Give me Enough Time to Rehearse: Presentation Rate Modulates the Production Effect

Ian Dauphinee[1], Mathis Roy[1], Dominic Guitard[2], James M. Yearsley[3], Marie Poirier[3], & Jean Saint-Aubin[1]

[1] Université de Moncton, Moncton, NB, Canada
[2] Cardiff University, Cardiff, Wales, United Kingdom
[3] City, University of London, United Kingdom

Authors' Note

This research was supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada to JSA. While working on this manuscript, ID was supported by an undergraduate student research award from NSERC and DG was supported by a post-doctoral fellowship from NSERC.

Correspondence concerning this article should be addressed to Jean Saint-Aubin, École de psychologie, Université de Moncton, Moncton, New Brunswick, E1A 3E9, Canada. Email: jean.saint-aubin@umoncton.ca

Open Practice Statement

The data, stimuli and model codes are available on the Open Science Framework project page https://osf.io/37dgt/?view_only=4a3dee74db5f49d7a0b74a1e79bf1ba0 [Note to the reviewers: This page will be made public once the manuscript has been accepted.]

Abstract

This paper uses the production effect to test one of the important predictions of a view of memory that is embodied in the Revised Feature Model (RFM). When to-be-recalled lists contain items both read aloud and silently, words read aloud are less well recalled at the beginning of the list and better recalled at the end. According to the RFM, producing the items by reading them aloud adds distinctive features which supports recall, but production also interferes with rehearsal – a process that operates more significantly at the start of a list. This critical role assigned to rehearsal has never been systematically tested. We do this here through a systematic literature review and an experiment that manipulates presentation rate. With a faster presentation rate, rehearsal is less likely; the implication is that the advantage observed for silently read items in the primacy positions should vanish, while the recency advantage for produced items should remain. The systematic review collected an initial sample of 422 unique articles on the production effect in immediate serial recall and revealed the predicted pattern. In addition, in our experiment, the presentation rate was manipulated within an immediate serial recall task (500, 1000, and 2000 ms/word). As predicted, the recency advantage for produced items was observed for all presentation speeds. Critically, the production disadvantage for early serial positions was only present for the two slowest rates, but not at the fastest speed. Results were successfully modeled by calling upon the RFM.

Keywords. Production Effect, Serial Position, Presentation Rate, Short-Term Memory, Immediate Serial Recall, Revised Feature Model.

Give me Enough Time to Rehearse: Presentation Rate Modulates the Production Effect

In an ideal world, memory strategies would reliably improve remembering. However, beneficial strategies are typically associated with extra processing or other resource costs. These opposing trends can generate confusing patterns of results with strategies either generating a benefit, a cost or no effect (see, e.g., MacLeod & Bodner, 2017; Mulligan & Lozito, 2004; Serra & Nairne, 1993). The well-established production effect is no exception; in a recall task, when participants are asked to produce items within a list by reading them aloud, relative to silently, overall, there is no benefit of this encoding strategy (Fawcett et al., 2023). However, production interacts with serial position: producing the items hinders recall for early serial positions while improving it for later ones—exhibiting another complex pattern of costs and benefits (see, e.g., Gionet et al., 2022).

Although frustrating in some regards, the complexities outlined above involve two positive aspects. First, they can be seen as illustrating important principles in learning and memory such as elaborative processing and the cost-benefit realities related to encoding (Mulligan & Lozito, 2004). Second, they provide a crucible for testing theories and models of memory (Mulligan et al., 2019).

Here we use the intricacies of the production effect to test a view of memory embodied in a recent recall model: the Revised Feature Model (RFM; Saint-Aubin et al., 2021, 2023). According to the RFM, elaborative encoding is useful if it generates distinctive features—i.e. features that support retrieval because they are unique to the items being memorized. More specifically, the RFM suggests that recall is determined by the *relative* distinctiveness of the link between the retrieval cue and memoranda. If a cue is more similar to the correct retrieval candidate than to competitors, then the probability of successful retrieval is heightened.

However, this beneficial distinctiveness can come at a cost if it disrupts other processes – namely, in this case, rehearsal.

In the RFM, the cost of production is modelled by assuming that reading aloud interferes with subvocal rehearsal (Murray, 1967). Saint-Aubin et al. (2021) implemented this idea by adding a rehearsal process to the model (see also Nairne, 1988, 1990; Neath & Nairne, 1995; Neath & Surprenant, 2007). Moreover, within the RFM, the representations of produced items have more features—generated by articulation and auditory feedback—than silently read items. These additional features increase distinctiveness and the probability of successful retrieval. However, procuring these additional features comes at a cost, as production interferes with rehearsal. Because the first list items are rehearsed more often than later ones, production has more impact on the primacy section of the serial position curve (Bhatarah et al., 2009; Rundus, 1971; Tan & Ward, 2008).

Despite the key role assigned to rehearsal in explaining the interaction between production and serial position, there is only one relevant experimental demonstration. In their fifth experiment, Saint-Aubin et al. (2021) asked participants in the control condition, involving silent reading, to say an irrelevant word aloud after each item was presented. This was an attempt to better equate rehearsal opportunities for the silent control relative to the production condition. They observed a clear benefit of production across all serial positions. However, saying a long irrelevant word aloud probably disrupts rehearsal more than reading to-be-remembered items aloud. Accordingly, they modelled the results by reducing the efficiency of the rehearsal process in the control condition compared to the production condition.

In a nutshell, more evidence is required to establish whether, as predicted by the RFM, the detrimental effect of production on primacy positions can be abolished by better equating rehearsal opportunities across conditions. Presentation rate is known to have a major impact on rehearsal frequency (e.g., Bhatarah et al., 2009; Camos et al., 2019; Glanzer & Cunitz, 1966; Tan & Ward, 2000). For instance, in an immediate serial recall task, Tan and Ward (2008) manipulated presentation rate while asking their participants to rehearse aloud. Results showed that participants rehearsed at slow, but not at fast presentation rates. In fact, fast presentation has often been used to prevent rehearsal (e.g., Landry et al., 2022; Macken et al., 2016). In this context, Macken's et al. study is of particular interest. They noted that compared to visually presented items, aurally presented items are better recalled for the recency portion of the curve—the well-known modality effect—but sometimes less well recalled on the pre-recency portion of the curve, an effect they labelled the inverse modality effect. Akin to our reasoning with the production effect, they hypothesized that the inverse modality effect was due to greater rehearsal opportunities for visually presented items. In their third experiment, they manipulated presentation speed to impede rehearsal. As hypothesised, a rapid presentation rate eliminated the inverse modality effect.

Study 1: Systematic review of the literature

To further assess this rehearsal hypothesis, we systematically reviewed the literature on production in immediate serial recall. We specifically targeted articles from which serial position curves could be extracted.

Method

We conducted our literature search on June 2nd, 2023, on the PsycINFO and Scopus databases, for which the following search terms were used ("Modality effect" OR "Vocalization effect" OR "Vocalization" OR "Vocalisation" OR "Production effect" OR "Reading aloud") AND ("Order Reconstruction" OR "Immediate Recall" OR "Serial Recall" OR "Verbal memory" OR "Serial learning" OR "Short Term Memory"). In addition, to be considered for the review, the article, conference paper, or book section had to: (a) report an empirical study, (b) use human participants, (c) avoid including a clinical population, (d) compare recall performance on words read aloud and silently, (e) include a serial recall task, and (f) present a serial position curve or a table.

The systematic review was conducted via the Covidence systematic review software (Veritas Health Innovation, Melbourne, Australia), and the flow diagram depicting our search and screening procedure is presented in Figure 1. After removing duplicates, from a total of 573 studies, 452 were retained for the primary screening phase. Of the former, 307 were obtained on Scopus and 264 on PsycINFO. Two additional studies were identified based on their mention by Saint-Aubin et al. (2021). For the primary screening, we examined if these studies met our inclusion criteria based on the title and abstract. After this procedure, 348 studies were removed, leaving 104 articles to be assessed by full-text review during the secondary screening phase. After the full-text review, of the 104 articles remaining, 88 were excluded, while the remaining 16 met all inclusion criteria. The excluded articles comprised 37 articles that did not compare performance for items read aloud and read silently, 36 studies in which there was no immediate serial recall task, 5 articles did not include a serial position curve or table from which data could be extracted, 5 articles with no empirical study, 2 articles for which it was impossible to retrieve the full text, 2 articles with an inadequate study design for the current purpose, and 1 only included a clinical population. From the 16 studies published between 1968 and 2021, we extracted 21 data sets using WebPlot Digitizer, version 4.5 (Rohatgi, 2020). Overall, the presentation rates range from 420 ms per word to 2000 ms per word. Given the distribution of presentation rates, we grouped the studies in three categories. The rapid presentation rate category included studies where items were presented at 500 ms per item or less. When the presented rate was between 500 ms and 1000 ms per item, they were placed in the intermediate presentation rate category. The slow category included studies with a presentation rate of at least 2000 ms per item. We then combined the 21 data sets into ten figures according to their list length and presentation rate category**.**

Results and Discussion

As shown in Figure 2, irrespective of presentation rate, produced items are better recalled than silently read items for the last positions. Importantly, with the long and the intermediate presentation rates, produced items were less well recalled for the first serial positions, but this disadvantage was not observed with the fastest presentation rate. Therefore, the systematic review supports the rehearsal hypothesis imbedded in the RFM (Saint-Aubin et al., 2021, 2023).

Study 2

The systematic review clearly showed a relationship between presentation rate and the production disadvantage for early serial positions. However, this is a correlational finding relying on comparisons across studies; no previous study has directly compared the impact of different presentation rates. This situation is suboptimal because the method used across studies varied on elements beyond presentation rate and some of them may account for the observed effects. Therefore, we decided to test the production effect with three different presentation rates.

Method

Sample size calculation. We used G*Power 3.1.9.6 (Faul et al., 2007) and the results of Experiment 1A of Saint-Aubin et al. (2021) to select our sample size. More specifically, we used the effect size for the critical interaction between presentation modality (aloud vs. silent) and serial position (1 to 6) observed in their analysis of the pure lists condition (η_p^2 = .43). With that information, an a priori interaction between the two repeated measure factors was computed with α = .05, 1 – β = .95, and the default parameters were used for the correlation among the repeated measures and the nonsphericity correction. The results from the analysis revealed that only 4 participants would be needed for our design. However, the suggested sample size being too small to achieve reliable estimates, we decided to be cautious and used 24 participants per presentation speed as used by Saint-Aubin et al. (2021) in their study in which there was a single presentation speed (2000 ms / word). We therefore overpowered our design and calculated a sensitivity analysis which revealed that a sample of 24 participants with $\alpha = .05$, $1 - \beta = .95$, and the default parameters would allow us to detect a medium effect size (Cohen's $f = 0.27$).

Participants. Seventy-two students (42 women, 29 men, 1 "other", $M_{age} = 20.85$, $SD = 2.85$) from Université de Moncton took part in this experiment and received course credits or were entered in a draw of $100. To take part in the experiment, participants had to be native French speakers and have normal or corrected to normal vision. Participants were randomly and evenly assigned to one of three presentation rate groups (500 ms, 1000 ms or 2000 ms). One participant was removed and replaced as her data could not be retrieved. Free and informed consent was given by all participants prior to the study, which was approved by the research ethics board of Université de Moncton.

Materials. The stimuli were 432 French words, selected from the Lexique 3 database (New et al., 2004). The words were all nouns comprised of a single phonological and orthographic syllable, with frequency ranging from 0 to 1289.39 occurrences per million ($M = 66.13$, $SD = 132.96$). For each participant, 72 lists, each containing six words, were generated by drawing without replacement from this pool. Therefore, each participant was presented with 72 different lists of words.

Design. A 3 × 2 × 6 mixed design was used with presentation rate (500 ms, 1000 ms or 2000 ms per word) as the between-participants factor, and presentation modality (silent vs.

aloud) and serial position (1–6) as repeated factors. The experiment included six practice trials, which were followed by 66 experimental trials.

Procedure. Participants were tested remotely in a single approximately 45-minute online session that took place on their personal computer. To ensure compliance with the instructions, participants had to turn on their camera and microphone and share their screen via Microsoft Teams or Zoom with the experimenter, who was present throughout the session. The experiment was controlled with PsyToolkit (Stoet, 2010, 2017), and the stimuli were presented against a white background on a computer screen in lowercase 20-point Times New Roman font. The experiment was self-controlled by the participants, who pressed the space bar key of the keyboard to initiate each trial. At the start of the experiment, instructions were presented on the screen. In both presentation modality conditions (aloud and silent), the six to-be-remembered words were sequentially displayed in the centre of a computer screen immediately after the start of a trial. The words were presented at a rate of 2 words per second (500 ms on, 0 ms off), 1 word per second (1000 ms on, 0 ms off) or 1 word per 2 seconds (2000 ms on, 0 ms off). The presentation modality condition varied randomly from trial to trial. Participants were instructed that lists of words presented in blue had to be read aloud, whereas those in black had to be read silently, without moving their lips or whispering the words. Participants were told to memorize, and later recall, as many words as possible in the order in which they were presented. Three question marks were presented at the top of the screen, following the presentation of the last word, to serve as an indication of the recall period. Participants recalled the words by typing them with the keyboard and were told that they would not be penalized for spelling errors. The enter key had to be pressed after each word to register their answer, which remained on the screen once typed, and backtracking was not permitted. Furthermore, if they forgot a word at a given serial position, they were told to leave the space blank.

Data analysis. Before conducting any statistical analyses, we first checked participants' responses for misspellings, which were then corrected with the proper word if they could be unambiguously identified (e.g., letter substitutions: telegrem instead of telegram; letter repetitions: telegraam instead of telegram; or letter omissions: telegrm instead of telegram). Overall, spelling was corrected for 609 words (3.91% of trials) in the silent condition and for 649 words (4.17% of trials) in the production condition. Both raw and corrected data are available on OSF, but only corrected data were analysed. Participants' responses were then scored using a strict recall criterion whereby a word is considered correct only if it is recalled at its original presentation position.

To guide our statistical inferences, we used Bayes factor (*BF*) ANOVA analyses computed with the "BayesFactor" package in R with the default parameters (Version 0.9.12–4.4; see Morey et al., 2022; Rouder et al., 2012). For all *BF*s, which were estimated using 100,000 iterations via Monte Carlo simulations, the proportional error was below 5%. In our *BF* ANOVAs, participants were entered as a random factor, while main effects and interaction models were tested by successively omitting these effects from the full model. In the results section, we use the following nomenclature in which *BF*s with values representing evidence in favour of an effect are denoted by BF_{10} and values representing evidence against an effect are denoted by BF_{01} ($1/BF_{10}$). We interpret a value between 3 and 10 as indicating substantial evidence; a value between 10 and 30 as strong evidence; values between 30 and 100 as very strong evidence; and values greater than 100 as decisive evidence (Wetzels et al., 2011). We also computed Bayesian factor paired samples *t*-tests. Finally, we reported the corresponding *F* ratios and partial eta squares as descriptive information for all analyses by use of the “ez” package (Version 4.4–0; Lawrence, 2016).

Results

Figure 3 shows the critical interaction between presentation rate, production and serial position. For the two slowest rates, produced items were better recalled at later positions; the reverse was observed for first positions. As expected, the typical disadvantage for produced words in the primacy positions was not observed with the fastest presentation rate (500 ms/item).

The 3 × 2 × 6 mixed-design ANOVA provided very strong evidence of a main effect of presentation modality, $F(1, 69) = 7.37$, $\eta^2_p = .10$, $BF_{10} > 10{,}000$, and serial position, $F(5, 345) = 82.54$, $\eta^2_p = .54$, $BF_{10} > 10{,}000$. However, the main effect of presentation rate was absent, $F(2, 69) = 3.14$, $\eta^2_p = .08$, $BF_{01} = 3.12$. There was decisive evidence in favor of three two-way interactions between presentation modality and presentation rate, $F(2, 69) = 6.94$, $\eta^2_p = .17$, $BF_{10} > 10{,}000$, serial position and presentation rate, $F(10, 345) = 3.62$, $\eta^2_p = .09$, $BF_{10} > 10{,}000$, and serial position and presentation modality, $F(5, 345) = 57.98$, $\eta^2_p = .46$, $BF_{10} > 10{,}000$. The two-way interaction between presentation modality and presentation rate was decomposed by computing an ANOVA with presentation rate as the only factor for each presentation modality. Results indicated strong evidence in favor of an effect of presentation rate for silent items, $F(2, 69) = 8.12$, $\eta_p^2 = .19$, $BF_{10} = 11.62$, and against an effect for aloud items, $F(2, 69) = 0.82$, $\eta_p^2 = .01$, $BF_{01} = 16.35$. As expected, there was decisive evidence in favor of a three-way interaction between presentation modality, serial position, and presentation rate, $F(10, 345) = 5.05$, $\eta^2_p = .13$, $BF_{10} > 10{,}000$. The three-way interaction was decomposed by computing a 2 × 6 repeated-measures ANOVA for each presentation rate.

500 ms. The ANOVA revealed a main effect of presentation modality, $F(1, 23) = 24.16$, $\eta^2_p = .51$, $BF_{10} > 10{,}000$, of serial position, $F(5, 115) = 34.14$, $\eta^2_p = .60$, $BF_{10} > 10{,}000$, and an interaction between presentation modality and serial position, $F(5, 115) = 8.71$, $\eta^2_p = .27$, $BF_{10} >$ 10,000. Paired sample Bayesian *t*-tests revealed that positions 2 and 3 yielded anecdotal evidence in favor of a superior performance for produced words. As expected, there was strong evidence that words read aloud were better recalled relative to silently read words at Position 5, $BF_{10} =$ 10.89, and decisive evidence at Position 6, $BF_{10} > 10{,}000$. Importantly, at Position 1, there was substantial evidence in favor of an absence of an effect of presentation modality, $BF_{01} = 3.89$ (1/ $BF_{10} = 0.26$). Position 4 yielded anecdotal evidence in favor of an absence of an effect of presentation modality

1000 ms. The ANOVA revealed decisive evidence in favor of a main effect of serial position, $F(5, 115) = 27.24$, $\eta^2_p = .54$, $BF_{10} > 10{,}000$, and of an interaction between presentation modality and serial position, $F(5, 115) = 27.78$, $\eta^2_p = .55$, $BF_{10} > 10{,}000$. However, there was strong evidence for an absence of a main effect of presentation modality, $F < 1$, $\eta^2_p = .00$, $BF_{01} =$ 40.61. Paired sample Bayesian *t*-tests revealed very strong evidence, decisive evidence, and substantial evidence, respectively, that silently read words were better recalled relative to words read aloud at Position 1, $BF_{10} = 64.15$, Position 2, $BF_{10} = 192.98$, and Position 3, $BF_{10} = 3.14$. Position 4 yielded substantial evidence in favor of an absence of an effect of presentation modality, $BF_{01} = 3.53$ (1/ $BF_{10} = 0.28$). Position 5 yielded very strong evidence that words read aloud were better recalled relative to silently read words, $BF_{10} = 83.42$, and Position 6 yielded decisive evidence, $BF_{10} > 10{,}000$.

2000 ms. As observed for the 1000 ms presentation duration, there was decisive evidence in favor of a main effect of serial position, $F(5, 115) = 25.45$, $\eta^2_p = .53$, $BF_{10} > 10{,}000$, and an interaction between serial position and presentation modality, $F(5, 115) = 29.15$, $\eta^2_p = .56$, $BF_{10} > 10{,}000$, but strong evidence against the effect of presentation modality, $F < 1$, $\eta^2_p = .00$, $BF_{01} = 42.78$. Bayesian *t*-tests revealed strong evidence at Position 1, $BF_{10} = 12.64$, very strong evidence at Position 2, $BF_{10} = 72.39$, and strong evidence at Position 3, $BF_{10} = 28.95$ that silently read words were better recalled relative to words read aloud. Further, there was substantial evidence in favor of an absence of an effect of presentation modality at Position 4, $BF_{01} = 4.30$ (1/ $BF_{10} = 0.23$). There was substantial evidence that words read aloud were better recalled relative to silently read words at Position 5, $BF_{10} = 83.42$, and decisive evidence at Position 6, $BF_{10} = 7340.84$.

Discussion

Results are straightforward: as in the systematic review, produced items are less well recalled than silent items in the first serial positions only when the presentation rate provides enough time for rehearsal. Results of the current study fit well with those of Macken et al. (2016) who manipulated presentation rate in the context of the modality effect. Furthermore, as observed in previous studies manipulating presentation rate with silent items, recall performance improved with slower presentation rates (see, e.g., Bhatarah et al., 2009; Oberauer, 2022; Tan & Ward, 2008). For produced items, the null effect of presentation rate on recall is consistent with prior results covered in the systematic review. For instance, as shown in Figure 2, with 6-word lists, on average, mean recall increased from 0.27 with fast presentation to 0.40 with slow presentation for silent items, while performance remained the same for produced items with a performance of 0.42 and 0.41, respectively.

The Revised Feature Model

Although our results align well with the expectations derived from the RFM, demonstrating good quantitative fits is also required. We therefore fitted the RFM to the data from Study 2. The RFM has previously been used to model immediate serial recall for lists of silently read and spoken aloud words, and no modification to the model as described in Cyr et al., (2022) and Saint-Aubin et al. (2021) was required. A full description of the model is available on OSF (https://osf.io/37dgt/?view_only=4a3dee74db5f49d7a0b74a1e79bf1ba0).

The RFM states that items possess modality-independent and modality-dependent features. Internal categorization and identification generate the modality-independent features, while the physical presentation, such as the color of the item, produce the modality-dependent features. Item presentation creates equivalent traces in primary and secondary memory. Traces in secondary memory are assumed to remain unchanged. In primary memory, vectors of features deteriorate through similarity-based retroactive interference, meaning that later items may overwrite earlier ones. At the end of presentation, a final overwriting of modality-independent features results from internal thought activity. Degraded traces in primary memory can be restored through rehearsal with some probability. As shown by Rundus (1971) and Bhatarah et al. (2009), items at initial serial positions are rehearsed more frequently than the last ones. Within the RFM, the efficiency of rehearsal decreases across serial positions, in particular after the first few items. At the point of retrieval, the degraded primary memory traces are used as cues to probe secondary memory.

The production effect is modelled by assuming that produced items possess more modality-dependent features than silently read items, resulting in better recall for the produced items for the final serial positions. This is mainly because the final overwriting of modality-independent features cannot overwrite the extra modality-dependent features of the last items within a list. However, reading items aloud is assumed to interfere with rehearsal. As a result, for the first items of the list, where rehearsal is more likely, produced items are at a particular disadvantage. With a rapid presentation rate, rehearsal opportunities are much reduced. Therefore, any differences between produced and silent items will be minimized. This explains the three-way interaction between production effect, serial position, and presentation rate.

The RFM contains parameters that can vary in the fitting process. However, our aim is to test the specific prediction that differences in performance between the 500ms, 1000ms, and 2000ms conditions can be explained through the varying effectiveness of rehearsal. Hence, we only allowed the rehearsal parameter to vary. For silent conditions, rehearsal was a free parameter. However, to simplify, we required that rehearsal be identical across presentation rates for the aloud conditions, implying that no more rehearsal is attempted at 2000ms than at 500ms when producing items. This means all parameters in the model are equal for the aloud conditions, which would imply, except for noise, that the serial position curves ought to be identical. Clearly this is not exactly the case in the data, however, the RFM should still be able to capture data patterns by assuming rehearsal is the main difference between presentation conditions (see Appendix A for details).

Overall, the fits are indeed good. The presence or absence of a cross-over between silent and aloud conditions are reproduced well (Figure 4). The quantitative fits for the silent conditions are also good, although, as explained above, the fits to the aloud conditions are less accurate since the parameters are not allowed to vary while the data do clearly show some variation (Figure 5).

We can also ask about the values of the best fitting parameters, specifically the rehearsal rate that we allowed to vary between conditions, and which we propose is responsible for the patterns seen in the data. We plot means of the posteriors and 90% Highest Density Intervals in Figure 6. There is a clear pattern of increasing rehearsal strength in the silent conditions as the presentation time increases from 500 to 2000ms, supporting our hypothesis. The best fitting value for the aloud conditions was close to the value for the silent 500ms condition.

General Discussion

The intricacies of the production effect, within both short- and long-term memory paradigms, have been explained and modelled by calling upon the RFM (Cyr et al., 2022; Saint-Aubin et al., 2021). The model suggests that producing the items enriches encoding, a form of elaboration, making the produced items more likely to hold distinctive features. At retrieval, these extra features are likely to improve recall, relative to competitors, especially for the recency items. Importantly, the addition of these useful features comes at a cost, in that production interferes with rehearsal. In other words, to explain the production effect, the RFM invokes a tradeoff between the benefits of distinctive features and the cost of generating those features. In the model, the latter cost is specified: reading aloud reduces the opportunities for rehearsal and rehearsal is beneficial as it helps to repair the damage done by retroactive interference.

Based on the predictions of the RFM, we asked whether increasing presentation speed would eliminate the disadvantage of produced items at the beginning of lists while maintaining their advantage at the end of the list. As a first test of this prediction, the systematic review showed a pattern that was entirely coherent with the expectations derived from the RFM. At slow presentation rates, silently read items were better recalled than produced ones in early serial positions, while for later positions it was the reverse. This cross-over pattern was eliminated at fast presentation rates where produced items continued to be better recalled for end of list positions, without a disadvantage for the primacy items. In Study 2, the direct experimental comparison of produced and silent conditions revealed that from a rate of 1000 ms per item, performance for silently read items is superior to performance for produced items for initial items. More importantly, this advantage for silent words is absent at 500 ms per item. Additionally, manipulation of the presentation rate had little impact on the production advantage for the recency positions.

The role of rehearsal in working memory is surrounded by some degree of controversy. For example, Souza and Oberauer (2018) suggested that there is little experimental evidence for a causal link between rehearsal and serial recall performance (see also Oberauer, 2019; Lewandowsky & Oberauer, 2015). They tested the causal effect of rehearsal in immediate serial recall by requiring that participants perform cumulative rehearsal (where all items presented so far are rehearsed aloud and in order). Their results failed to show an overall benefit of cumulative rehearsal. Importantly, when comparing cumulative rehearsal to simple repetition (with typed recall), their findings replicate the ones we report here, with production being equivalent to simple repetition, and our silent condition being like cumulative rehearsal. In other words, their findings produce an interaction between cumulative rehearsal and serial position; in the RFM that interaction is explained by calling upon subvocal rehearsal. Barrouillet et al. (2020) revisited the causal role of rehearsal. They argued that a limited number of items (about 4) can be effectively rehearsed. When they instructed participants to rehearse a subset of items corresponding to the hypothesized limit, they observed a large benefit. Under this view, Souza and Oberauer's (2018) cumulative rehearsal of six mainly multisyllabic words would have disrupted performance by exceeding rehearsal capacity. Interestingly, in the usual implementation of RFM, the rehearsal process is mostly restricted to the first four items.

Conclusion

Consistent with the results of our systematic review, our study indicates presentation rate changes the disadvantage of produced items at the beginning of the list while having little effect on the end of the list. Importantly, by systematically manipulating presentation rate, our results confirm our hypothesis that the interaction between production and serial position curve is due to rehearsal. The RFM can account for this complex pattern of results while being quite specific in terms of the mechanisms involved. The model argues for a central role of relative distinctiveness in retrieval, while suggesting that during encoding a number of processes compete. Production can be seen as a process that increases the number of encoded features and, depending on conditions, this can generate improved distinctiveness. However, production also interferes with rehearsal. The latter is important in the RFM as it is thought of as a form of covert retrieval from LTM based on a degraded cue held in working memory. Importantly, this covert retrieval can undo the damage to representations that retroactive interference inflicts. Importantly, the relatively simple architecture of the RFM can account for the findings reported here, as well as for those observed when other factors are manipulated and other tasks called upon – including tasks related to long-term memory. This architecture encompasses feature-based representation, similarity-based retroactive interference and rehearsal at encoding, and relative distinctiveness at the point of retrieval.

References

* Denotes articles included in the systematic review.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. Journal of Verbal Learning & Verbal Behavior, 14(6), 575–589. https://doi.org/10.1016/S0022-5371(75)80045-4

Barrouillet, P., Gorin, S., & Camos, V. (2021). Simple spans underestimate verbal working memory capacity. *Journal of Experimental Psychology: General*, *150(4)*, 633–665. https://doi.org/10.1037/xge0000957

*Battacchi, M. W., Pelamatti, G. M., & Umiltà, C. (1990). Is there a modality effect? Evidence for visual recency and suffix effects. *Memory & Cognition*, *18*(6), 651–658. https://doi.org/10.3758/BF03197107

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

Bhatarah, P., Ward, G., Smith, J., & Hayes, L. (2009). Examining the relationship between free recall and immediate serial recall: Similar patterns of rehearsal and similar effects of word length, presentation rate, and articulatory suppression. *Memory & Cognition*, *37*, 689–713. https://doi.org/10.3758/MC.37.5.689

*Bunt, A. A. (1976). Context effects in short-term memory: Confirmatory evidence from recall of visually presented lists. *Acta Psychologica*, *40*(6), 423–430. https://doi.org/10.1016/0001-6918(76)90021-4

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*, *61*(3), 457–469. https://doi.org/10.1016/j.jml.2009.06.002

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

Covidence Systematic Review Software. Veritas Health Innovation Ltd., Melbourne, Australia. https://www.covidence.org

*Crowder, R. G. (1970). The role of one's own voice in immediate memory. *Cognitive Psychology*, *1*(2), 157–178. https://doi.org/10.1016/0010-0285(70)90011-3

*Crowder, R. G. (1978). Memory for phonologically uniform lists. *Journal of Verbal Learning and Verbal Behavior*, *17*(1), 73–89. https://doi.org/10.1016/S0022-5371(78)90540-6

*Crowder, R. G. (1986). Auditory and temporal factors in the modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(2), 268–278. https://doi.org/10.1037/0278-7393.12.2.268

Cyr, V., Poirier, M., Yearsley, J. M., Guitard, D., Harrigan, I., & Saint-Aubin, J. (2022). The production effect over the long term: Modeling distinctiveness using serial positions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48(12)*, 1797–1820. https://doi.org/10.1037/xlm0001093.supp (Supplemental)

*Engle, R. W., Cantor, J., & Turner, M. (1989). Modality Effects : Do They Fall on Deaf Ears? *Quarterly Journal of Experimental Psychology Section A*, *41*(2), 273-292. https://doi.org/10.1080/14640748908402366

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fawcett, J. M., Baldwin, M. M., Whitridge, J. W., Swab, M., Malayang, K., Hiscock, B., Drakes, D. H., & Willoughby, H. V. (2023). Production improves recognition and reduces intrusions in between-subject designs: An updated meta-analysis. *Canadian Journal of Experimental Psychology*, *77*(1), 35–44. https://doi.org/10.1037/cep0000302

*Frick, R. W. (1989). Recency and the modality effect in immediate ordered recall. *Canadian Journal of Psychology*, *43*(4), 494–511. https://doi.org/10.1037/h0084231

*Gathercole, S. E., Gardiner, J. M., & Gregg, V. H. (1982). Modality and phonological similarity effects in serial recall: Does one's own voice play a role? *Memory & Cognition*, *10*(2), 176–180. https://doi.org/10.3758/BF03209219

Gionet, S., Guitard, D., & Saint-Aubin, J. (2022). The production effect interacts with serial positions: Further evidence from a between-subjects manipulation. *Experimental Psychology*, *69(1)*, 12–22. https://doi.org/10.1027/1618-3169/a000540

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning & Verbal Behavior*, *5(4)*, 351–360. https://doi.org/10.1016/S0022-5371(66)80044-0

*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*, 1909–1933. https://doi.org/10.1037/xlm0000430.supp (Supplemental)

*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. https://doi.org/10.1016/S0022-5371(84)90259-7

*Kappel, S., Harford, M., Burns, V. D., & Anderson, N. S. (1973). Effects of vocalization on short-term memory for words. *Journal of Experimental Psychology*, *101*(2), 314. https://doi.org/10.1037/h0035247

Landry, É. R., Guitard, D., & Saint-Aubin, J. (2022). Arousal affects short-term serial recall. *Canadian Journal of Experimental Psychology*, *76(2)*, 99–110. https://doi.org/10.1037/cep0000272

Lawrence, M. A. (2016). *ez: Easy analysis and visualization of factorial experiments*. *R package version 4.4-0*. https://CRAN.R-project.org/package=ez

Lewandowsky, S., & O berauer, K. (2015). Rehearsal in serial recall: An unworkable solution to the nonexistent problem of decay. *Psychological Review*, *122*(4), 674–699. https://doi.org/10.1037/a0039684

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

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science*, *26(4)*, 390–395. https://doi.org/10.1177/0963721417691356

Marin, J-M., Pudlom P., Robert, C. P., & Ryder, R. J. (2012). Approximate Bayesian computational methods. S*tatistics and Computing. 22*, 1167-1180. doi: 10.1007/s11222-011-9288-2

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau

(2005). *Tutorial in Quantitative Methods for Psychology, 4*, 61-64. http://dx.doi.org/10.20982/tqmp.04.2.p061?

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2022). BayesFactor: *Computation of Bayes Factors for Common Designs.* R package version 0.9.12-4.4. https://CRAN.R-project.org/package=BayesFactor

Mulligan, N. W., & Lozito, J. P. (2004). Self-generation and memory. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory, Vol 45*. (pp. 175–214). Elsevier Academic Press.

Mulligan, N. W., Smith, S. A., & Buchin, Z. L. (2019). The generation effect and experimental design. *Journal of experimental psychology. Learning, memory, and cognition*, *45*(8), 1422–1431. https://doi.org/10.1037/xlm0000663

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, *78*(4, Pt.1), 679. https://doi.org/10.1037/h0026641

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology*, *21*(3), 263–276. https://doi.org/10.1037/h0082978

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition*, *16*, 343–352. doi: https://doi.org/10.3758/BF03197045

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

*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. https://doi.org/10.1016/S0022-5371(83)90300-6

Neath, I., & Nairne, J. S. (1995). Word-Length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review*, *2*, 429-441. doi:10.3758/BF03210981

Neath, I., & Surprenant, A. M. (2007). Accounting for age-related differences in working memory using the feature model. In N. Osaka, R. H. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory: Behavioural and neural correlates* (pp. 165-179). Oxford, UK: Oxford University Press.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2 : A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 516–524. https://doi.org/10.3758/BF03195598

Oberauer, K. (2019). Is Rehearsal an Effective Maintenance Strategy for Working Memory? *Trends in Cognitive Sciences*, *23*(9), 798–809. https://doi.org/10.1016/j.tics.2019.06.002

Oberauer, K. (2022). When does working memory get better with longer time? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48(12)*, 1754–1774. https://doi.org/10.1037/xlm0001199

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958. https://doi.org/10.1037/bul0000153

Poirier, M., Yearsley, J.M., Saint-Aubin, J., Fortin, C., Gallant, G. and Guitard, D. (2019). Dissociating visuo-spatial and verbal working memory: It’s all in the features. *Memory and Cognition, 47(4)*, 603–618. doi:10.3758/s13421-018-0882-9.

Rohatgi, A. (2020). WebPlotDigitizer (Version 4.5) [computer software]. https://automeris.io/WebPlotDigitizer

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

*Routh, D. A. (1970). 'Trace strength' modality, and the serial position curve in immediate memory. *Psychonomic Science*, *18*(6), 355–357. https://doi.org/10.3758/BF03332397

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, *89*, 63–77. https://doi.org/10.1037/h0031185

*Saint-Aubin, J., Yearsley, J. M., Poirier, M., Cyr, V., & Guitard, D. (2021). A model of the production effect over the short-term: The cost of relative distinctiveness. *Journal of Memory and Language*, *118*, 104219. https://doi.org/10.1016/j.jml.2021.104219

Saint-Aubin, J., Poirier, M., Yearsley, J. M., Robichaud, J.-M., & Guitard, D. (2023). Modeling verbal short-term memory: A walk around the neighborhood. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49(2)*, 198–215. https://doi.org/10.1037/xlm0001226

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

Sisson, S. A., Fan, Y., & Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, *104(6)*, 1760–1765. https://doi.org/10.1073/pnas.0607208104

Sisson, S. A., Fan, Y., & Tanaka, M. M. (2009). Correction for Sisson et al., sequential Monte

Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, *106(39)*, Article 16889. https://doi.org/10.1073/pnas.0908847106

Souza, A. S., & Oberauer, K. (2018). Does articulatory rehearsal help immediate serial recall? *Cognitive Psychology*, *107*, 1–21. https://doi.org/10.1016/j.cogpsych.2018.09.002

Stoet, G. (2017). PsyToolKit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*, 24–31. https://doi.org/10.1177/0098628316677643

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*, 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1589–1625. https://doi.org/10.1037/0278-7393.26.6.1589

Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review*, *15(3)*, 535–542. https://doi.org/10.3758/PBR.15.3.535

Turner, B. M., & Van Zandt, T. (2012). A tutorial on approximate Bayesian computation. *Journal of Mathematical Psychology*, *56*, 69-85. doi:10.1016/j.jmp.2012.02.005

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, *6(3)*, 291–298. https://doi.org/10.1177/1745691611406923

Figure 1

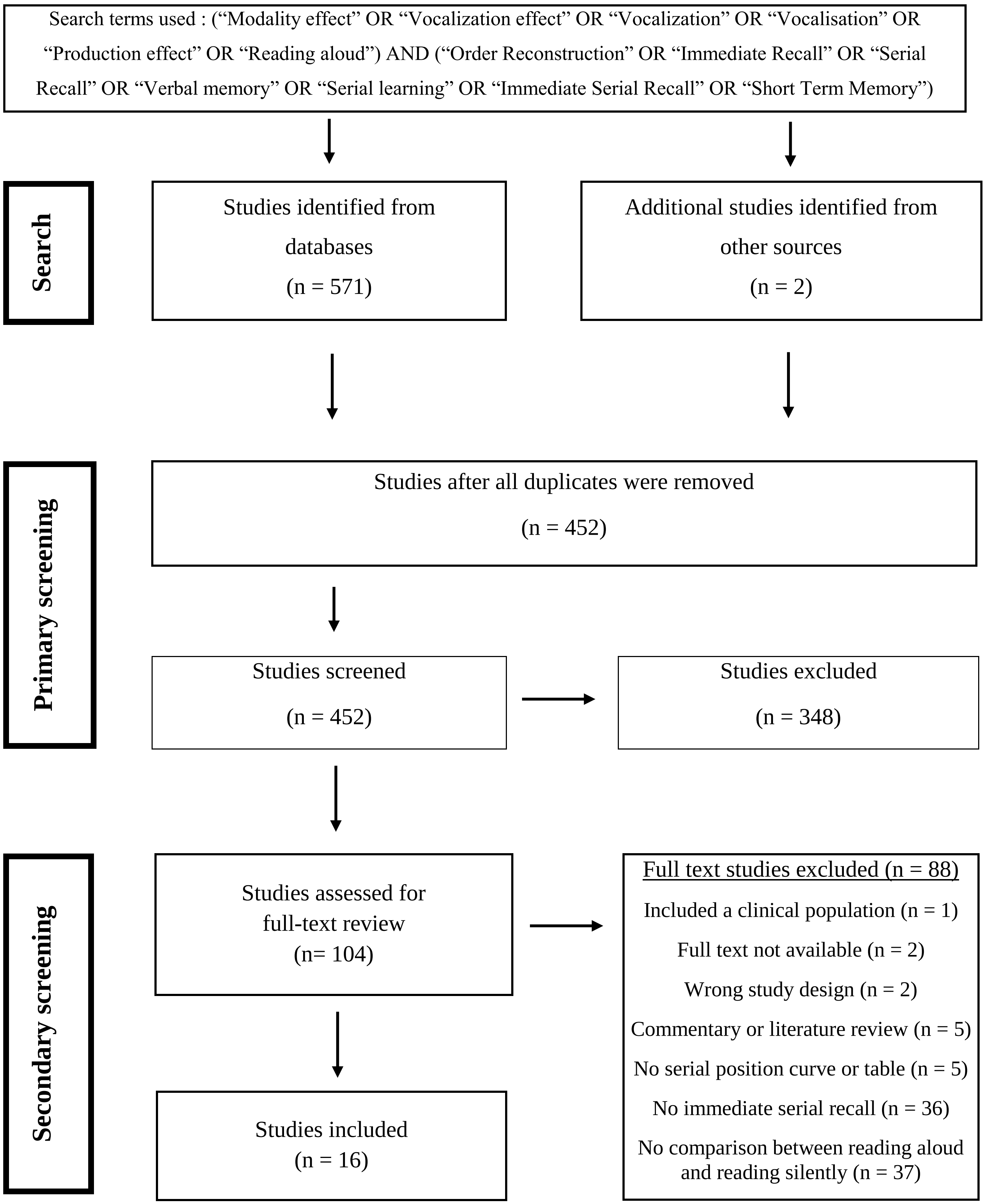
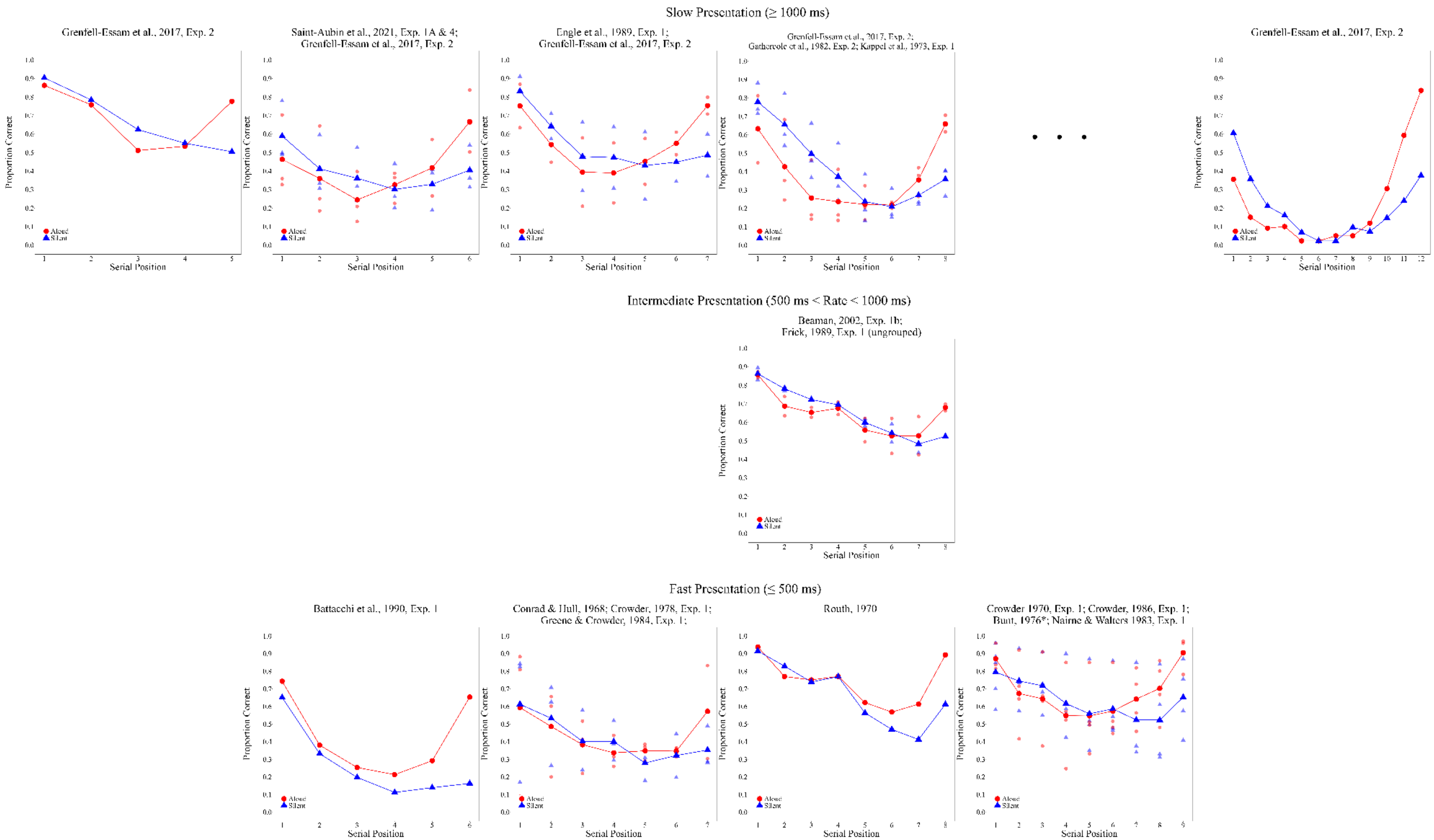

Search terms used : ("Modality effect" OR "Vocalization effect" OR "Vocalization" OR "Vocalisation" OR "Production effect" OR "Reading aloud") AND ("Order Reconstruction" OR "Immediate Recall" OR "Serial Recall" OR "Verbal memory" OR "Serial learning" OR "Immediate Serial Recall" OR "Short Term Memory")
Search
Studies identified from databases (n = 571)
Additional studies identified from other sources (n = 2)
Primary screening
Studies after all duplicates were removed (n = 452)
Studies screened (n = 452)
Studies excluded (n = 348)
Secondary screening
Studies assessed for full-text review (n= 104)
Studies included (n = 16)
Full text studies excluded (n = 88)
Included a clinical population (n = 1)
Full text not available (n = 2)
Wrong study design (n = 2)
Commentary or literature review (n = 5)
No serial position curve or table (n = 5)
No immediate serial recall (n = 36)
No comparison between reading aloud and reading silently (n = 37)

Figure 2

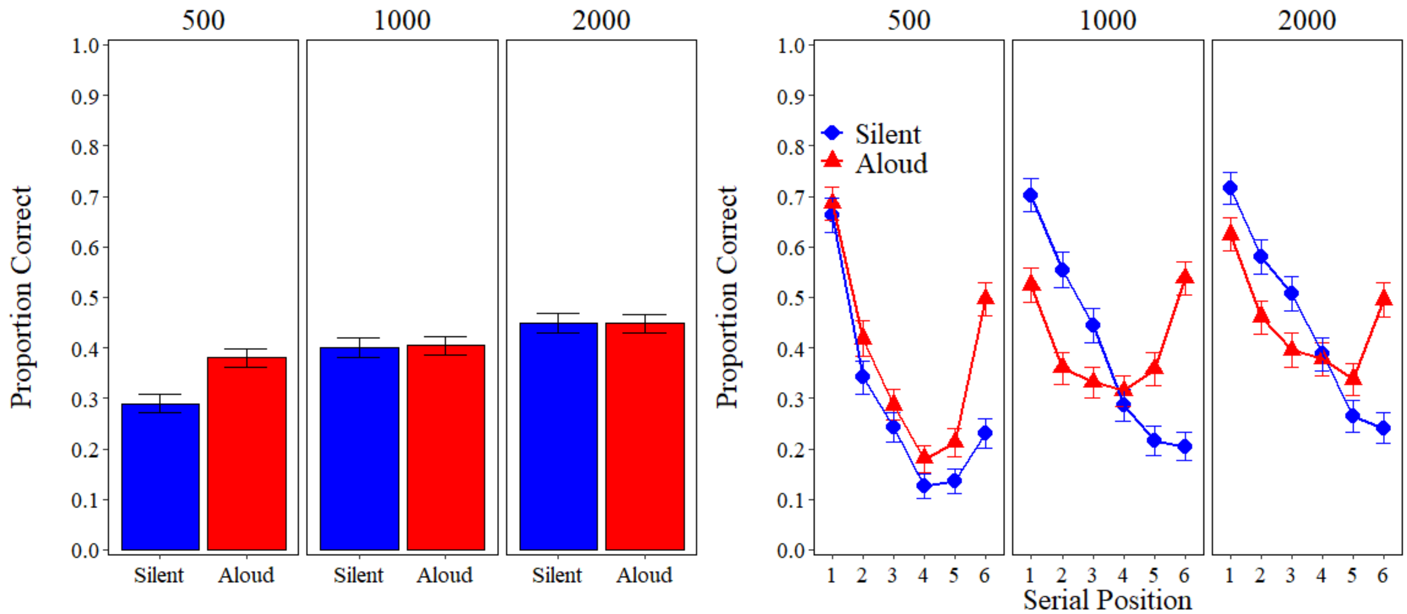


*Free-position recall

Proportion of correct response as a function of presentation modality (silent, aloud) and presentation rate (lesser or equal to 500 ms, greater than 500 and lesser than 1000 ms, greater or equal to 2000 ms), and serial position (1 to 5; 1 to 6; 1 to 7; 1 to 8; 1 to 9; 1 to 12).

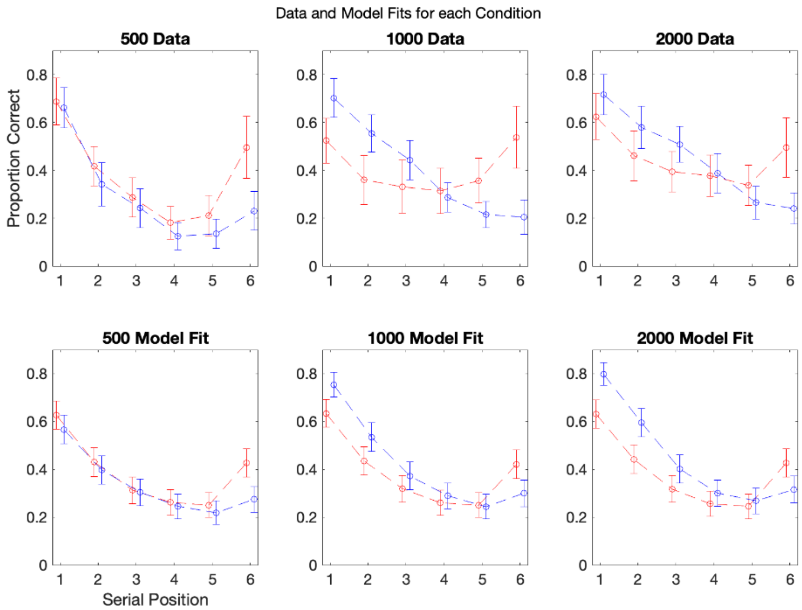
Figure 3

Proportion of correct response as a function of presentation modality (silent, aloud) and presentation rate (500 ms, 1000 ms, 2000 ms), and serial position (1 to 6).



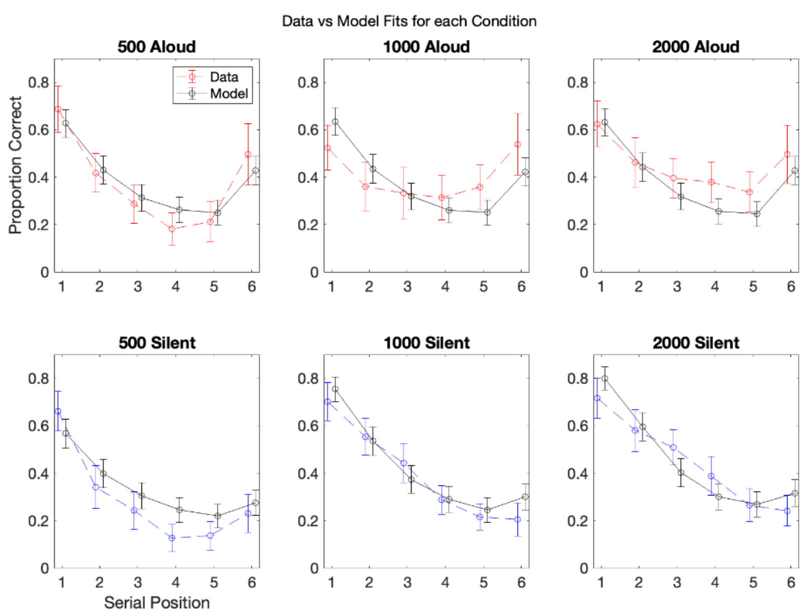
Note. **Left column**, results as a function of presentation modality (silent, aloud) and presentation rate (500 ms, 1000 ms, 2000 ms), **right column**, results as a function of presentation modality (silent, aloud) and presentation rate (500 ms, 1000 ms, 2000 ms), and serial position (1 to 6). Error bars represent 95% within-participant confidence intervals computed according to Morey's (2008) procedure.

Figure 4:



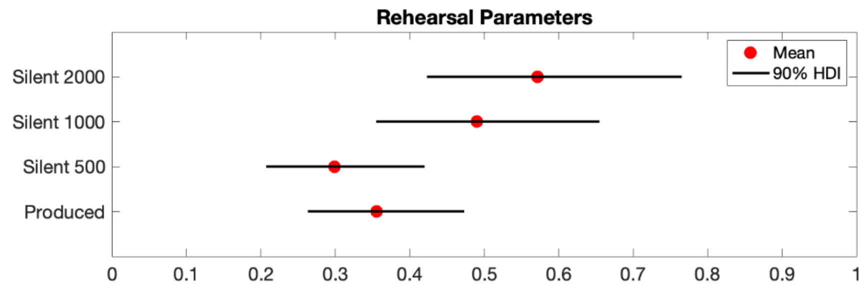
Comparison of serial position curves for the aloud (red) and silent (blue) conditions. Top row shows data, bottom row shows curves generated from the RFM with best fitting parameters. The important comparison here is that the qualitative features, e.g. the presence or absence of a crossover, are reproduced by the model fits.

Figure 5:



Comparison of data (colored lines) and curves generated from the RFM with best fitting parameters (black lines), for each condition. Given the strict demands on how the parameters were allowed to vary between conditions, fits are overall good.

Figure 6:



Comparison of the key parameters, those governing rehearsal, for each of the four distinct conditions. Dots show means of the posteriors, lines show 90% Highest Density Intervals (HDI).

Appendix A: Modelling Details

The RFM is too complex for an analytic expression for the likelihood to be derived, so as with all previous attempts to fit the model to data we used a version of Approximate Bayesian Computation (ABC) (see Turner & Van Zandt, 2012, or Marin et al., 2012, for a review). Following Poirier et al. (2019), Saint-Aubin et al. (2021), and Cyr et al (2021) we used ABC Partial Rejection Control (ABC-PRC) (Sisson et al., 2007, 2009). ABC-PRC works by repeatedly sampling from a prior over the parameter space until it finds a set of parameters which generate a set of summary statistics (in our serial position curves) sufficiently close to the data, as determined by the discrepancy function. When this happens, the algorithm stores these parameter values, and moves on to the next particle in the generation. Once all particles in a generation have been associated with parameter sets, the algorithm gives each particle a weight depending on the prior, and then begins a new generation, sampling from the previous generation with probabilities given by the weights, and repeatedly perturbing around the previous parameter values until a set is found producing summary statistics even closer to the data. Once the required number of generations have elapsed posterior estimates for the parameters can be obtained as the fraction of particles in the final generation with that parameter value. Posterior predicted distributions of the summary statistics are also easily obtained. For full details see Sisson et al. (2007) (Note also the errata, Sisson et al., 2009).

As explained in the main text, we are fitting six data sets, one for each condition, but our hypothesis is that only the rehearsal rate varies between the different presentation rate conditions. We therefore split the parameters into two groups, we fit lambda and tau but demanded these be the same for each condition, and we fit a number of rehearsal rates which we allowed to vary between conditions. In addition, as is standard, we assumed the aloud conditions had more modality dependent features. Full details are given in Table A1.

The important parameters for ABC-PRC are the number of particles (set to 1000 for all fits reported here), the details of the prior, the proposal distributions, and the minimum tolerances for each fit. The proposal and tolerances can be found in the code on the OSF. Priors, and resulting posterior distributions are summarized in Table A1.

Parameter	**Prior**	**Posterior Mean (90% HDI)**
Lambda (Overwriting)	Normal(1,0.2)	0.935 [0.749, 1.129]
Tau (Temperature)	Normal(0.1,0.02)	0.117 [0.095, 0.139]
Rehearsal (Aloud)	Beta(2,2)	0.361 [0.263, 0.473]
Rehearsal (Silent 500ms)	Beta(2,2)	0.305 [0.207, 0.420]
Rehearsal (Silent 1000ms)	Beta(2,2)	0.494 [0.355, 0.655]
Rehearsal (Silent 2000ms)	Beta(2,2)	0.579 [0.423, 0.765]

Table A1: Priors and descriptives for the posteriors for the fitted parameters in the model fitting.