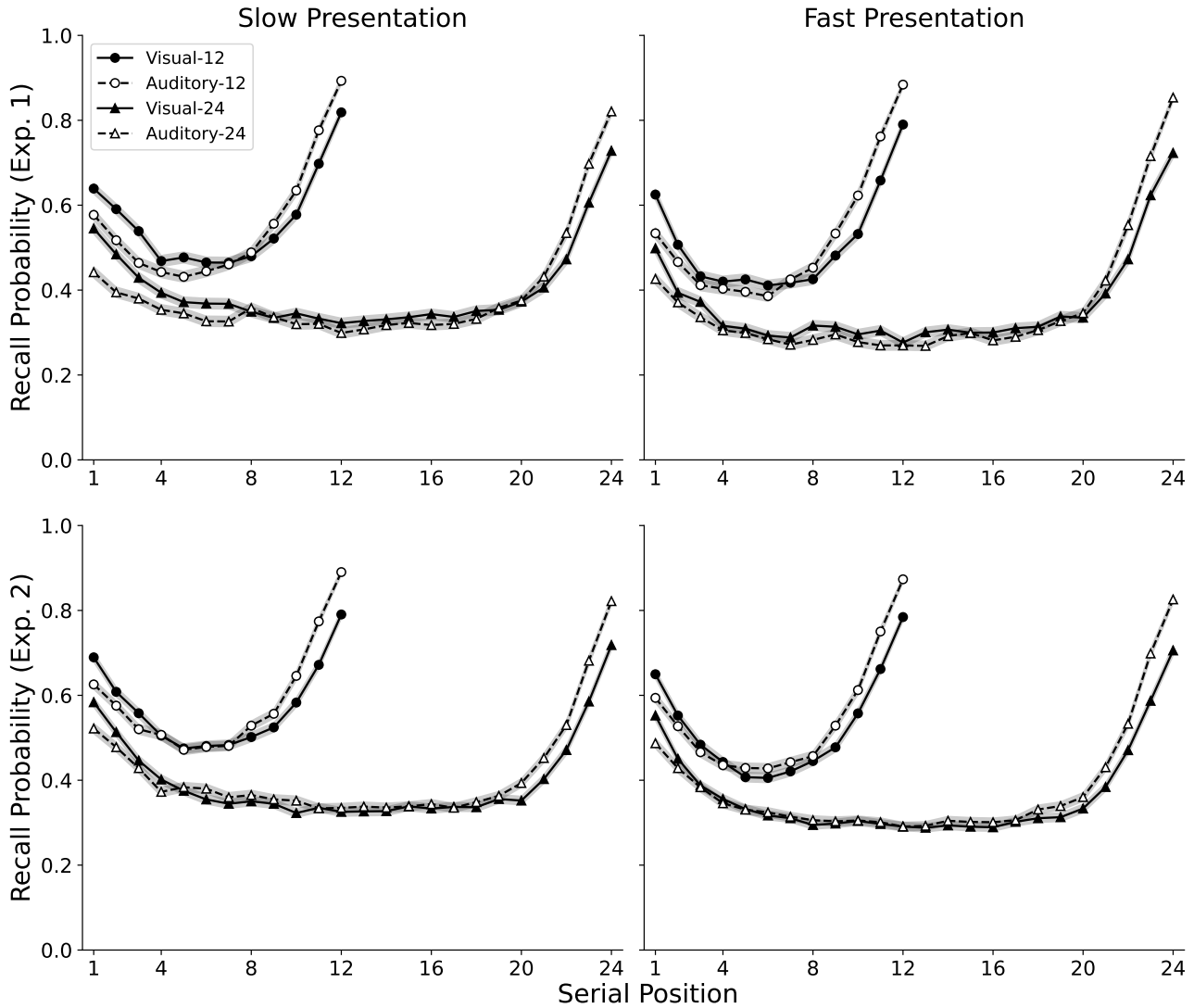


Figure 1

Recall Probability. Serial position curves for trials of each combination of modality, list length, and presentation rate in Experiments 1 (top row) and 2 (bottom row). Shaded regions indicate one standard error of the mean.

In both experiments, primacy performance showed significant main effects of modality, list length, and presentation rate, with no significant interactions between them. Recency performance also showed significant main effects of modality, list length, and presentation rate, but additionally showed a significant interaction between list length and presentation rate. No other significant interactions were observed. We observed classic modality and inverse modality effects, such that auditory presentation conferred an advantage for recalling recent items, whereas visual presentation conferred an advantage for recalling early-list items. Both primacy and recency performance were higher on 12-item lists than on

24-item lists, reflecting a general tendency for the proportion of items recalled to decrease as list length increases (e.g., Grenfell-Essam et al., 2017; Roberts, 1972). A slower presentation rate improved recall for both primacy and recency items, with the effect on recency performance being larger on 12-item lists than on 24-item lists.

Recall initiation

Figure 2 shows the probability of first recall (PFR) during trials of each combination of modality, list length, and presentation rate. Based on the analysis by Grenfell-Essam et al. (2017) of recall initiation behavior and its relation to the

Variable	Experiment 1			Experiment 2		
	χ^2	df	p	χ^2	df	p
Primacy Performance						
Modality	97.54	1	<.001	11.63	1	<.001
List Length	265.41	1	<.001	410.07	1	<.001
Presentation Rate	74.08	1	<.001	95.71	1	<.001
Modality \times List Length	0.33	1	.564	0.97	1	.324
Modality \times Presentation Rate	2.59	1	.108	0.41	1	.521
List Length \times Presentation Rate	1.47	1	.225	0.73	1	.392
Modality \times List Length \times Presentation Rate	0.26	1	.609	1.57	1	.210
Recency Performance						
Modality	189.68	1	<.001	114.60	1	<.001
List Length	291.78	1	<.001	512.92	1	<.001
Presentation Rate	5.07	1	.024	12.14	1	<.001
Modality \times List Length	1.08	1	.300	0.02	1	.881
Modality \times Presentation Rate	3.62	1	.057	<0.01	1	.953
List Length \times Presentation Rate	12.61	1	<.001	10.59	1	.001
Modality \times List Length \times Presentation Rate	0.16	1	.687	2.18	1	.140

Table 1*Wald Chi-Squared Tests for Factors Predicting Primacy and Recency Performance*

modality effect, we analyzed the proportion of trials where participants initiated recall from the first list item, as well as the proportion of trials where they initiated recall from one of the final four list items. To test the effects of modality, list length, and presentation rate on the probability of first recall we fit two mixed effects logistic regression models to the data from each experiment—the first to predict recall initiation from Serial Position 1 and the second to predict recall initiation from any of the final four serial positions. Both models incorporated fixed effects of modality (Visual and Auditory), list length (12 and 24), presentation rate (Slower and Faster), and all two- and three-way interactions, as well as a random intercept for each participant. Significance testing was performed using a series of Type III Wald chi-squared tests, which are summarized in Table 2.

In both experiments, modality and list length significantly affected the probability of initiating recall from Serial Position 1. Participants were more likely to initiate recall from the first item on visual lists than on auditory lists, and they were more likely to initiate recall from the first item of 12-item lists than of 24-item lists. Experiment 2 additionally showed a significant effect of presentation rate that was not present in Experiment 1, such that participants were more likely to initiate recall from Serial Position 1 on slower trials than on faster trials. No interaction effects reached significance.

Modality significantly affected recall initiation from the final four serial positions in both experiments, such that participants were more likely to initiate recall from the final four

items of auditory lists than of visual ones. Experiment 1 identified additional significant main effects of list length and presentation rate, such that participants were more likely to initiate recall from the last four items of 12-item lists than of 24-item lists and the last four items of faster lists than of slower lists. Furthermore, list length and presentation rate interacted such that slower item presentation increased the effect of list length. In Experiment 2, these effects of list length and presentation rate were replaced by a three-way interaction, characterized by a larger effect of modality with slower presentation on 12-item trials, but a larger effect of modality with faster presentation on 24-item trials.

Recall performance by first item recalled

Given that modality significantly impacted recall initiation in both experiments, we further investigated the possibility that differences in recall initiation produced the modality and inverse modality effects we observed. To do so, we calculated separate start-conditional SPCs for trials where participants initiated recall from Serial Position 1 (SP1) or from any of the final four serial positions (L4). Figure 3 shows the average SPCs for each modality and list length, conditional on whether recall began from the first or final four words of the list. Consistent with Craik (1969) and Grenfell-Essam et al. (2017), we observed modality and inverse modality effects regardless of the first item recalled.

To determine whether the modality and inverse modality effects varied with recall start position, we fit two mixed effects logistic regression models to the data from each exper-

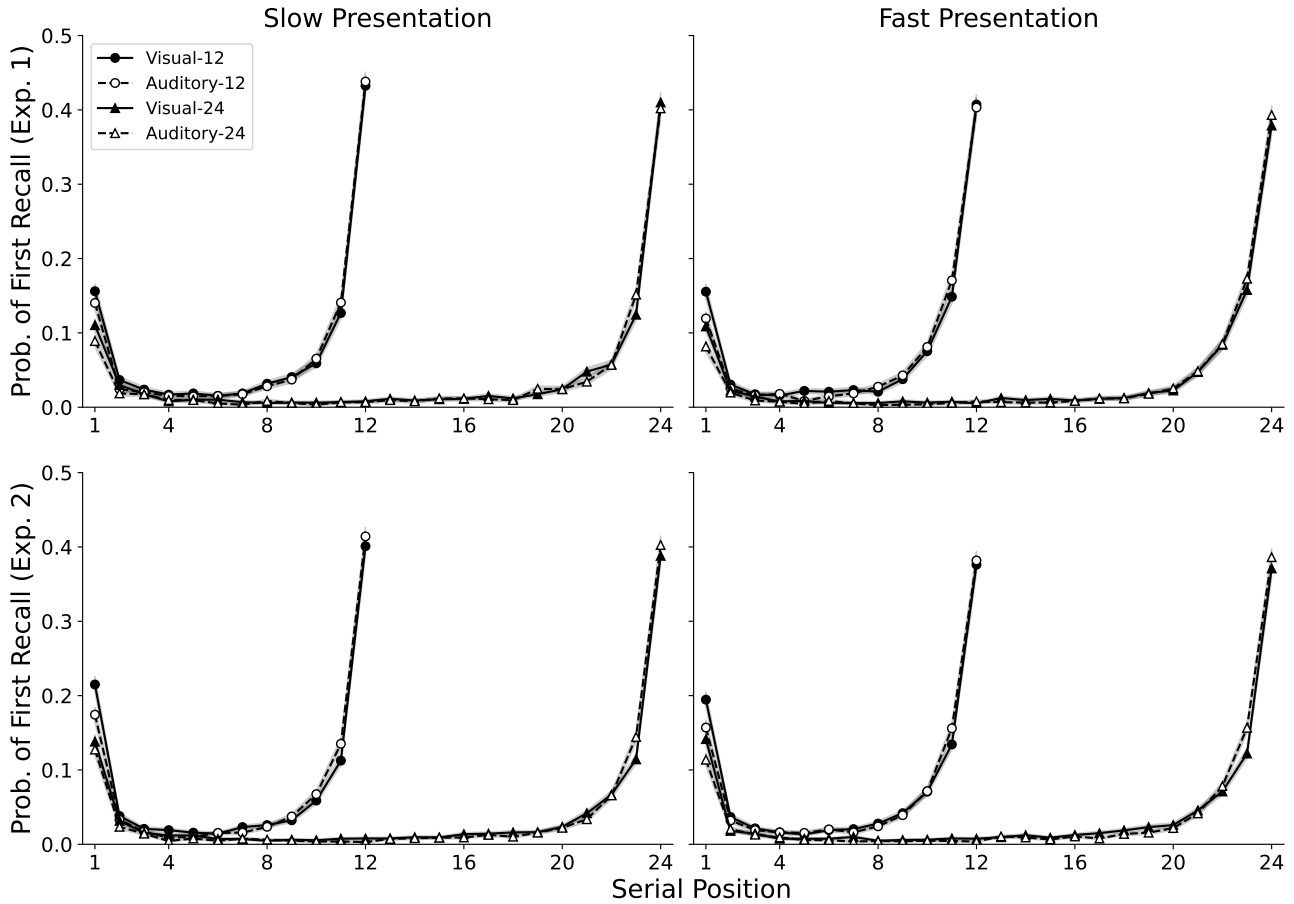


Figure 2

Recall Initiation. Probability of first recall on trials of each combination of modality, list length, and presentation rate in Experiments 1 (top row) and 2 (bottom row). Shaded regions indicate one standard error of the mean.

iment to predict primacy and recency performance, respectively. These models included modality (Visual and Auditory), recall start position (SP1 and L4), and the interaction between them as fixed effects, with a random intercept for each participant. Type III Wald chi-squared tests conducted on both models for both experiments revealed significant main effects of modality (Experiment 1 only for primacy performance) and start type, with no significant interaction (see Table 3). We observed a larger primacy effect when recall initiated from the beginning of the list and a larger recency effect when recall began from one of the last four items of the list. However, recall start position did not significantly impact the size of the modality effect, nor the inverse modality effect. It therefore does not appear that the modality and inverse modality effects originate solely from differences in recall initiation behavior between modalities.

Prior-list intrusions (PLIs) and PLI recency

Prior-list intrusions (PLIs) occur when a participant erroneously recalls a word that was presented on a previous trial. In analyzing the effect of modality on intrusion behaviors in our experiments, we examined both the quantity of PLIs that participants made and the recency of those intrusions (cf. Zaromb et al., 2006) under each modality condition. As modality varied between trials in Experiment 1, PLIs may have been influenced not only by their original presentation modality, but also by the modality of the trial during which they were recalled. As such, our PLI analyses for Experiment 1 followed a 2 (Modality at Encoding: Visual and Auditory) \times 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 presentation modality of the trial where the intrusion took place. For con-

Variable	Experiment 1			Experiment 2		
	χ^2	df	p	χ^2	df	p
Probability of First Recall = Serial Position 1						
Modality	30.41	1	<.001	9.66	1	.002
List Length	94.73	1	<.001	191.88	1	<.001
Presentation Rate	2.83	1	.092	8.94	1	.003
Modality \times List Length	0.35	1	.556	1.97	1	.160
Modality \times Presentation Rate	2.07	1	.150	1.74	1	.187
List Length \times Presentation Rate	0.22	1	.639	1.70	1	.192
Modality \times List Length \times Presentation Rate	0.38	1	.540	1.69	1	.194
Probability of First Recall = Last Four						
Modality	11.27	1	<.001	6.44	1	.011
List Length	4.19	1	.041	<0.01	1	.983
Presentation Rate	17.01	1	<.001	1.89	1	.169
Modality \times List Length	0.31	1	.577	0.46	1	.497
Modality \times Presentation Rate	1.66	1	.198	0.10	1	.757
List Length \times Presentation Rate	4.40	1	.036	0.03	1	.864
Modality \times List Length \times Presentation Rate	0.60	1	.438	4.22	1	.040

Table 2

Wald Chi-Squared Tests for Factors Predicting Probability of First Recall

Variable	Experiment 1			Experiment 2		
	χ^2	df	p	χ^2	df	p
Primacy Performance						
Modality	16.16	1	<.001	3.61	1	.057
Start Position	558.87	1	<.001	959.57	1	<.001
Modality \times Start Position	0.23	1	.632	0.43	1	.514
Recency Performance						
Modality	92.37	1	<.001	135.96	1	<.001
Start Position	344.83	1	<.001	568.17	1	<.001
Modality \times Start Position	0.68	1	.410	2.03	1	.154

Table 3

Wald Chi-Squared Tests for Recall Start Position as a Predictor of Modality Effects

venience, we will refer to the four modality conditions 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. Manipulating modality between-subjects in Experiment 2 produced a simpler design in which the modality at encoding always matched the modality at retrieval. As such, there were only two modality conditions (Visual and Auditory) for PLIs in Experiment 2.

To determine whether modality affected the quantity of PLIs participants made in Experiment 1, we first counted how many PLIs each participant made from each of the four modality conditions. We then divided each participant's totals by the number of trials on which it was possible for them

to make each type of PLI. This procedure revealed the expected number of PLIs per trial, conditional on those PLIs being possible. Such a normalization scheme is necessary in Experiment 1 because no greater than two of the four modality types were possible for PLIs on any given trial. For example, only V-A and A-A intrusions could have been made during auditory trials. The values resulting from this procedure appear in Figure 4A. We examined differences in the number of PLIs by condition using a 2 (Modality at Encoding: Visual and Auditory) \times 2 (Modality at Retrieval: Visual and Auditory) repeated measures ANOVA. Neither the main effect of modality at encoding, nor the effect of modality at retrieval was significant, both $F(1, 887) < 1$. However, the two-way interaction between modality at encoding and modality at re-

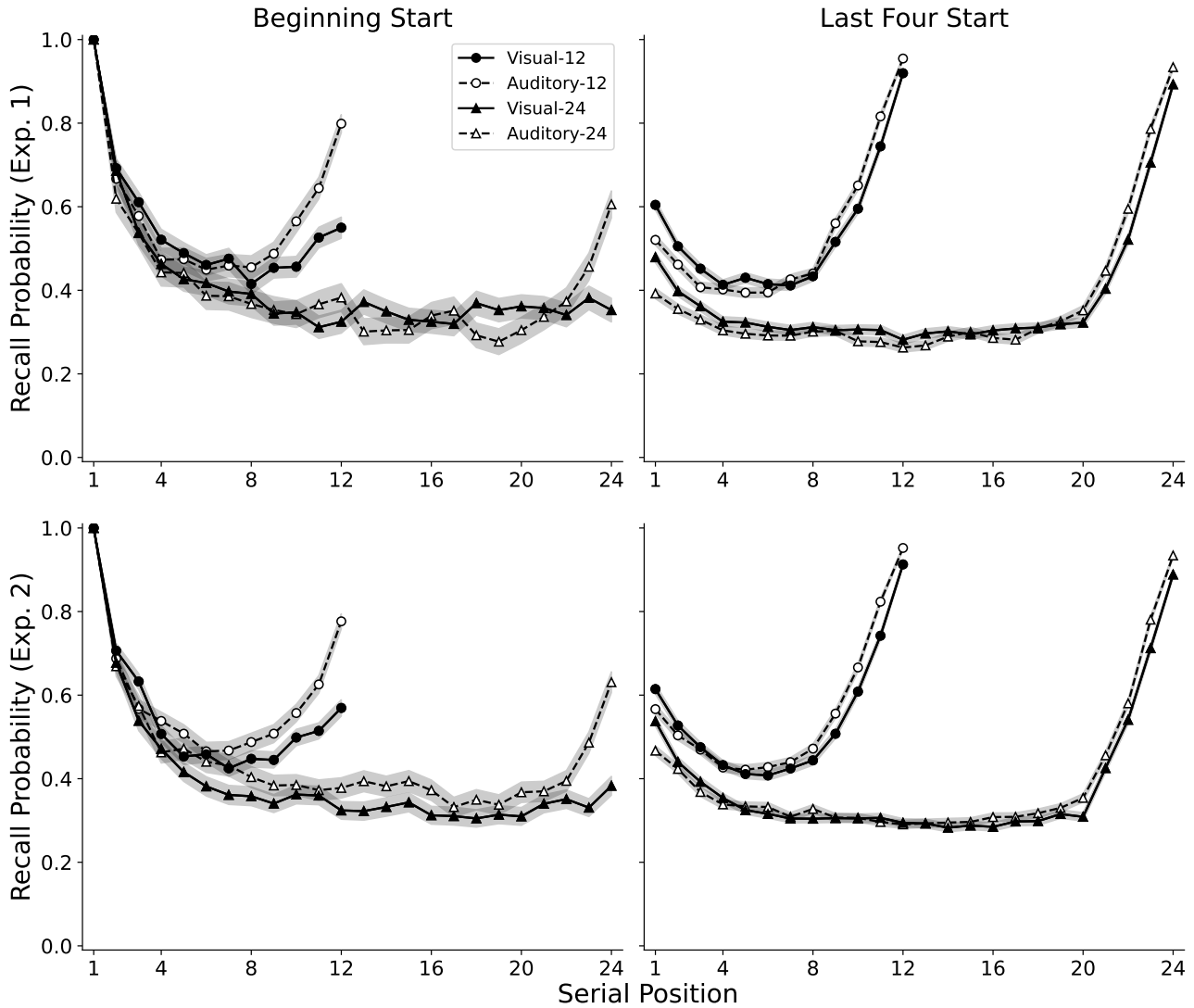


Figure 3

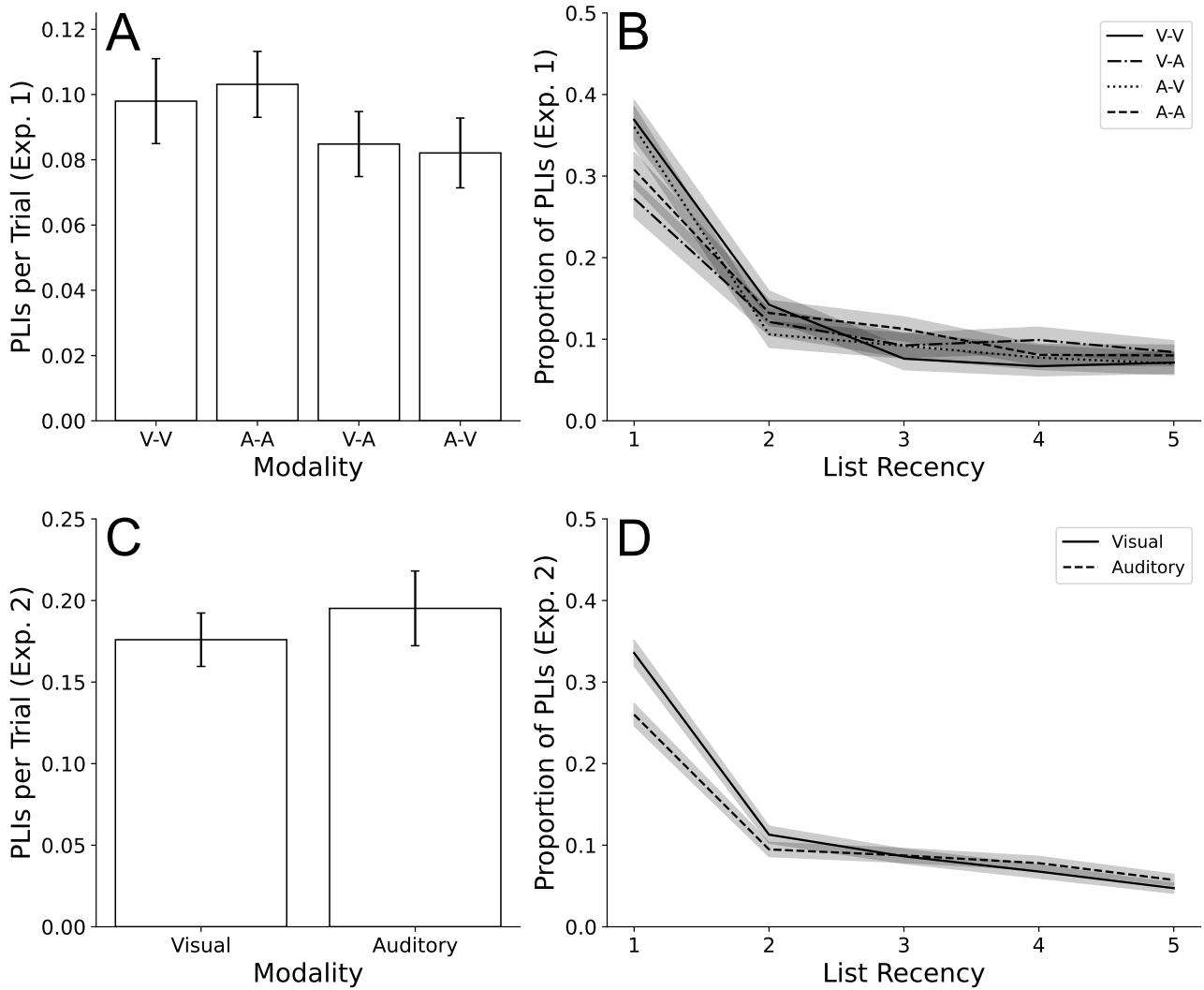
Recall Probability by First Item Recalled. Average serial position curves across all 12- and 24-item trials where recall started from the first list item versus any one of the last four list items in Experiments 1 (top row) and 2 (bottom row). Shaded regions indicate one standard error of the mean across trials.

trieval was significant, $F(1, 887) = 15.65$, $MS_e = 0.26$, $p < .001$, $\eta_p^2 = .017$. These results indicate that participants were similarly likely to intrusively recall visual and auditory words, and made similar numbers of PLIs on visual and auditory trials. However, participants made significantly more same-modality (V-V and A-A) PLIs than cross-modality (V-A and A-V) ones.

Figure 4C shows the number of PLIs committed per trial by participants in each modality condition in Experiment 2. An independent samples t -test did not detect a significant difference in the number of PLIs made by participants in

the visual and auditory presentation conditions, $t(1467) = 1.34$, $p = .18$. Consistent with Experiment 1, presentation modality did not affect the number of PLIs that participants recalled.

We next calculated PLI recency curves for each modality type. To do so, we first counted how many PLIs of that type each participant made from each lag (ranging from 1 to 15 lists back). Then, we counted the number of trials on which it was possible to make a PLI of that modality and lag. Next, we divided the PLI totals by the trial counts, giving the expected number of PLIs from each lag, conditional on such

**Figure 4**

Prior-List Intrusions by Modality. **A.** Number of PLIs per trial in Experiment 1. **B.** PLI recency in Experiment 1. **C.** Number of PLIs per trial in Experiment 2. **D.** PLI recency in Experiment 2. For Experiment 1, the first letter in each modality label indicates the original presentation modality of the intruding word and the second letter indicates the modality of the trial where the intrusion occurred. In Experiment 2, the modality at encoding and retrieval were always identical. Scores in Experiment 1 are conditional on PLIs of that modality being possible. Shaded regions indicate one standard error of the mean, and error bars indicate 95% confidence intervals.

PLIs being possible. Finally, we divided these 15 expected values by their sum, giving the proportion of PLIs expected to originate from each number of lists back. These final values comprise the PLI recency curve. Each participant's first two experimental trials were excluded from our PLI recency calculations.

Figures 4B and 4D illustrate these PLI recency curves for list lags up to five in Experiments 1 and 2, respectively. To analyze the effects of modality on the recency of PLIs, we

compared the proportion of PLIs expected to originate from a list lag of 1 under each condition. We fit a linear mixed effects regression model to the data from Experiment 1, including modality at encoding, modality at retrieval, and the interaction between them as fixed effects, as well as a random intercept for each participant. A series of Type III Wald chi-squared tests revealed a significant effect of modality at retrieval, $\chi^2(1) = 10.84, p < .001$, with a nonsignificant effect of modality at encoding and a nonsignificant two-way

interaction, both $\chi^2(1) < 1$. This test confirmed that PLIs on visual trials were more likely than those on auditory trials to originate from one list back. Therefore, although participants made similar numbers of PLIs during visual and auditory trials, PLIs during visual trials tended to be more recent than those during auditory trials. In contrast, the original presentation modality of the intruding word had no reliable influence on intrusion recency.

In Experiment 2, an independent samples *t*-test confirmed that PLIs committed by participants in the visual presentation condition were significantly more likely than those of participants in the auditory condition to originate from one list prior, $t(1126) = 3.66, p < .001$, replicating the novel effect of modality at retrieval that was observed in Experiment 1.

Discussion

We investigated the effects of modality on recall behavior, as well as modality's interaction with list length and presentation rate in two large-scale, immediate free recall experiments. At list lengths of 12 and 24 and across two presentation rates, we observed robust modality and inverse modality effects, characterized by auditory presentation producing reduced primacy and improved recency performance, when compared with visual presentation (Figure 1). Furthermore, we found in both experiments that auditory presentation reduced the probability of initiating recall from the first item of a list, while increasing the probability of initiating recall from one of the final four items of a list (Figure 2). This finding corrects previous conclusions that presentation modality does not affect probability of first recall (Grenfell-Essam et al., 2017). These modality differences in recall initiation did not account for the differences in primacy and recency performance, however, as modality and inverse modality effects persisted regardless of recall start location (Figure 3), confirming similar findings from prior work (Cowan et al., 2002; Craik, 1969; Grenfell-Essam et al., 2017; Harvey & Beaman, 2007).

The list lengths used in these experiments affected neither the modality effect nor the inverse modality effect. This result contrasts with recent observations that both effects expand as list length increases from two to 12 (Grenfell-Essam et al., 2017), but is consistent with earlier observations that the modality effect remains stable between list lengths of ten and 40 (Roberts, 1972). We therefore suggest that the modality and inverse modality effects may expand with list length until reaching an upper limit at list lengths of approximately 12.

Conspicuous by its absence was an interaction between modality and presentation rate in determining primacy and recency performance. Prior research has suggested that faster item presentation results in a larger modality effect (e.g., Murdock & Walker, 1969; Murray, 1966), and possibly an attenuated inverse modality effect (Macken et al., 2016). Our

data did not replicate these findings. One possible explanation is that our faster presentation rate may not have been fast enough to evoke the effects observed by prior studies.

Analyses of prior-list intrusion errors in our experiments revealed four notable findings (Figure 4). First, participants intrusively recalled a similar number of visual and auditory words as PLIs. Second, visual and auditory trials produced similar quantities of PLIs. Third, words were more likely to intrude on trials of their own modality than on trials of a different modality. Fourth, PLIs were more likely to originate from one list prior on visual trials than on auditory trials.

The first three intrusion results are consistent with prior observations from a cued recall study by Murdock (1966), which found that intrusions tended to originate from the same modality as the test pair, but that visual and auditory presentation produced equal numbers of intrusions overall. Our results extend these effects to free recall, and indicate that trials and items of the same modality are less distinct in memory than those of different modalities. This effect may be driven by the encoding of modality-specific information, as captured in Nairne's (1990) feature model. Alternatively, it may reflect the coding of source features, whereby items form associations with the type of task in which they appeared (e.g., Polyn, Norman, & Kahana, 2009b).

Our fourth PLI finding is a novel discovery which suggests that visual trials are more confusable than auditory trials with their immediately preceding lists. Within a retrieved-context framework, we suggest two possible explanations for this effect. The first possibility is that the inverse modality effect increases PLI recency on visual trials as a side-effect of improved primacy performance. The contextual states of early-list items are relatively similar to the context of the immediately preceding trial; therefore, an increased probability of recalling primacy items should increase the reinstatement of early-list contexts, which in turn would increase the likelihood of making lag-1 PLIs. The second possibility is that context shifts to a greater extent over the course of an auditory list than over the course of a visual list. This effect would result in the recall period of an auditory list having a context that is especially distinct from the previous trial's context, potentially reducing the probability of making lag-1 PLIs on auditory trials.

A Retrieved-Context Model of the Modality Effect

We next asked whether the Context Maintenance and Retrieval model (CMR2; Lohas et al., 2015; Healey & Kahana, 2016) could account for the effects of modality in our data. The sections below provide an overview of CMR2, describe three sets of simulations, and present our retrieved-context account of the modality effect.

Model overview

CMR2 assumes that items form associations with the contexts in which they appear. As in earlier papers, we model items as basis vectors in a high-dimensional feature space, \mathbf{f} . We model context as a vector, \mathbf{c} , containing a recency-weighted sum of past contextual states. Features and context interact via a pair of associative matrices, M^{FC} and M^{CF} . Presenting item i to the model creates a new input to context,

$$\mathbf{c}_i^{\text{IN}} = \frac{M^{FC}\mathbf{f}_i}{\|M^{FC}\mathbf{f}_i\|}, \quad (1)$$

following which context evolves according to the equation:

$$\mathbf{c}_i = \rho_i \mathbf{c}_{i-1} + \beta \mathbf{c}_i^{\text{IN}}. \quad (2)$$

The parameter β determines the rate at which context changes with each studied item. ρ_i is a constant defined such that $\|\mathbf{c}_i\| = 1$. The rate of contextual drift may differ between encoding and recall events, denoted β_{enc} and β_{rec} , respectively.

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 Hebbian learning rule,

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

where parameters γ_{FC} and γ_{CF} scale the strength of newly learned associations. We model the decay of attention over the course of a trial by scaling associations in M^{CF} by a primacy factor, ϕ , where the factor ϕ_k for the k^{th} item of a list is:

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

Model parameters ϕ_s and ϕ_d govern the strength and decay of the primacy gradient, respectively. To represent the semantic associations between items, we scale the initial values of M^{CF} by the cosine similarity between the Word2vec vector representations of each pair of items (Mikolov, Chen, Corrado, & Dean, 2013) multiplied by the model’s semantic scaling parameter, s .

To recall items, the model uses the current context, \mathbf{c}_t , as a retrieval cue to activate a vector of items in memory, $\mathbf{a} = M^{CF}\mathbf{c}_t$. The activated items enter a leaky, competitive accumulator process that determines which item the model recalls (Usher & McClelland, 2001). A vector of evidence, \mathbf{x} , tracks the accumulating activation of each element in \mathbf{a} . The vector of evidence evolves at each step of the accumulator process according to the equation:

$$\begin{aligned} \mathbf{x}_n &= \mathbf{x}_{n-1} - \tau \kappa \mathbf{x}_{n-1} - \tau \lambda \mathbf{N} \mathbf{x}_{n-1} + \tau \mathbf{a} + \boldsymbol{\epsilon}, \\ \boldsymbol{\epsilon} &\sim \mathcal{N}(0, \tau \eta), \\ \mathbf{x}_n &\rightarrow \max(\mathbf{x}_n, 0). \end{aligned} \quad (5)$$

Thus, at each step, previous evidence decays due to a leak parameter, κ . Evidence also decays due to lateral inhibition from other items, which we scale by the model parameter λ (\mathbf{N} is a square matrix with 0 on the diagonal and 1 elsewhere, such that $\mathbf{N}\mathbf{x} = \sum_i x_i - \mathbf{x}$). New evidence accumulates from incoming activation, \mathbf{a} , as well as from a noise vector, $\boldsymbol{\epsilon}$, which we draw from a normal distribution with mean 0 and standard deviation determined by the model parameter η . τ is a fixed time constant that scales the change in evidence on each time step. Values in the accumulator cannot decay below zero.

The accumulator process runs iteratively until the recall period ends or one of the accumulating elements crosses a threshold, Θ_i , signifying its retrieval from memory. To filter out undesirable recalls (e.g., prior-list intrusions), the model compares the current context to the context associated with the retrieved item, $u = \mathbf{c}_{t+1}^{\text{IN}} \cdot \mathbf{c}_t$, and overtly recalls the item only if u exceeds a threshold defined by the model parameter c_{thresh} . Regardless of whether the model overtly recalled the retrieved item, the retrieval activates the associated element of the feature vector, and context updates with a drift rate of β_{rec} according to Equation 2. The updated state of context, \mathbf{c}_{t+1} , then cues a new activation vector, \mathbf{a} , and a new accumulator process begins.

CMR2 incorporates repetition suppression by increasing an item’s retrieval threshold to a maximum value after its retrieval, determined by the model parameter ω . The threshold then decays back towards the baseline value of $\Theta_i = 1$, as a function of the number of subsequently retrieved items, j , and the model parameter α :

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

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 $\beta_{\text{post}}^{\text{recall}}$. 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.

Modeling the modality effect

Two prominent accounts of modality effects have suggested that auditory processing of items either leads to enhanced temporal discriminability (Glenberg & Swanson, 1986) or to richer encoding of features (Nairne, 1990). Here we consider ways in which each of these hypotheses can be instantiated within CMR2.

We propose that the contextual drift rate, β_{enc} , determines the temporal distinctiveness of items in memory. A larger drift rate reduces the representational similarity between

studied items by decaying past contexts more quickly (Equation 2). Therefore, temporal distinctiveness theory (Glenberg & Swanson, 1986) suggests a larger β_{enc} during auditory presentation than during visual presentation.

Whereas Nairne (1990) hypothesized that auditory stimuli carry more modality-specific features than do visual stimuli, CMR2 uses a 'localist' representation as a simplifying assumption, whereby a single (orthogonal) feature represents a given item. We therefore selected context-to-item association strength, γ_{CF} , as a proxy for richness of encoding (see Equation 3). Nairne's (1990) feature model suggests that auditory processing should be consistent with a greater γ_{CF} compared to visual processing.

We conducted three sets of simulations focusing on the between-subject¹ modality-effect data from Experiment 2. Given that we observed similar modality and inverse modality effects across list lengths and presentation rates, we simplified Simulations 1 and 2 by pooling across presentation rates and fitting data from only the shorter (12-item) lists. Simulation 3 then evaluated CMR2's generality across varied list lengths, incorporating data from our longer (24-item) lists, as well as data from shorter lists reported by Grenfell-Essam et al. (2017). We implemented the CMR2 model and all three simulations in Python, and have made the associated code publicly available on the Open Science Framework at <https://osf.io/4rz7k/>, as well as on GitHub at <https://github.com/jpazdera/PazdKaha22>.

Simulation 1: Pooled-modality data

We first sought to identify a set of parameters under which CMR2 fit the grand average data from Experiment 2, pooled across modalities and presentation rates. We used a constriction-based particle swarm optimization algorithm (Clerc & Kennedy, 2002; Eberhart & Shi, 2000) to update and evaluate 200 particles for 200 iterations, using the parameter set associated with each particle's location in a 14 dimensional space to simulate all 1,469 sessions from Experiment 2. Simulated sessions used the exact sets of word lists presented to our participants, with the exception that we held list length constant at 12 items by presenting only the first 12 items of each experimental trial to our models.

On each iteration of the particle swarm, we 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 by the mean squared error between model and data, weighted by the inverse of the standard error of the data (analogous to a χ^2 goodness-of-fit statistic). The following 41 data points contributed to the model's fitness: points two through 12 of the serial position curve for trials where recall began from Serial Position 1, all 12 points of the serial position curve for trials where recall initiated from one of the final four serial positions, the probability of first recall

from all 12 serial positions, the average number of PLIs per trial, and the PLI recency score for list lags of one through five.

After completing the particle swarm, we identified the best parameter set discovered by each particle and used it to simulate all sessions of Experiment 2 five times, in order to precisely distinguish between models with similarly good fits. We then recalculated each model's goodness of fit based on this more precise simulation and selected the parameter set with the lowest error as our final best fit.

Simulation 2: The modality effect

Beginning from the more general model of our data developed in Simulation 1, we next endeavored to simulate the effects of modality by systematically varying β_{enc} and 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 fit in Simulation 1, allowing β_{enc} to vary by up to ± 0.1 and γ_{CF} to vary by up to ± 0.15 . Within this search space, we selected parameter values evenly spaced at 101 intervals along each dimension, producing a total of 10,201 unique parameter sets.

We used each of these parameter sets to simulate all 1,469 sessions of Experiment 2 five times, and we calculated each model's start-conditional serial position curves for trials where recall began from the first or final four list items. To determine which parameter set best simulated visual recall, we computed the mean squared error (weighted by the inverse of the data standard error) between each model's simulated start-conditional SPCs and the average data from visual trials in Experiment 2. We selected the combination of β_{enc} and γ_{CF} that produced the closest fit as our model of visual recall. We then repeated this process to identify the parameter set that best simulated auditory recall.

Simulation 3: Modality effects across list lengths

Our final set of simulations generalizes our account of modality effects across list lengths. We focused on simulating two data sets: the six-item immediate free recall data from Experiment 1 of Grenfell-Essam et al. (2017) and the 24-item trials from our own Experiment 2.

We began by repeating the particle swarm optimization algorithm from Simulation 1 on each of these two datasets separately, in order to identify parameter sets that approximated the pooled-modality behavioral data from each. The optimization process was identical to that used for Simulation 1, with the exception that we instead simulated 500 sessions containing 16 lists of six or 24 unique, random words

¹We focused exclusively on simulating between-subject modality conditions to avoid the technical complexities associated with defining separate model parameters for different modalities within a single experimental session.

drawn from the present study’s word pool. The same 500 sessions were simulated by all models across all iterations of the particle swarm. Following the particle swarm, the final best fit was selected based on a simulation of 2,500 unique sessions.

This procedure produced two sets of parameters that approximated the average performance observed on six- and 24-item lists in these two datasets. We next repeated the procedure of Simulation 2 for each dataset, fitting modality-specific behavior by allowing only β_{enc} and γ_{CF} to vary. We used an identical grid search procedure to Simulation 2, and used each parameter set to simulate the same 2,500 sessions used for finalizing the best fit to the average data.

For the simulation of six-item lists, we assessed goodness of fit as the mean squared error between each model’s behavior and the following 17 data points extracted from Figure 4, Figure 5, and Appendix C of Grenfell-Essam et al. (2017) and pooled across modalities: points two through six of the SPC for trials where recall began from Serial Position 1, all six points of the SPC for trials where recall began from one of the final four serial positions, and all six points of the PFR curve. Akin to Simulation 2, only the 11 points of the start-conditional SPCs contributed to the modality-specific fits.

For the simulation of 24-item lists, we assessed goodness of fit as the mean squared error (weighted by the inverse of the data standard error) between the model’s behavior and the following 77 data points from Experiment 2 of the present study: points two through 24 of the SPC for trials where recall began from Serial Position 1, all 24 points of the SPC for trials where recall began from one of the final four serial positions, all 24 points of the PFR curve, the average number of PLIs per list, and the PLI recency score for list lags of one through five. Only the 47 points of the start-conditional SPCs contributed to the modality-specific fits.

Results

Simulation 1 examined CMR2’s fit to the pooled-modality data from Experiment 2. Simulation 2 examined our model’s ability to simulate the modality effect by varying two parameters: the contextual drift rate, β_{enc} and the strength of learned context-to-item associations, γ_{CF} , with the remaining parameters set to the values obtained in Simulation 1. We additionally examined the model’s predictions regarding the effects of modality on recall initiation, intrusion behavior, and performance by recall start position. Simulation 3 examined CMR2’s ability to account for the modality effect for short and long lists.

Simulation 1

Table 4 lists the set of parameters that best fit data from 12-item trials, pooled across modalities. The behavioral predictions of the best-fitting model can be found in Figure 5. CMR2 closely replicated the serial position curves

Table 4

Best-Fitting CMR2 Parameters by List Length and Modality

6-Item Lists			
Param.	Pooled	Visual	Auditory
β_{enc}	0.420	0.344	0.510
β_{rec}	0.011	-	-
γ_{FC}	0.536	-	-
γ_{CF}	0.249	0.234	0.288
ϕ_s	3.198	-	-
ϕ_d	1.441	-	-
κ	0.225	-	-
λ	0.045	-	-
η	0.017	-	-
s	1.971	-	-
β_{post}^{recall}	0.654	-	-
ω	12.224	-	-
α	0.960	-	-
c_{thresh}	0.437	-	-
12-Item Lists			
Param.	Pooled	Visual	Auditory
β_{enc}	0.439	0.427	0.457
β_{rec}	0.206	-	-
γ_{FC}	0.165	-	-
γ_{CF}	0.533	0.518	0.569
ϕ_s	2.515	-	-
ϕ_d	0.536	-	-
κ	0.069	-	-
λ	0.063	-	-
η	0.007	-	-
s	1.912	-	-
β_{post}^{recall}	0.691	-	-
ω	7.682	-	-
α	0.842	-	-
c_{thresh}	0.022	-	-
24-Item Lists			
Param.	Pooled	Visual	Auditory
β_{enc}	0.416	0.404	0.416
β_{rec}	0.102	-	-
γ_{FC}	0.495	-	-
γ_{CF}	0.447	0.417	0.450
ϕ_s	3.981	-	-
ϕ_d	0.182	-	-
κ	0.107	-	-
λ	0.111	-	-
η	0.074	-	-
s	1.761	-	-
β_{post}^{recall}	0.683	-	-
ω	12.489	-	-
α	0.853	-	-
c_{thresh}	0.073	-	-

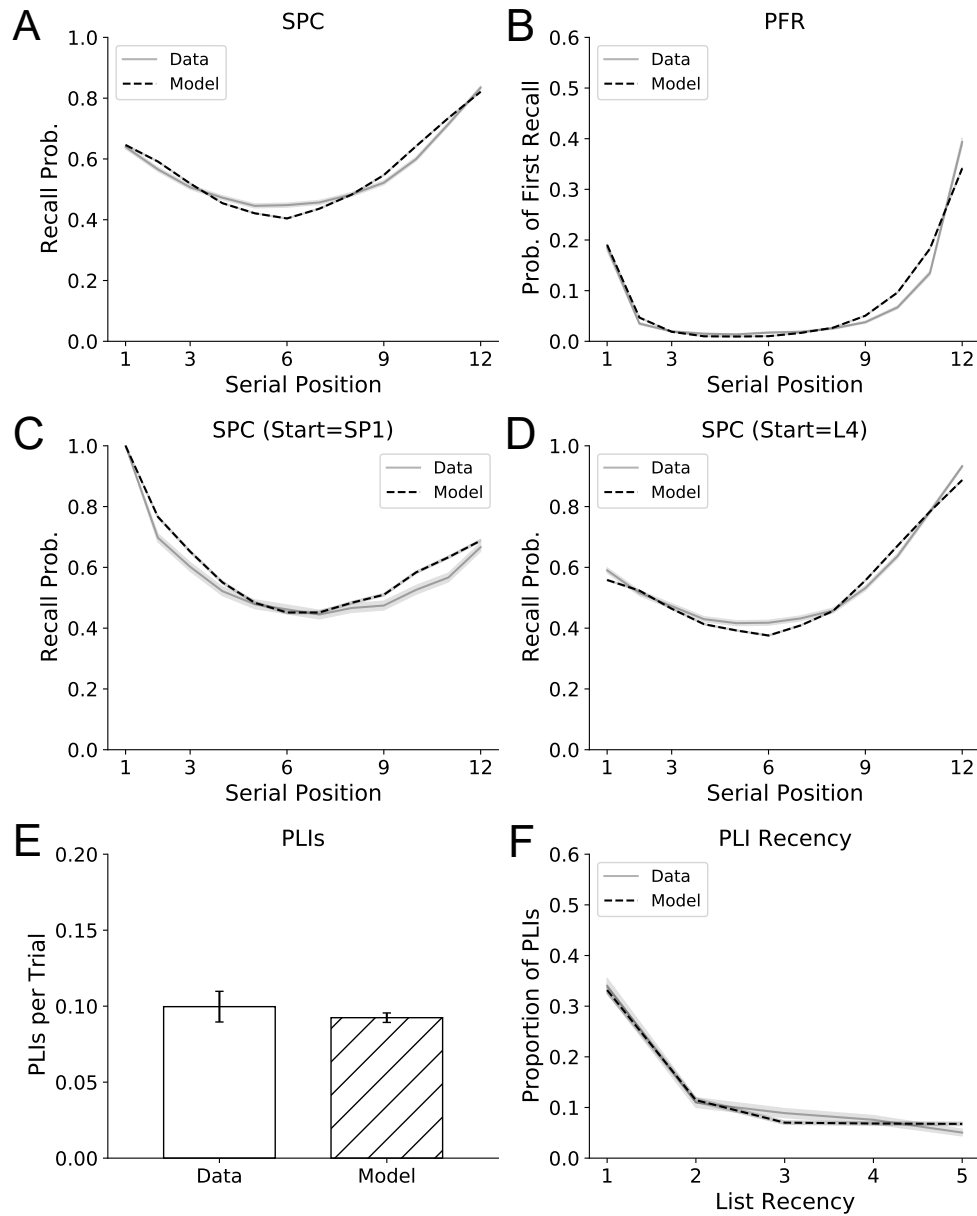


Figure 5

Simulation of Pooled-Modality Data. Empirical data from Experiment 2, pooled across modalities, are plotted against the best-fitting model predictions from Simulation 1. The best fit was determined by the mean squared error (inversely weighted by the data standard error) between model and data for the probability of first recall (**B**), start-conditional serial position curves (**C** and **D**), average prior-list intrusions per list (**E**), and PLI recency curve (**F**). The serial position curve across all trials (**A**) did not contribute to the goodness-of-fit metric, but is included for completeness. Shaded regions indicate one standard error of the mean, and error bars indicate 95% confidence intervals.

(Figures 5A, 5C, and 5D), probabilities of first recall (Figure 5B), and intrusion behavior of participants (Figure 5E). CMR2 also successfully captured participants' tendency to preferentially initiate recall from either the first item or last four items of the list, as well as their tendency to make more intrusions from the immediately preceding trial than from earlier trials (Figure 5F). Overall, CMR2 provided a good qualitative fit to the pooled-modality data from the 12-item lists in Experiment 2. Simulation 2 used the best-fitting parameters obtained from these fits to evaluate a two-parameter model of the modality effect.

Simulation 2

As noted above, we evaluated whether CMR2 could account for the modality effect by varying only two parameters: β_{enc} and γ_{CF} . The heat maps in Figure 6 illustrate the results of a grid search over these parameters. The color of these maps indicates the goodness of fit to each modality's serial position curves (conditional on the first item recalled) at each point in the parameter space. The heat maps show global minima for both modalities lying near the diagonal of the parameter space, suggesting that a two-parameter account best explains the modality effect.

Table 4 shows the values of β_{enc} and γ_{CF} that best fit the behavior of participants in the visual and auditory conditions. As predicted, recall performance in the auditory task was best simulated by a faster drift rate at encoding and strengthened context-to-item associations ($\beta_{enc}^{aud} = 0.457$, $\gamma_{CF}^{aud} = 0.569$), relative to performance in the visual task ($\beta_{enc}^{vis} = 0.427$, $\gamma_{CF}^{vis} = 0.518$). Figure 7 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 regardless of recall start position. This auditory advantage was similar in extent and magnitude to the modality effects observed in our empirical data. The model showed little to no early-list visual advantage in the start-conditional SPCs (Figures 7D and 7F), but predicted a prominent visual primacy advantage in the general SPC (Figure 7B) that was not directly fit by the grid search.

By fitting our modality-specific models to serial position effects only, we allowed these models to make their own predictions regarding the range of additional effects identified in our study. The models' predictions regarding the effects of modality on probability of first recall and PLI behavior appear in Figure 8. CMR2 predicted that auditory presentation would reduce the probability of initiating recall from Serial Position 1 and increase the probability of initiating recall from the final few items of the list, though it overestimated the magnitude of these effects. In contrast, the model incorrectly predicted that visual presentation would result in more PLIs than auditory presentation, whereas our data show modality having little effect on PLI counts. Furthermore, it

predicted stronger lag-1 PLI recency with auditory presentation than with visual presentation—the opposite of our empirical findings. We explain the reasons for this conflicting behavior in our Modeling Discussion below.

Simulation 3

Table 4 lists the parameter sets that best fit the pooled-modality and modality-specific data from six-item immediate free recall trials in Experiment 1 of Grenfell-Essam et al. (2017) and from 24-item trials in Experiment 2 of the present study. Consistent with Simulation 2, faster contextual drift rates during encoding and stronger context-to-item associations best captured the difference between auditory and visual free recall performance, regardless of list length. Figure 9 and Figure 10 compare the serial position curves generated by our modality-specific models to those observed empirically.

Replicating the results of Simulation 2, our models of six-item and 24-item free recall both demonstrated auditory recency advantages regardless of recall start position. Our six-item model consistently showed an auditory advantage for the final three items of the list, whereas our 24-item model showed an auditory advantage for the final three to seven items, depending on start position. When recall began from the final four list items, the six-item model showed an auditory advantage for the first list item, whereas the 24-item model showed a small visual advantage for the first list item. Thus, the model successfully reproduced both the growth in the extent of the modality effect with list length, as well as the emergence of the inverse modality effect as list length increases beyond six (Grenfell-Essam et al., 2017).

Modeling Discussion

We find that the modality effect in free recall can arise as a product of increased contextual drift rates and stronger context-to-item binding under conditions of auditory relative to visual presentation. The CMR2 model (Lohnas et al., 2015) successfully accounted for both the magnitude and extent of the modality effect across list lengths ranging from six to 24. Furthermore, our analyses show that the modality effect can arise from these two parameters regardless of recall order, consistent with our data as well as that of Grenfell-Essam et al. (2017).

We now consider how differences in drift rate and context-to-item associations might interact to produce the effects we observed in our experiments and simulations. Faster contextual drift during encoding reduces the contextual overlap between all memory items, effectively increasing the distances between them. Therefore, all items' contexts lie further from the end-of-list context than they would under conditions of slower contextual drift. This increased distance produces a tilt in the serial position curve, characterized by a steeper recency curve that drops to a lower asymptote and rises to

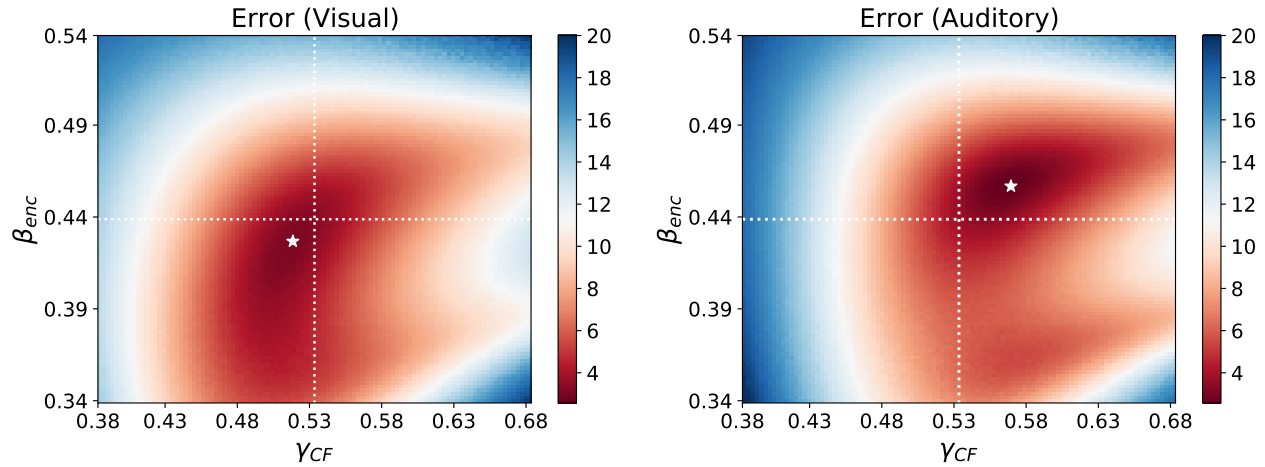


Figure 6

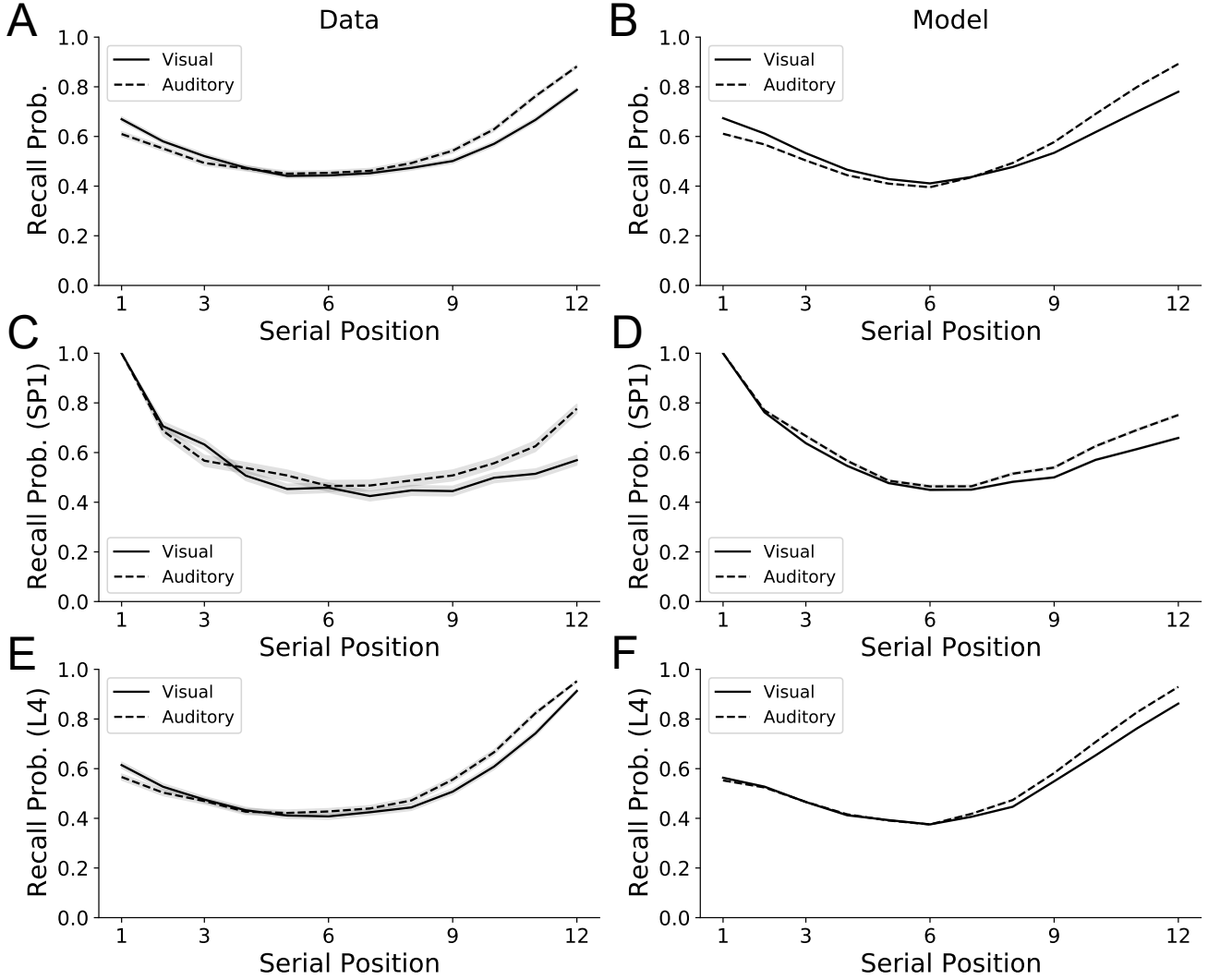
Parameter Space Analysis of the Modality Effect. Heat maps illustrate the square root of the weighted mean-squared-error between simulated and empirical data for each modality, as a function of β_{enc} and γ_{CF} in Simulation 2. Dotted lines indicate the parameter values that produced the best fit to the grand average data, pooled across modalities in Simulation 1. The best-fitting parameter combination for each individual modality is indicated by a star.

a relatively shallow primacy curve. This distance also increases the probability of initiating recall from recent items, relative to the first item of the list. The likelihood of recalling an item first depends largely on the similarity between its associated context and the state of context at the start of the recall period. An increased drift rate reduces the similarity between the end-of-list context and the contexts of earlier items, thereby decreasing their accessibility at the start of the recall period. On its own, then, a faster contextual drift rate can account for the recency advantage associated with auditory presentation.

However, the tilt in the serial position curve caused by increased contextual drift would produce auditory disadvantages for all prerecency items, whereas our empirical data show auditory disadvantages isolated to early-list items only. In our two-parameter account, enriched context-to-item binding for aurally presented items counterbalances the memory impairment caused by an increased drift rate, thereby negating the drop in the asymptote of the serial position curve. Together these factors produce a reliable auditory recency advantage (i.e., the modality effect), coupled with the potential for an auditory primacy disadvantage (i.e., the inverse modality effect) if enriched encoding insufficiently counterbalances the negative impacts of faster contextual drift. Across mid-serial items, the two effects exist in approximate equilibrium, and neither modality shows a consistent advantage.

CMR2's predictions regarding prior-list intrusions (PLIs) challenge our account of modality effects. Although we initially set out to develop an account of the classic modality

effect, our experiments uncovered novel effects of modality on intrusion behavior. Our simulations contradict two of our findings, as CMR2 predicted that auditory presentation would produce fewer PLIs and a steeper PLI recency curve relative to visual presentation. These predictions conflict with our empirical observation that modality does not affect the quantity of PLIs, as well as our observation that auditory presentation decreases PLI recency. To better understand the effects of β_{enc} and γ_{CF} on intrusions, we examined the topography of intrusion rates across our grid search from Simulation 2 (see Figure S1 in Supplemental Materials). The distribution of intrusion rates suggests that the quantity of PLIs is highly sensitive to γ_{CF} , yet relatively insensitive to β_{enc} . Strengthened context-to-item binding reliably decreased the prevalence of PLIs, resulting in our auditory model producing fewer intrusions than our visual model. Increasing γ_{CF} amplifies the importance of temporal context (e.g., relative to semantic associations) in guiding memory retrieval. We suggest that this effect improves a person's ability to differentiate words presented in the current list context from those presented earlier in the session, reducing the probability of making PLIs. Furthermore, intrusions that do occur will be especially likely to come from the most temporally similar list, increasing PLI recency. Indeed, the topography of PLI recency scores across our grid search suggests that PLI recency tends to increase dramatically with increased γ_{CF} ; in contrast, it tends to decrease slightly with increases in β_{enc} , as we hypothesized in our discussion of our empirical results. However, because γ_{CF} produces a larger effect than does β_{enc} , our model shows a net increase in PLI recency during audi-

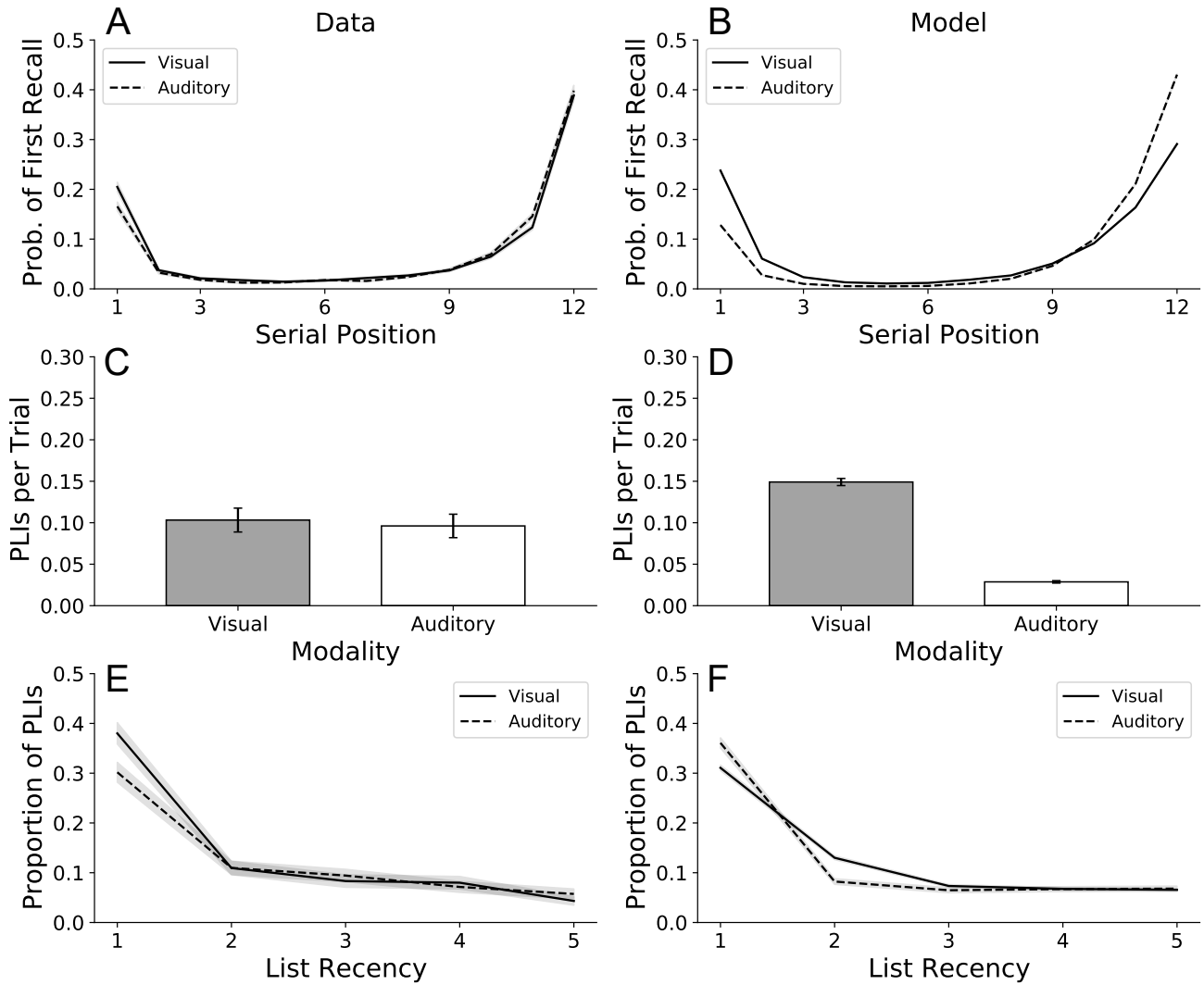
**Figure 7**

Simulated Modality Effects. Recall performance by modality as observed in Experiment 2 and as simulated by CMR2 in Simulation 2. Performance is shown across all trials (**A** and **B**), across trials where the first list item was recalled first (**C** and **D**), and across trials where one of the last four items was recalled first (**E** and **F**). Shaded regions indicate one standard error of the mean. The best fit was determined by the weighted mean squared error between model and data for the start-conditional serial position curves only.

tory presentation (as compared to visual) whereas our data show a decrease in PLI recency.

Given that increases in β_{enc} enhanced recall for recent items, reduced recall for early-list items, increased the probability of initiating recall from the final items of a list, and reduced PLI recency without substantially altering the total number of PLIs made, our simulations strongly support the hypothesis that differences in contextual drift rates underpin the variety of modality-driven effects in our data. In contrast, we found only mixed support for the hypothesis that auditory

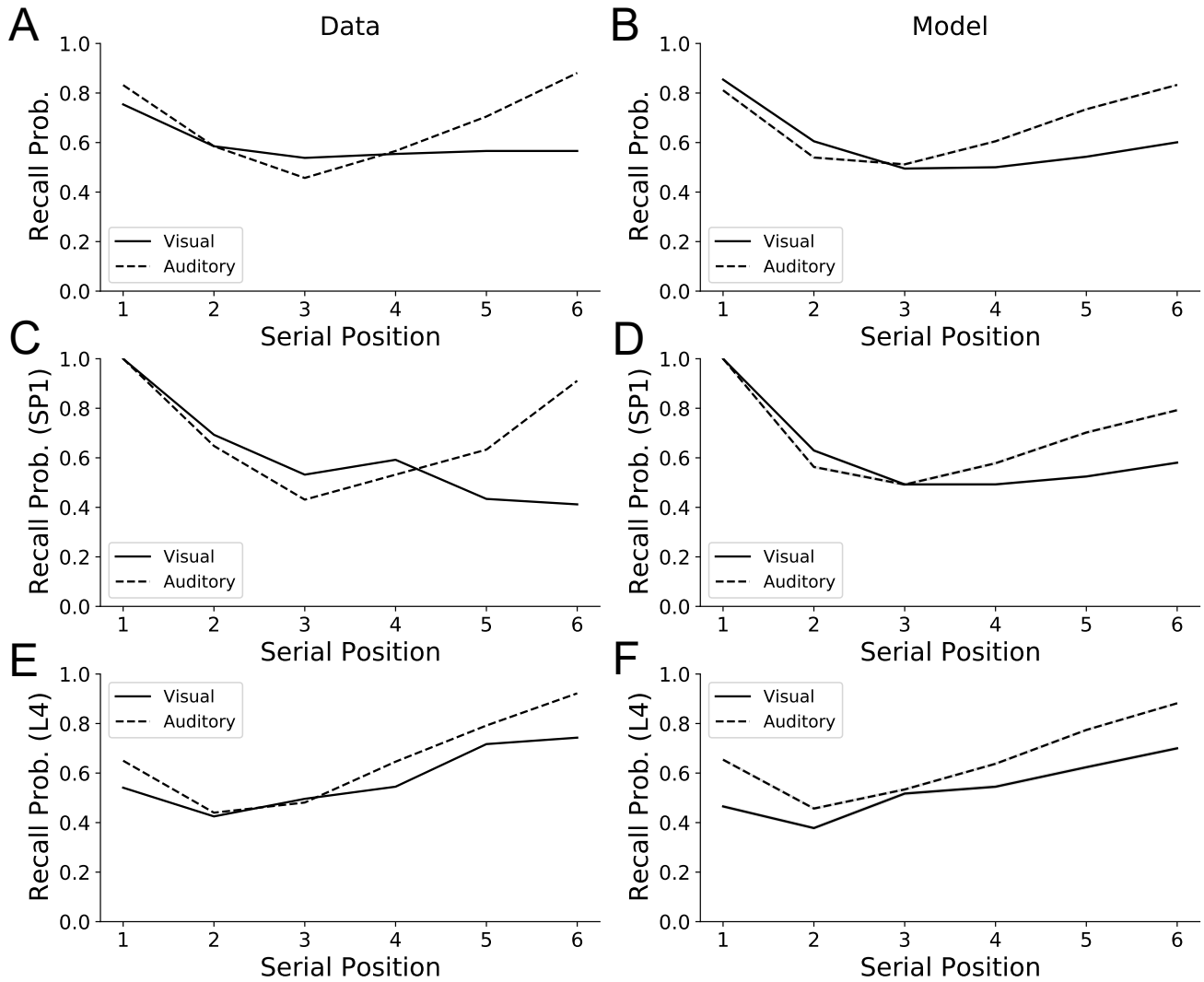
presentation facilitates enhanced encoding relative to visual presentation. Although such an effect would help to counterbalance the prerecency disadvantage conferred by faster contextual drift, it would also produce differences in prior-list intrusion behavior that contradict those we empirically observed. Future work should assess whether β_{enc} may combine with a parameter other than γ_{CF} to better account for intrusion behavior across modalities. For example, modalities might conceivably differ in ϕ_s or ϕ_d depending on their capacity to capture attention across a trial.

**Figure 8**

Additional Predicted Effects of Modality. Effects of modality on recall initiation (**A** and **B**), prior-list intrusions (**C** and **D**), and PLI recency (**E** and **F**) as observed in Experiment 2 and as predicted by Simulation 2. Shaded regions indicate one standard error of the mean and error bars indicate 95% confidence intervals.

To explain why contextual drift rates and association strengths might differ by presentation mode, we consider the nature of the stimuli themselves. The modality effect literature frequently overlooks a potential confound of modality, in that auditory presentation differs from visual presentation not only in the sensory mode of the stimuli, but also in its temporal dynamics. Under traditional serial and free recall paradigms, the physical features of visual items remain static for the duration of the presentation interval, whereas 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

**Figure 9**

Simulated Modality Effects in Short Lists. Recall performance by modality is illustrated for 6-item lists as observed by Grenfell-Essam et al. (2017) and as simulated by CMR2 in Simulation 3. Performance is shown across all trials (**A** and **B**), across trials where the first list item was recalled first (**C** and **D**), and across trials where one of the last four items was recalled first (**E** and **F**). Shaded regions indicate one standard error of the mean. The best fit was determined by the mean squared error between model and data for the start-conditional serial position curves only.

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 propose a retrieved-context account of modality effects in free recall. Data from approximately three thousand participants allowed us to precisely quantify the effects of modality on

free recall under conditions of varying list length and presentation rate. These data not only replicated recent and classic findings, but also identified new effects such as participants' increased tendency to commit lag-1 prior-list intrusions during visual lists, relative to auditory. We fit the Context Maintenance and Retrieval model (Lohnas et al., 2015) to our empirical data, accounting for the classic modality effect across list lengths as a product of differential contextual drift rates (faster drift for auditory items) and context-to-item associative encoding (stronger for auditory items).

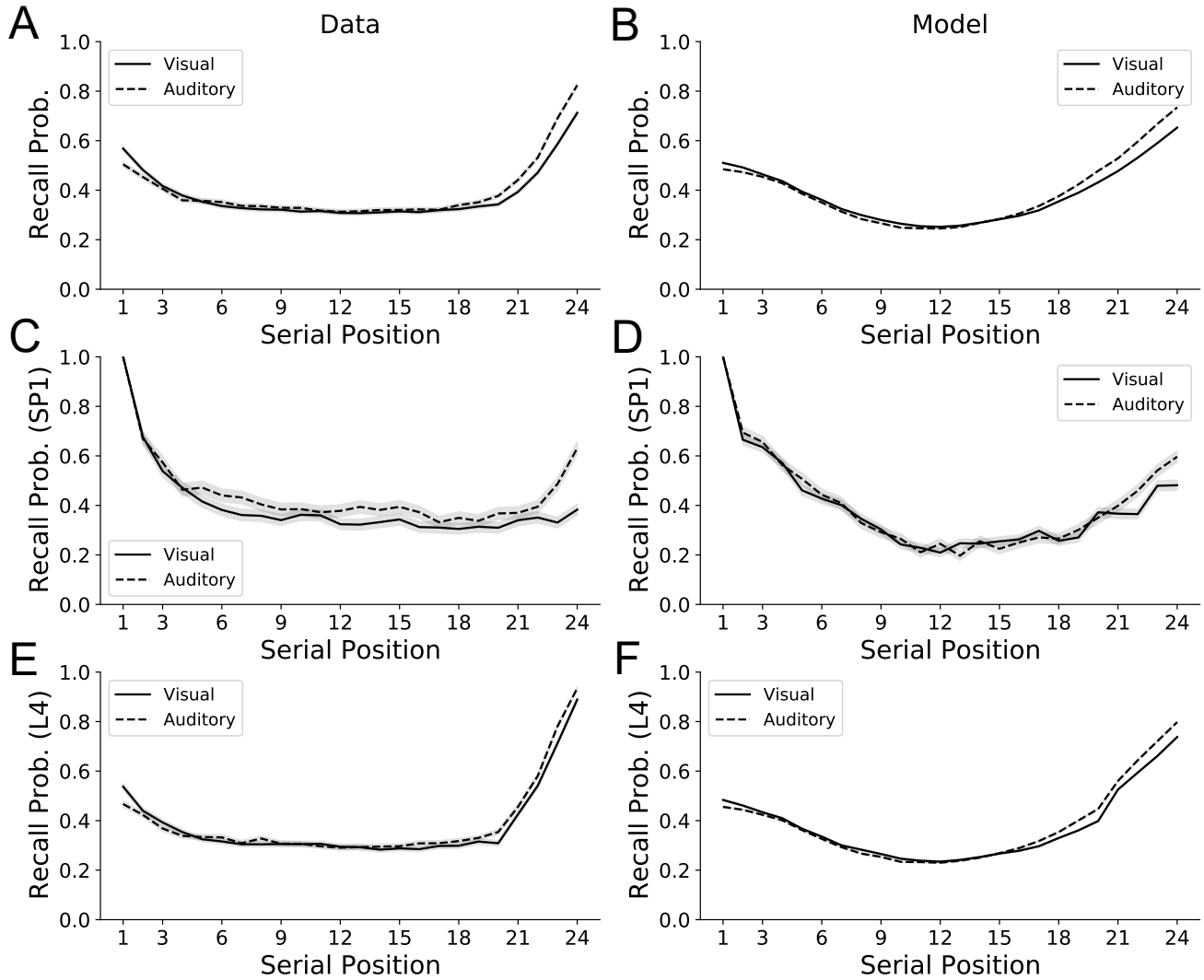


Figure 10

Simulated Modality Effects in Long Lists. Recall performance by modality is illustrated for 24-item lists as observed in Experiment 2 and as simulated by CMR2 in Simulation 3. Performance is shown across all trials (A and B), across trials where the first list item was recalled first (C and D), and across trials where one of the last four items was recalled first (E and F). Shaded regions indicate one standard error of the mean. The best fit was determined by the weighted mean squared error between model and data for the start-conditional serial position curves only.

Across 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). Although we found that participants were more likely to initiate recall from end-of-list items on auditory trials than on visual trials, visual primacy and auditory recency advantages persisted regardless of whether recall began from the beginning or end of the list (see Figure 3). Our findings suggest that

presentation modality does affect the relative accessibility of early and recent items at the start of recall, but support previous evidence that the modality and inverse modality effects do not originate from differences in recall initiation (Cowan et al., 2002; Craik, 1969; Grenfell-Essam et al., 2017; Harvey & Beaman, 2007). Our findings do not, however, support previous suggestions that faster presentation rates strengthen modality effects and attenuate inverse modality effects (e.g., Macken et al., 2016; Murdock & Walker, 1969; Murray, 1966; Roberts, 1972). We observed similarly large modal-

ity and inverse modality effects across presentation rates and across list lengths of 12 and 24. Our list length results augment those of Grenfell-Essam et al. (2017) and Roberts (1972), and suggest that modality effects may extend with list length across shorter lists, but remain stable across list lengths longer than 12.

In addition to clarifying and augmenting past findings, we observed in both experiments a novel effect of modality on the recency of PLIs (Figure 4), such that visual presentation elicited more intrusions from the immediately preceding list 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 PLIs; 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) in cued recall. That modality affects prior-list intrusion behavior suggests that modality exerts a longer-lasting impact on memory than theories such as PAS (Crowder & Morton, 1969) and the feature model (Nairne, 1990) have often assumed.

To elucidate the mechanistic underpinnings of modality effects in free recall, we developed an extension of the CMR2 model (Lohnas et al., 2015; Healey & Kahana, 2016) 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 effect across a wide range of list lengths as a function of these two parameters. According to our model, the combination of faster contextual drift and stronger context-to-item binding biases recall towards recent items while boosting recall across all serial positions, the net effect of which is an auditory recency advantage and potential visual primacy advantage. 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 accounts for the persistence of the modality effect across different recall start positions (e.g., Grenfell-Essam et al., 2017). 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 dynamic visual presentation modes (e.g., sign language, finger spelling, and lip reading) can produce similar 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., 2009a). 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. Finally, whereas our findings regarding the rates and recency of intrusions support the hypothesis that auditory presentation produces faster contextual drift than does visual presentation, they conflict with the hypothesis of strengthened auditory context-to-item binding. Future modeling work should seek to identify whether an alternative parameter to γ_{CF} may better account for our novel results.

Our retrieved-context account also raises new questions for behavioral research. To explain why our model parameters might vary by modality, we suggested that changes in the features of a stimulus across 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 contextual drift and stronger contextual coding, 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. Perhaps the modality effect derives from temporal dynamics rather than sensory mode.

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 work shows how modality effects can arise without hypothesizing distinct memory stores responsible for retrieval of recent and remote information.

References

- Aguinis, H., Villamor, I., & Ramani, R. S. (2020). MTurk research: Review and recommendations. *Journal of Management*, 47(4), 823–837. doi: 10.1177/0149206320969787
- Aka, A., Phan, T. D., & Kahana, M. J. (2021). Predicting recall of words and lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(5), 765–784. doi: 10.1037/xlm0000964
- 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
- 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., Sauls, 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., Sauls, 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
- Dennis, S. A., Goodson, B. M., & Pearson, C. (2019). *Virtual private servers and the limitations of IP-based screening procedures: Lessons from the MTurk quality crisis of 2018*. (Working Paper) doi: 10.2139/ssrn.3233954
- 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
- Gardiner, J. M. (1983). On recency and echoic mem-

- ory. *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). psi-Turk: 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
- 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., 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, R., Clifford, S., Burleigh, T., Jewell, R., & Waggoner, P. (2018). *The shape of and solutions to the MTurk quality crisis*. (Working Paper) doi: 10.2139/ssrn.3272468
- 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
- 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: Modeling 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
- Long, N. M., & Kahana, M. J. (2017). Modulation of task demands suggests that semantic processing interferes with the formation of episodic associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(2), 167-176. doi: 10.1037/xlm0000300
- 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. *Cog-*

- tion, 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. (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., & 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 Experimental 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. (2009a). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116(1), 129-156. doi: 10.1037/a0014420
- Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009b). Task context and organization in free recall. *Neuropsychologia*, 47, 2158-2163. doi: 10.1016/j.neuropsychologia.2009.02.013
- 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.
- 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 proac-

tive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, 104(4), 442-452. doi: 10.1037/0278-7393.14.442

- 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

Appendix A Participant Exclusion

We excluded participants in both experiments based on the following criteria, which were designed to filter out individuals who failed to stay on task or who may have cheated:

1. Making no correct recalls on two or more trials.
2. Correctly recalling more than 95% of the words in the study.
3. Averaging fewer than three correct distractor problems per trial.
4. Demonstrating extremely low (z -score < -2) overall accuracy on the distractor task.
5. Indicating in the post-experiment questionnaire that they had written or typed notes as a memory strategy.

In Experiment 1, we excluded a total of 201 participants for violating at least one of these criteria. Ten participants (5 unique) violated the first, 2 (both unique) violated the second, 42 (24 unique) violated the third, 40 (21 unique) violated the fourth, and 132 (126 unique) violated the fifth. In Experiment 2, we excluded 531 participants. Thirteen (5 unique) violated the first criteria, 12 (5 unique) violated the second, 97 (42 unique) violated the third, 54 (12 unique) violated the fourth, and 430 (387 unique) violated the fifth.

Recent studies have raised concerns regarding online participants' usage of virtual private servers (VPSs) to circumvent MTurk's IP screening methods, and have associated VPS usage with poor data quality (Dennis, Goodson, & Pearson, 2019; Kennedy, Clifford, Burleigh, Jewell, & Waggoner, 2018). In light of these concerns, we tested for VPS

usage among our participants, using the tools available at <https://ipinfo.io> to identify the Internet service provider associated with each respondent's IP address. We found that only a small pool of participants used VPS services to access our experiment (including, but not limited to those identified by Dennis et al., 2019). In total, only 34 participants in Experiment 1 (3.1% of respondents) and 33 participants in Experiment 2 (1.7% of respondents) were identified as VPS users. Given these low usage rates, we filtered participants only based on their task performance, as described above, and did not exclude participants based solely on their use of a virtual private server.

Appendix B List Construction

Although we did not examine semantic similarity effects in these studies, we designed our lists following the methodology of Long and Kahana (2017) and Aka et al. (2021) to more evenly distribute semantically similar items across adjacent and non-adjacent list positions. We used Google's Word2vec model (Mikolov et al., 2013) to estimate semantic similarities among all of the words in our word pool. We then used these (cosine) similarity scores to group word pairs into low ($\cos \theta \leq 0.188$), medium ($0.188 < \cos \theta < 0.515$), and high similarity ($\cos \theta \geq 0.515$) bins. 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 C Effects of Manipulating Distractor Duration

Although not the primary focus of our report, we included an arithmetic distractor task between the end of each recall interval and the presentation of the subsequent list. We manipulated the duration of this distractor task within subjects, with half of the lists being preceded by a 12-second distractor period and the other half preceded by a 24-second distractor period. As one might expect intertrial distractor tasks to affect the tendency to make prior-list intrusions (PLIs), we evaluated the effects of distractor duration on the average number of prior-list intrusion errors per trial and PLI recency for a list lag of 1 via a series of paired sample t -tests. These tests revealed no significant effects of distractor duration on the number of PLIs per trial, nor PLI recency in either experiment (all $t < 1$).