Phonological Similarity and Recall Dynamics

Haydn G. Herrema and Michael J. Kahana

University of Pennsylvania

Author Note

Support for this work was provided by the National Institutes of Health (MH55687). Correspondence concerning this article should be addressed to Michael J. Kahana. E-mail: kahana@psych.upenn.edu

Abstract

Similarities among items govern how individuals organize their memories, as seen in the order and timing of successively recalled words in free recall memory tasks. Prior work has documented the influence of similarity in temporal, semantic, emotional, and physical space on recall dynamics. We asked whether phonological similarity also impacts the dynamics of free recall. Analyzing data from a large delayed free recall experiment, we find that participants cluster their recalls based on alliterative and rhyming patterns in the encoded words, and such organization predicts recall performance. We also discover that phonological features induce false memories, and that phonological similarity interacts with temporal and semantic features to accelerate recall transitions. Our results further catalog how similarity, in its many forms, dictates the dynamics of free recall.

Keywords: memory, phonological similarity, free recall, recall organization, false memory

Phonological Similarity and Recall Dynamics

Introduction

Decades of research on the free recall paradigm have elucidated the overarching theme that similarity supports memory. Considerable work has highlighted temporal contiguity as an operative mechanism of the memory system. During retrieval, transitions between items experienced close in time occur with greater probability and shorter latencies (Kahana, 1996; Healey, Long, & Kahana, 2019). The temporal contiguity effect represents a universal principle of memory (Healey & Kahana, 2014) and participants who exhibit stronger temporal organization achieve superior overall recall (Sederberg, Miller, Howard, & Kahana, 2010; Healey, Crutchley, & Kahana, 2014). Likewise, many analyses have exemplified that semantic features play a role in organizing memory. Recall transitions between semantically related items occur with greater probability and faster inter-response times (IRTs) (Howard & Kahana, 2002; Sederberg et al., 2010). Healey and Kahana (2014) demonstrated that the semantic proximity effect represents another universal principle of memory, and Healey et al. (2014) found that semantic organization predicts overall recall. Temporal and semantic associations represent "laws" of memory(Kahana, Diamond, & Aka, 2024), but they are not the only forms of similarity-based clustering during recall.

Emotional characteristics organize recall, as transitions between words with similar emotional valence are more likely than transitions between words with different valences (Long, Danoff, & Kahana, 2015; Cohen & Kahana, 2022). In spatial memory tasks, recall transitions between items encoded close in space occur with greater probability and shorter latencies (Miller, Lazarus, Polyn, & Kahana, 2013; Dougherty et al., in press). In experiments that manipulate how participants encode studied items, participants cluster their recalls based on the encoding task (Polyn, Norman, & Kahana, 2009). Even the recency effect (Murdock, 1962) represents a form of similarity, as participants recall with greater probability items whose encoding contexts bear (temporal) similarity to the context at the time of the test.

The foregoing examples illustrate how similarity promotes correct recall transitions, but similarity can also hinder recall, inducing participants to commit false memories. Zaromb et al. (2006) demonstrated how temporal features influence memory errors, as incorrect recalls of items that were previously studied tend to come from recent lists. Moreover, Zaromb et al. (2006) revealed that semantic features incite intrusions, with greater semantic relatedness for transitions from a correct recall to an incorrect recall than transitions between correct recalls. Herz, Bukala, Kragel, and Kahana (2023) even showed that neural activity differentiates false memories of varying semantic similarity to studied items.

This paper investigates the potential role of phonological similarity as an organizing principle of human memory search. We hypothesize phonological similarity can both explain data on recall transitions and recall errors. To investigate these questions, we examine recall dynamics and intrusion errors in a large dataset of delayed free recall of word lists (Kahana, Lohnas, et al., 2024). We also consider how our analyses relate to other known phonological similarity effects on memory.

Considerable prior work has examined the role of phonological similarity in memory paradigms other than free recall, with significant evidence of a phonological similarity detriment to memory. Phonologically similar sequences of words decrease serial recall performance, regardless of whether item presentation is visual or auditory (Baddeley, 1966b, 1966a, 1968; Spurgeon, Ward, & Matthews, 2014). When the studied items are letters, removing possible confounds of semantic relations, such a phonological similarity decrement still exists (Conrad, 1964). Likewise, in memory span tasks, performance worsens with sequences of phonologically similar words (Conrad, 1963; Mueller, Seymour, Kieras, & Meyer, 2003; Copeland & Radvansky, 2001) or letters (Conrad & Hull, 1964). In probed recall, recall probability for end-of-list items is worse on lists of homophone pairs than lists of unrelated words (Kintsch & Buschke, 1969). In a continuous recognition task, words phonologically similar to previous items produce more false alarms (Anisfeld, 1969) and using lists of phonologically similar words in a standard recognition paradigm, lures that are phonologically similar to the studied items elicit greater false alarm rates than unrelated lures (Lim & Goh, 2019). Crafting lists of words phonologically similar to a specific unstudied item, recognition of the phonologically similar critical item is greater than of other unstudied items (Sommers & Lewis, 1999), and the false recognition of the critical item increases with instructions to attend to the sound of the words when encoding rather than the meaning (Chan, McDermott, Watson, & Gallo, 2005). Participants frequently misjudge pairs of words as semantically related when one word is a homophone of a truly related word (Luo, Johnson, & Gallo, 1998). In naturalistic settings, phonological similarity impacts the misnaming of familiar individuals (Deffler, Fox, Ogle, & Rubin, 2016).

However, some work has demonstrated conditions for a phonological similarity benefit to memory. In reading span paradigms, performance increases in the phonologically similar condition (Copeland & Radvansky, 2001) and this facilitation persists in reading span tests with scrambled sentences or varying distractor pool sizes (Chow, Macnamara, & Conway, 2016). While phonologically similar lists hinder performance in reconstruction of order tasks with a 2-second distractor interval between encoding and retrieval, increased performance occurs with a 24-second distractor interval (Nairne & Kelley, 1999; Fournet, Juphard, Monnier, & Roulin, 2003). This phonological similarity benefit following longer distractor intervals also occurs in serial recall tasks (Fournet et al., 2003). In a cued recall task where some cues rhymed with studied words and others with non-studied words, participants give higher recognition ratings for cues corresponding to studied items, even without successful identification of the target items (Cleary, 2004)

Moreover, there have been some investigations of the influence of phonological similarity in free recall. Considerable evidence points to phonological similarity benefits. Fournet et al. (2003) observed better recall performance on short lists of phonologically similar items regardless of the retention interval (2, 8, or 24 seconds). Comparing longer lists of unrelated words to lists with clusters of phonologically similar words in the middle or end of the list, Craik and Levy (1970) found that recall probability increases for the phonologically similar items. Watkins, Watkins, and Crowder (1974) had participants encode lists of unrelated or phonologically similar words either by reading silently or aloud and found that the recall of pre-recency items is greater in the phonologically similar condition regardless of modality. Gupta, Lipinski, and Aktunc (2005) created lists of alliterative, rhyming, phonologically similar (mix of alliterative and rhyming) and dissimilar lists and, using free recall scoring of a serial recall task, discovered that performance on rhyming and alliterative lists is greater than on dissimilar lists, so long as words are drawn from an open set. However, if words are repeated across lists, the greatest performance occurs on dissimilar lists (Gupta et al., 2005), which is not the only evidence of phonological similarity deficits in free recall. Spurgeon et al. (2014) found worse recall performance on lists of phonologically similar words when visual encoding occurs silently. When lists are constructed to be phonologically similar to a specific word, Sommers and Lewis (1999) found that incorrect recall of the critical item occurs more often than of unrelated words and Chan et al. (2005) found the same false recall under instructions to pay attention to the sound of the words rather than their meaning during encoding.

Building on prior work on similarity effects in free recall and phonological similarity in various memory paradigms, we aim to address several questions about the role of phonological similarity in free recall. First, can we detect a phonological clustering effect, whereby participants are more likely to transition between items that are phonologically similar? Mirroring temporal, semantic, and other forms of similarity-based organization, we hypothesize that participants display significant phonological clustering. Second, can phonological similarity impair free recall by inducing false memories? Temporal and semantic associations are known to generate incorrect responses, and phonological similarity leads to errors in recognition, recall, and naturalistic retrieval. Therefore, we hypothesize that phonological features play a part in generating recall errors. Third, how does phonological clustering, in conjunction with temporal and semantic organization, relate to recall performance? Evidence from serial recall and memory span tasks suggests a phonological similarity detriment, but work in free recall has shown temporal and semantic organization relate positively with overall recall. Thus, if highly practiced participants display phonological organization of memory, we'd expect such order to be beneficial to recall performance. Finally, do transitions between phonologically similar items occur faster? We hypothesize that phonological similarity accelerates IRTs in the same way that temporal contiguity and semantic relatedness do.

Methods

Participants

Participants were 127 native English speaking, neurocognitively healthy young adults. The mean age among participants was 20.86 (SD = 2.96) and the oldest participant was 30. Participants received a base salary for participation in the experiment, as well as modest bonuses for performance and completion of all sessions. The University of Pennsylvania Institutional Review Board reviewed and approved all recruitment and testing procedures.

Materials

The wordpool consisted of a 576 nouns. Word lists were constructed such that adjacent items would have a variable degree of semantic similarity to one another. Using the Word Association Spaces (Steyvers, Shiffrin, & Nelson, 2004), semantic similarity scores were calculated for all word pairs, which were divided into four semantic similarity bins. On each list of 24 words, there were two pairs of words drawn from each bin, one of which was presented contiguously and the other with at least two intervening items. The other eight words on the list were chosen randomly.

For our analyses of phonological similarity, we used the Natural Language Toolkit (Bird, Klein, & Loper, 2009) corpus to break words into their constituent phonemes. For our analyses of semantic similarity, we used Google's Word2Vec pre-trained Google News model to transform words into 300-dimensional vector representations and calculated the semantic similarity of two words as the cosine similarity between the vector representations. We preferred Word2Vec over Word Association Spaces for our analyses as the Word2Vec model allowed us to easily obtain semantic representations for over 2,000 recalls that were outside of the original wordpool.

For all analyses, statistical testing, and visualization, we used Python and R.

Design and Procedure

We analyzed data from Experiment 4 of the Penn Electrophysiology of Encoding and Retrieval study (Kahana, Lohnas, et al., 2024). In this experiment, participants completed up to 23 sessions of free recall, and each session consisted of 24 experimental trials. Every trial contained a list of 24 words presented visually one at a time, then a 24-second distractor period during which participants completed simple arithmetic problems of the form "X + Y + Z = ?", where X, Y, and Z were single-digit integers, and then a 75-second free recall test. There was an additional 24-second math distractor preceding list presentation on a random half of the trials. Participants were given a short break after every eight trials in a session. All 576 words in the wordpool were presented in each session, but the order was randomized. The experiment also contained a 24th session that used 288 words from the previous sessions and introduced 288 novel words, which we did not analyze to avoid confounds of differential familiarity with the studied items.

Phonological Similarity

In order to measure the impacts of phonological similarity on memory behavior, we must first define a measure of phonological similarity. Early studies defined phonological similarity qualitatively (Conrad, 1963; Baddeley, 1966b, 1966a, 1968; Craik & Levy, 1970) – simply based on whether words sound the same. While some more recent work has utilized similar subjective classifications (Chow et al., 2016; Copeland & Radvansky, 2001; Cleary, 2004), there have been several attempts at a more rigorous, quantitative definition of phonological similarity (Mueller et al., 2003; Vitz & Winkler, 1973; Li & MacWhinney,

2002) that often take inspiration from linguistics (Chomsky & Hall, 1968).

Given our work represents a preliminary investigation of phonological similarity in free recall, we sought to develop a rigorous yet simple measure. To quantify the phonological similarity of words, we used the Natural Language Toolkit (Bird et al., 2009) to break words into their constituent phonemes. We defined two measures of phonological similarity, which allowed us to test slightly different hypotheses.

A priori, we believed the most salient phonological features of a word are the start and the end, which align with alliteration and rhyme, respectively. The primacy and recency effects in free recall (Murdock, 1962) indicate that participants best remember the start and end of lists. Analogously, we suspected participants best remember the start and end of words. Therefore, we defined a binary representation of phonological similarity, classifying two words as phonologically similar if they are alliterative or rhyming – they share the same first phoneme or the same last two phonemes. Formally, with words i and j that have n and m phonemes, respectively:

$$phonological\ similarity = \left\{ \begin{array}{ll} 1, & \text{if } p_{i,1} = p_{j,1} \\ 1, & \text{if } p_{i,n-1} = p_{j,m-1} \text{ and } p_{i,n} = p_{j,m} \\ 0, & \text{otherwise} \end{array} \right\} \tag{1}$$

However, canonical analyses of temporal and semantic clustering (Polyn et al., 2009) require a ranking of relative similarities that our binary metric does not afford. Additionally, we sought to test the hypothesis that all phonemes of words, not just the first and last two, determine the phonological similarity. Therefore, we calculated a continuous measure of phonological similarity using the Jaccard index (Jaccard, 1912), which is the ratio of the size of the intersection and union of two sets:

$$phonological\ similarity = \frac{|p_i \cap p_j|}{|p_i \cup p_j|} \tag{2}$$

where p_i and p_j are the sets of phonemes in words i and j. Notably, the Jaccard index does not enforce the order of the phonemes – two words that share phonemes sound similar even if the shared phoneme is not in the same position.

Phonological Clustering Scores

For our binary metric that classifies words as phonologically similar if they are alliterative or rhyming, we defined a novel *phonological clustering score* to measure whether the first and last two phonemes of words governed recall transitions to a degree greater than chance. For every transition between correct recalls i and j, we determined if there was a remaining list word (i.e., not already recalled) that was phonologically similar to i. If so, we calculated the phonological similarity of recalls i and j, which would either be 1 or 0 by our binary representation. Then, we determined the random probability of transitioning from i to a phonologically similar word, which was the ratio of the number of remaining phonologically similar words left from the list to the total number of remaining words on the list. The difference between the phonological similarity of the actual transition and the random probability of transitioning to a phonologically similar word represents the *phonological clustering score (pcs)* for the transition. Formally,

$$pcs = psim(i,j) - \frac{\sum_{w \in W} psim(i,w)}{|W|} \text{ if } \exists\, k \in W \mid psim(i,k) = 1 \tag{3}$$

where i and j are neighboring recalls and W is the set of remaining (not yet recalled) words from the list. By this measure, a 0 clustering score represents chance.

For our continuous metric that calculates the phonological similarity of words using the Jaccard index, we utilized the same ranking method as for the semantic and temporal clustering scores from (Polyn et al., 2009). By this measure, a 0.5 clustering score represents chance.

Statistical Testing

In our analysis of the relation of phonological, temporal, and semantic clustering to recall performance, we fit a linear mixed effects model to jointly model recall probability as a function of the three forms of organization and their pairwise interactions. We fit two models and used the Akaike Information Criteria (AIC) (Akaike, 1974) and Bayes Information Criteria (BIC) (Schwarz, 1978) for model selection. Both models included random intercepts for participant to account for individual differences and for session and list to account for learning effects. In one model we also allowed the slopes for phonological, temporal, and semantic clustering to vary by participant, and it was this model that provided a better fit to the data. Formally, the chosen model specification was:

$$\begin{aligned} \text{P(Recall)} \sim\ & \text{pcs} + \text{scs} + \text{tcs} + \\ & \text{pcs} : \text{scs} + \text{pcs} : \text{tcs} + \text{scs} : \text{tcs} + \\ & (1 + \text{pcs} + \text{scs} + \text{tcs} \mid \text{participant}) + \\ & (1 \mid \text{session}) + (1 \mid \text{list}) \end{aligned} \tag{4}$$

where *pcs*, *scs*, and *tcs* represent phonological, semantic, and temporal clustering scores, respectively.

In our analysis of inter-response times (IRTs), we sought to understand if phonological similarity (*psim*) modulates IRTs. To account for other known factors of IRTs, we fit a linear mixed effects model that contained fixed effects for semantic similarity (*ssim*), absolute temporal lag *(abs_lag)*, output position in the recall sequence (*outpos*), number of correct recalls on the trial (*ncr*), and the number of phonemes in the first word of the transition pair (*len_ph*). We also included all pairwise interactions between phonological similarity, semantic similarity, and absolute temporal lag. We fit two models and used the AIC and BIC for model selection. Both models included random intercepts for participant to account for individual differences and for session and list to account for learning effects. In one model we also allowed the slope for phonological similarity to vary by participant, as unlike our other well-founded factors, it was unknown whether phonological similarity had a universal effect on IRTs, and it was this model that provided a better fit to the data. Formally, the chosen model specification was:

$$\begin{aligned}\text{IRT} \sim\ & \text{psim} + \text{ssim} + \text{abs\_lag} + \\ & \text{outpos} + \text{ncr} + \text{len\_ph} + \\ & \text{psim} : \text{ssim} + \text{psim} : \text{abs\_lag} + \text{ssim} : \text{abs\_lag} + \\ & (1 + \text{psim} \mid \text{participant}) + (1 \mid \text{session}) + (1 \mid \text{list})\end{aligned} \tag{5}$$

We used the *lme4* package in R for the linear mixed effects models. We used the *pingouin* package in Python for repeated measures ANOVA and subsequent pairwise tests. For other statistical tests, which included one sample t-tests, paired t-tests, and Pearson correlations, we utilized methods from the *scipy* statistics library and corrected for multiple comparisons with the Benjamini-Hochberg method (Benjamini & Hochberg, 1995).

Data and Code Availability

All experimental and participant data is available open source at `https://openneuro.org/datasets/ds004395`. The public repository `https://github.com/hherrema/pho_sim_fr` contains all code for data processing, analysis, visualization, and statistical testing. This study was not preregistered.

Results

Phonological clustering occurs based on alliteration and rhyme

We first report phonological clustering scores for our different phonological similarity metrics. Much work has demonstrated the organization of memory by temporal (Kahana, 1996; Healey et al., 2019) and semantic (Howard & Kahana, 2002) features in free recall. Polyn et al. (2009) characterized these effects with clustering scores, which compare the temporal or semantic associations of actual recall transitions to the distributions of associations of possible transitions. We hypothesized that participants organized their recalls by phonological features as well. First, we applied a binary classification, considering words to be phonologically similar if they were alliterative (same first phoneme) or rhyming (same last two phonemes) and calculated our novel *phonological* *clustering score* (see Methods). To subsequently test whether alliteration or rhyme contributed differently to any phonological organization, we separately considered words as phonologically similar if they shared the same first phoneme or the same last two phonemes. We also used the Jaccard index (Jaccard, 1912) as a continuous measure of phonological similarity to test whether all phonemes of words play a role in guiding memory, calculating phonological clustering scores with the same ranking method as for semantic and temporal clustering scores (Polyn et al., 2009). To directly compare the importance of the first and last two phonemes to all phonemes using the same metric, we ran our analysis with the continuous Jaccard index on only the first and last two phonemes of words. We calculated phonological clustering scores for each list, then averaged lists within session, and then averaged sessions within participant.

Figure 1A illustrates the between-participant average phonological clustering scores for each similarity metric. We calculated one-sample t-tests against chance phonological clustering with a two-sided alternative hypotheses, and corrected for multiple comparisons with the Benjamini-Hochberg method (Benjamini & Hochberg, 1995). We found phonological clustering significantly greater then chance using the binary metric when defining phonologically similar words by alliteration or rhyme, $t(126) = 10.47, p < 0.001$, by the first phoneme only, $t(126) = 7.37, p < 0.001$, and by the last two phonemes only, $t(126) = 10.57, p < 0.001$. We did not find phonological clustering different than chance when calculating phonological similarity by the Jaccard index, considering all phonemes of words, $t(126) = 0.10, p = 0.918$. However, when using the Jaccard index on only the first and last two phonemes, we again found phonological clustering significantly greater than chance, $t(126) = 3.37, p = 0.001$. Therefore, two measures of phonological similarity converged to show that participants display phonological organization of recalls based on the first and last two phonemes of words.

Phonological features induce false memories

We next investigate the role phonological similarity plays in generating incorrect recalls. As in prior work (Kahana, Dolan, Sauder, & Wingfield, 2005), we partitioned false memories into prior-list intrusions (PLIs), which are recalled words that appeared on a previous list, and extra-list intrusions (ELIs), which consist of all other incorrect recalls. Analyses of PLIs have shown that such intrusions are more likely to come from prior lists that are more recent to the current list (Zaromb et al., 2006), suggesting that temporal similarity acts as a generating mechanism for intrusions. Comparing transitions from a correct recall to an intrusion to transitions between correct recalls, Zaromb et al. (2006) also found that intrusions were more semantically similar to the preceding recall than correct recalls. In a recognition task, Lim and Goh (2019) observed greater false alarm rates for lures that were phonologically similar to studied items, as did Anisfeld (1969) in a continuous recognition paradigm. Sommers and Lewis (1999) induced false recognition and recall of specific words by crafting lists of words phonologically similar to the critical item, which Chan et al. (2005) also achieved by instructing participants to attend to the sounds of the words during encoding. In naturalistic settings, Deffler et al. (2016) found that when people misnamed a familiar individual, they often did so with a name that was phonologically similar to the correct one. Thus, we hypothesized that intrusions would be phonologically similar to the words on the list – incorrect recalls are phonological confusions of would be correct recalls. Moreover, given the temporal contiguity effect already partly explains PLIs, we also hypothesized that phonological similarity played a greater role in inducing ELIs than PLIs.

To test how the presented list related to false memories, for every intrusion, we calculated the mean phonological similarity to the words presented on the immediately preceding list. As a control, we randomly selected n words presented on other lists during the session, where n equals the number of items presented per list, and determined the mean phonological similarity to the randomly drawn items. Running 100 permutations of this control for each intrusion, we got an estimate of the phonological similarity of the intrusion to a random list of words from the wordpool. Therefore, for each intrusion, we had observed and control values for the phonological similarity to a list of words, and subtracting the two gave us a measure of whether phonological features of the presented word list induced the intrusion. We ran this analysis with both our binary metric that classifies alliterative or rhyming words as phonologically similar and with our continuous metric that considers all phonemes.

For our binary metric, we illustrate the observed and control values for PLIs and ELIs, as well as explicitly depict the mean differences between observed and control values, in Figure 2A. We display the same analyses by the continuous metric in Figure 2B. The greater control values for PLIs than ELIs provided assurances of our methodology, as it was possible to randomly sample the PLI from a prior list in the permutation control. To test the hypothesis that phonological features induced intrusions, we ran one-sample t-tests with a two-sided alternative hypothesis on the delta (observed - control) values for PLIs and ELIs. To test the hypothesis that phonological similarity generated ELIs moreso than PLIs, we ran paired t-tests with a two-sided alternative hypothesis on the observed values for ELIs and PLIs. We applied a FDR correction with the Benjamini-Hochberg method (Benjamini & Hochberg, 1995) on all comparisons. For our binary metric, we found that phonological features induced both PLIs, $t(126) = 7.21, p < 0.001$, and ELIs, $t(126) = 14.63, p < 0.001$, and that ELIs were more phonologically similar to list items than PLIs, $t(126) = 4.58, p < 0.001$. For our continuous metric, we likewise found that phonological features induced both PLIs, $t(126) = 4.37, p < 0.001$, and ELIs, $t(126) = 15.45, p < 0.001$. While we observed the trend that ELIs were more phonologically similar to list items than PLIs using the continuous metric, this effect was not significant, $t(126) = 1.91, p = 0.058$.

Next, we tested whether the immediately preceding recall induced intrusions via phonological similarity. For every transition from a correct recall to another correct recall, PLI, or ELI, we calculated the phonological similarity of the two words, using both our binary and continuous representations. As a control, we calculated the mean phonological similarity between all pairs of words presented on the list. For our binary metric, we display the average phonological similarities of these three transition types along with the control values in Figure 2C. We also explicitly display the difference between the phonological similarities of the three response types and the control values. We do the same for our continuous metric in Figure 2D. To statistically test these results, we ran a repeated-measures ANOVA on the observed phonological similarity values, treating response type (correct recall, PLI, ELI) as a within-subjects factor. Using our binary metric, we did not observe a significant main effect of response type, $F(2, 252) = 2.18, p = 0.130$. However, using our continuous metric, we found a significant main effect of response type, $F(2, 252) = 3.87, p = 0.022$, and subsequent FDR corrected (Benjamini & Hochberg, 1995) pairwise tests showed significant differences between ELIs and correct recalls, $p = 0.039$ and ELIs and PLIs, $p = 0.039$, but not between PLIs and correct recalls, $p = 0.647$.

Considering alliteration and rhyme or all phonemes, recall intrusions display greater phonological similarity to just-studied items than other words in the wordpool, with ELIs moreso than PLIs. Also, ELIs display greater phonological similarity to the preceding recall than PLIs and correct recalls when calculating phonological similarity based on all phonemes of words. Taken together, it appears that phonological features played a role in generating false memories.

Phonological organization predicts recall performance

We have demonstrated a phonological clustering effect based on alliteration and rhyme, suggesting that participants used phonological features to guide memory. Presumably, highly practiced participants would only persist in exhibiting phonological organization if it were beneficial to recall. However, we also found that phonological similarity induces memory errors – attending to phonological features can be detrimental to recall. Therefore, we sought to understand how phonological clustering relates to overall recall performance. In an analysis across many free recall studies, Sederberg et al. (2010) found that temporal clustering related positively with recall performance, and in a study that ensured lists contained semantically related items, Healey et al. (2014) discovered that both temporal and semantic organization predicted overall recall. Some evidence suggests phonological similarity hampers free recall. Spurgeon et al. (2014) observed decreased performance on lists of phonologically similar words with silent visual encoding. Using free recall scoring of a serial recall task, Gupta et al. (2005) observed greatest performance on phonologically dissimilar lists when words repeated across lists; however, when words came from an open set, recall was greater on rhyming and alliterative lists. Numerous other examples demonstrate phonological similarity benefits in free recall. Fournet et al. (2003) found improved free recall performance on 5-item lists of phonologically similar words. With 12-item lists of phonologically similar words, Watkins et al. (1974) observed greater recall of pre-recency items compared to unrelated lists. With lists containing portions of adjacent phonologically similar words either in the middle or at the end of the list, Craik and Levy (1970) observed increased recall probability for the phonologically similar items. Therefore, we hypothesized that phonological clustering by alliteration and rhyme would relate positively with recall rate, given we observed a significant clustering effect. However, we did not expect phonological clustering based on all phonemes of words to relate significantly to overall recall.

For every list, we calculated the recall rate (proportion of items correctly recalled), the phonological clustering score (by our binary and continuous metrics), and the temporal and semantic clustering scores. We calculated Pearson correlations between the four clustering scores and recall rates across lists within session, and then averaged the correlation R values across sessions within participant. We illustrate the participant-average Pearson R correlations between the four clustering scores and recall probability in Figure 3A. However, the separate correlations with recall probability for the different clustering scores do not account for the fact that the recall sequence on any given trial results from the interplay of temporal, semantic, and phonological organization. Therefore, we ran a linear mixed effects model where we modeled recall rate as a function of phonological, semantic, and temporal clustering, as well as all pairwise interactions of the clustering scores. We also included random intercepts for participant, session, and list to account for individual differences in memory performance and potential learning effects across lists and sessions. Additionally, we allowed the slopes for phonological, semantic, and temporal clustering to vary by participant. We completed this process for our two phonological clustering metrics separately, reporting results from the models in Table 2 and showing the participant slopes for phonological clustering in Figure 3B. For our binary metric that classifies phonological similarity based on alliteration or rhyme, we observed a strong positive relation between phonological clustering and recall rate. $\beta = 0.080, CI = [0.033, 0.126]$, suggesting that phonological organization was beneficial to memory. For our continuous metric that considers all phonemes in the phonological similarity calculation and ignores order, we did not observe a discernible relation between phonological clustering and recall rate, $\beta = 0.012, CI = [-0.043, 0.068]$. Phonological clustering based on all phonemes of words did not promote overall recall, which aligned with the lack of phonological clustering by all phonemes of words. However, participants did exhibit phonological clustering using alliteration and rhyme, and this organization facilitated recall performance.

Phonological similarity interacts with temporal contiguity and semantic relatedness to accelerate recall transitions

Participants exhibited phonological clustering by the first and last two phonemes of words. Such organization of memory mirrors one half of the temporal contiguity and semantic similarity effects, with the other being that transitions between temporally contiguous (Kahana, 1996; Healey et al., 2019) and semantically related (Howard & Kahana, 2002; Sederberg et al., 2010) items occur with shorter latencies. Therefore, we hypothesized that phonological similarity accelerated recall transitions. Although we only found significant clustering based on alliteration and rhyme, we also found that all phonemes of words impact false memories, so we ran our inter-response time (IRT) analyses using both phonological similarity metrics.

Using our binary metric, for all transitions between correct recalls, we calculated the IRT and classified the phonological similarity of the words. We illustrate the participant-average IRTs as a function of phonological similarity and temporal lag (the serial position of recall $i+1$ minus the serial position of recall i) in Figure 4A. Using our continuous metric, for all transitions between correct recalls, we calculated the IRT and the phonological similarity of the words. Then, we binned transitions by phonological similarity, where the zero bin contains all transitions between words with zero phonological similarity (no shared phonemes) and bins one to five represent quintiles of the remaining transitions. We show the participant-average IRTs as a function of phonological similarity bin in Figure 4B. For both analyses, we omitted transitions with an IRT greater than 30 seconds, as we felt those instances to not represent a recall transition but rather a stopping and starting of recall.

However, for our statistical tests, we must also consider other factors that impact IRTs in free recall. We know temporal lag (Kahana, 1996; Healey et al., 2019) and semantic similarity (Howard & Kahana, 2002; Sederberg et al., 2010) modulate IRTs. Murdock and Okada (1970) and Rohrer and Wixted (1994) demonstrated that IRTs depend on the output position in the recall sequence together with the number of recalls on the trial. Mueller et al. (2003) included articulatory duration in their models when investigating phonological similarity in memory span tasks – the time it takes to say a recall provides a mechanical limitation on IRTs. Thus, we modeled IRTs as a function of phonological similarity, semantic similarity, temporal lag, output position, the number of correct recalls on the trial, and the number of phonemes in the first word of the transition (which served as a proxy for articulatory duration). We utilized a linear mixed effects model, so we linearized the IRT data by taking the natural logarithm of the IRTs and considered the absolute value of the temporal lag so the slope for lag goes in one direction. We also included all pairwise interactions between phonological similarity, semantic similarity, and absolute lag to understand how the organizational features of memory interacted with one another. We included random intercepts for participant, session, and list to account for individual differences and potential learning effects across lists and sessions. While the other factors have well understood impacts on IRTs, it was unknown whether phonological similarity had a universal effect on IRTs, so we also allowed the slope for phonological similarity to vary by participant. We completed this process for our two phonological clustering metrics separately, reporting results from the models in Table 3.

For our binary metric that classifies phonological similarity based on alliteration and rhyme, we did not observe a significant relation between phonological similarity and IRT, $\beta = -0.018, CI = [-0.036, 0.000]$. We did observe a significant interaction between phonological similarity and absolute lag, $\beta = -0.008, CI = [-0.010, -0.007]$. To better understand this interaction, we display the model predictions for IRTs as a function of phonological similarity and absolute lag in Figure 4C. The model predicted that phonological similarity does not impact recall transitions between temporally contiguous items – these transitions occurred quickly regardless of the phonological similarity label. However, for transitions between temporally distant items, phonological similarity generated faster IRTs.

For our continuous metric that considers all phonemes in the phonological similarity calculation, we did not find a discernible relation between phonological similarity and IRT, $\beta = -0.004, CI = [-0.047, 0.040]$. We did find significant interactions between phonological similarity and absolute lag, $\beta = -0.017, CI = [-0.020, -0.013]$, and between phonological and semantic similarity, $\beta = -0.122, CI = [-0.207, -0.038]$. We display the model predictions for IRTs as a function of phonological similarity and absolute lag in Figure 4E and phonological similarity and semantic similarity in Figure 4F. The phonological similarity bins consist of a zero bin with all transitions between words with no shared phonemes and five quintiles of the remaining transitions, and the six semantic similarity bins contain equal data. Again, the model predicted that phonological similarity speeds transitions only between temporally distant items. In contrast, the effect of phonological similarity on IRT was greater when the transition was between more semantically similar items, suggesting an additive effect of phonological and semantic similarity on quickening IRTs.

Discussion

In free recall, similarities among items govern the organization of memory, as seen in the order and timing of successively recalled words and in the characteristics of incorrect responses. Prior work has documented the influence of temporal, semantic, emotional, and spatial features on recall dynamics. In this paper, we asked whether phonological similarity also impacts the dynamics of free recall. Analyzing a large delayed free recall dataset, we found that participants clustered their recalls based on alliterative and rhyming words in the encoded lists. Moreover, the use of this phonological organization predicted recall performance. We also discovered that phonological features of the word lists and of participants' recalls induced false memories, and that phonological similarity interacted with temporal contiguity and semantic relatedness to accelerate recall transitions. Our results further exemplified how similarity, in its various forms, dictates the dynamics of free recall.

We believe our results complement the catalog of known similarity effects in free recall. First, we examined whether there exists a phonological clustering effect. Countless studies have demonstrated the temporal contiguity (Kahana, 1996; Healey et al., 2019) and semantic similarity (Howard & Kahana, 2002) effects in free recall, which Polyn et al. (2009) characterized with temporal and semantic clustering scores. Some past work has also shown emotional clustering effects (Long et al., 2015; Cohen & Kahana, 2022) and spatial navigation memory paradigms have elucidated spatial clustering effects (Miller et al., 2013) which Dougherty et al. (in press) quantified with a spatial clustering score. Consequently, we hypothesized there would exist significant phonological clustering, and we tested whether participants organized recalls based on alliterative (shared first phoneme) or rhyming (shared last two phonemes) words as well as whether participants attended to all phonemes of words. We found significant phonological clustering on the basis of the first and last two phonemes of words.

Having discovered a novel form of recall organization, a natural follow up was to ask if this organization related to recall performance. Sederberg et al. (2010) found that temporal clustering reliably predicted recall performance, and Healey et al. (2014) found that both temporal and semantic organization related positively to overall recall. We ran a linear mixed effects model to analyze the relation of phonological, temporal, and semantic clustering with recall performance at the list level and found significant positive relations of phonological (by our binary metric that attends to alliteration and rhyme), temporal, and semantic clustering with recall performance. Any sort of memory organization that our highly practiced participants utilized was beneficial to recall. There was not a significant relation between phonological clustering based on all phonemes of words and recall performance, which aligned with a lack of phonological clustering by this metric.

Similarity not only organizes, but also accelerates free recall. Transitions between temporally (Kahana, 1996; Healey et al., 2019), semantically (Howard & Kahana, 2002; Sederberg et al., 2010), and spatially (Dougherty et al., in press) similar recalls occur with shorter latencies. After accounting for numerous other factors in inter-response times (IRTs), we did not observe an accelerating effect of phonological similarity. However, we did find significant interactions of phonological similarity with temporal lag and semantic similarity. Linear mixed effects models predicted that phonological similarity, based on alliteration and rhyme or all phonemes, decreased IRTs between temporally distant items. Transitions between temporally contiguous items occurred quickly regardless, but phonological similarity sped up typically slow transitions between temporally distant items.

Moreover, using the continuous metric that calculates phonological similarity based on all phonemes, a linear mixed effects model predicted that phonological and semantic similarity had an additive effect on IRTs. Transitions between phonologically similar words occurred faster if they were also semantically similar.

We also feel our results complement prior analyses of phonological similarity effects in memory. Deffler et al. (2016) investigated misnamings of familiar individuals and found that incorrect names were often phonologically similar to the correct name. Sommers and Lewis (1999) and Chan et al. (2005) induced false recall and recognition of specific words that were phonologically similar to the studied words and Lim and Goh (2019) found greater false alarm rates for phonologically similar lures than unrelated lures. In these studies, the item retrieved from memory was a phonological confusion of the correct item. Likewise, we found that recall intrusions where phonologically similar to the words on the encoded list (would be correct recalls) based on alliteration and rhyme or all phonemes of words.

Several studies have examined the effect of lists of phonologically similar words on memory. Phonological similarity has been shown to be detrimental to performance in serial recall (Conrad, 1964; Baddeley, 1966b, 1966a, 1968; Spurgeon et al., 2014), memory span (Conrad, 1963; Conrad & Hull, 1964; Copeland & Radvansky, 2001; Mueller et al., 2003), and continuous recognition (Anisfeld, 1969) tasks. However, there exist several examples of phonological similarity facilitation in free recall. On 5-item lists, Fournet et al. (2003) observed improved free recall performance on phonologically similar lists, regardless of the length of the distractor period between encoding and retrieval. On 12-word lists, Watkins et al. (1974) found greater recall of pre-recency items on phonologically similar lists. On 20-item lists containing sections of neighboring phonologically similar words in the middle or at the end of the list, Craik and Levy (1970) saw increased recall probability for the phonologically similar items. Using free recall scoring on immediate serial recall data, Gupta et al. (2005) found greater recall performance on alliterative and rhyming lists than dissimilar lists, provided words were not repeated across lists. While these studies examined recall probability as a function of phonological list construction, we explicitly modeled recall rates as a function of phonological clustering of recalls on lists of words with no explicit phonological specification. We found that phonological clustering based on alliteration and rhyme related positively with recall performance, providing novel evidence that attending to phonological features of words can aid memory in free recall.

We found that participants exhibited phonological clustering by the first and last two phonemes of words, but not by all phonemes of words, which leads to the question of why did only the first and last two phonemes drive memory organization? Primacy and recency effects (Murdock, 1962) are fundamental principles of free recall. However, unlike the encoding of lists which typically occurs one item at a time, we do not imagine the encoding of words happened one phoneme at a time. Although, Cleary, Winfield, and Kostic (2007) had participants listen to words and then attempt to identify the words with phonemes removed and rate the likelihood the words were studied. Even for unidentified items, participants gave higher recognition ratings to studied items than non-studied items, suggesting phonemes are isolable features in memories of words. Moreover, alliteration and rhyme appear ubiquitously in literature and music, providing credence to the notion that similar sounds at the start or end of words are particularly salient. In studies that used qualitative measures of phonological similarity, it is no surprise that the phonologically similar conditions often contained rhyming or alliterative words (Conrad, 1963; Baddeley, 1966b, 1966a, 1968; Copeland & Radvansky, 2001; Chow et al., 2016; Cleary, 2004; Fournet et al., 2003; Gupta et al., 2005; Nairne & Kelley, 1999).

We also recognize that our study involved visual presentation of the stimuli, meaning our analyses of the influence of the sound of words on memory consider data in which participants do not actually hear the words. Several studies have demonstrated modality effects in free recall (Pazdera & Kahana, 2023; Murdock, 1968; Murdock & Walker, 1969; Craik, 1969) and any phonological similarity effects would seem prime candidates to be modality-dependent. Watkins et al. (1974) had participants read lists of phonologically similar or unrelated words either silently or aloud and found that phonologically similar lists eliminated the typical modality effect of better recall for recency items with auditory encoding. Lim and Goh (2019) did not observe modality differences between visual and auditory presentation for recognition performance on studied items or lures that were not phonologically related to encoded items, but found more false alarms for phonologically similar lures in the auditory condition.

However, we argue that visual presentation importantly removes the possibility of perceptual errors becoming entangled with memory errors, a concern Baddeley (1966b) raised. McDermott and Watson (2001) demonstrated the importance of perception, presenting lists of words phonologically similar to a specific item with presentation durations ranging from 20 to 5000 milliseconds. The probability of false recall of the critical item decreased with longer presentation durations that mitigated improper perception of the studied items. With auditory presentation, it would be possible that an incorrect recall that was phonologically similar to a presented word did represent successful retrieval, but of an item that was incorrectly perceived and thereby wrongly encoded. As such, the visual encoding in our study renders our findings that phonological similarity induces false memories even more noteworthy.

In free recall, associative similarities among items govern the patterns of memory retrieval. Temporal contiguity and semantic relatedness stand as laws of memory, organizing the order of recalls and accelerating recall transitions. Evidence of recall clustering by emotional, spatial, and task features further demonstrates the influence of similarity on recall dynamics. In this paper, we investigated phonological similarity as another potential governing factor in the dynamics of free recall. We found that participants clustered their recalls based on alliterative and rhyming features of the studied words, and that the use of this phonological organization facilitated recall performance. We also discovered that phonological similarity, based on alliteration and rhyme or all phonemes of words, induced false memories and interacted with temporal and semantic similarity to accelerate recall transitions. Our work further catalogs how similarity, in its various forms, dictates the dynamics of free recall.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*, 6.

Anisfeld, M. (1969). False recognition produced by semantic and phonetic relations under two presentation rates. *Psychonomic Science*, *17*(6), 366-&.

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

Baddeley, A. D. (1966b). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, *18*(4), 362-365. doi: 10.1080/14640746608400055

Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *Quarterly Journal of Experimental Psychology*, *20*(3), 249-264. doi: 10.1080/14640746808400159

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. *Journal of Royal Statistical Society, Series B*, *57*, 289-300.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python.* O'Reilly Media Inc.

Chan, J. C. K., McDermott, K. B., Watson, J. M., & Gallo, D. A. (2005). The importance of material-processing interactions in inducing false memories. *Memory & Cognition*, *33*(3), 389–395. doi: 10.3758/BF03193057

Chomsky, N., & Hall, M. (1968). *The sound pattern of english.* New York: Harper & Row.

Chow, M., Macnamara, B. N., & Conway, A. R. A. (2016). Phonological similarity in working memory span tasks. *Memory & Cognition*, *44*(6), 937–949. doi: https://doi.org/10.3758/s13421-016-0609-8

Cleary, A. M. (2004). Orthography, phonology, and meaning: Word features that give rise to feelings of familiarity in recognition. *Psychonomic Bulletin & Review*, *11*(3), 446–451. doi: 10.3758/BF03196593

Cleary, A. M., Winfield, M. M., & Kostic, B. (2007). Auditory recognition without identification. *Memory & Cognition*, *35*(8), 1869–1877. doi: 10.3758/BF03192921

Cohen, R. T., & Kahana, M. J. (2022). A memory based theory of emotional disorders. *Psychological Review*, *129*(4), 742-776. doi: 10.1037/rev0000334

Conrad, R. (1963). Acoustic confusions and memory span for words. *Nature*, *197*(4871), 1029–1030. doi: 10.1038/1971029a0

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, *55*(1), 75-84. doi: https://doi.org/10.1111/j.2044-8295.1964.tb00899.x

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, *55*, 429-432.

Copeland, D. E., & Radvansky, G. A. (2001). Phonological similarity in working memory. *Memory & Cognition*, *29*(5), 774–776. doi: 10.3758/BF03200480

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., & Levy, B. A. (1970). Semantic and acoustic information in primary memory. *Journal of Experimental Psychology*, *86*, 77-82.

Deffler, S. A., Fox, C., Ogle, C. M., & Rubin, D. C. (2016). All my children: The roles of semantic category and phonetic similarity in the misnaming of familiar individuals. *Memory & Cognition*, *44*(7), 989–999. doi: 10.3758/s13421-016-0613-z

Dougherty, M. R., Chang, W., Rudoler, J. H., Katerman, B. S., Halpern, D. J., Bruska, J. P., . . . Kahana, M. J. (in press). Neural correlates of memory in a naturalistic spatiotemporal context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: https://doi.org/10.1101/2022.11.30.518606

Fournet, N., Juphard, A., Monnier, C., & Roulin, J.-L. (2003). Phonological similarity in free and serial recall: The effect of increasing retention intervals. *International*

Journal of Psychology, *38*(6), 384-389. doi: https://doi.org/10.1080/00207590344000204

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: The roles of type of similarity, category cuing, and item recall. *Memory & Cognition*, *33*(6), 1001–1016. doi: 10.3758/BF03193208

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. (2014). Is memory search governed by universal principles or idiosyncratic strategies? *Journal of Experimental Psychology: General*, *143*(2), 575–596. doi: 10.1037/a0033715

Healey, M. K., Long, N. M., & Kahana, M. J. (2019). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, *26*(3), 699–720. doi: 10.3758/s13423-018-1537-3

Herz, N., Bukala, B. R., Kragel, J. E., & Kahana, M. J. (2023). Hippocampal activity predicts contextual misattribution of false memories. *Proceedings of the National Academy of Sciences*, *120*(40), e2305292120. Retrieved from `https://www.pnas.org/doi/abs/10.1073/pnas.2305292120` doi: 10.1073/pnas.2305292120

Howard, M. W., & Kahana, M. J. (2002). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, *46*(1), 85–98. doi: 10.1006/jmla.2001.2798

Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New Phytologist*, *11*(2), 37-50. doi: https://doi.org/10.1111/j.1469-8137.1912.tb05611.x

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

Kahana, M. J., Diamond, N. B., & Aka, A. (2024). Laws of human memory. In M. J. Kahana & A. D. Wagner (Eds.), *Oxford handbook of human memory* (Vol. 1).

Oxford University Press.

Kahana, M. J., Dolan, E. D., Sauder, C. L., & Wingfield, A. (2005). Intrusions in episodic recall: Age differences in editing of overt responses. *Journal of Gerontology: Psychological Sciences*, *60*(2), 92–97. doi: 10.1093/geronb/60.2.P92

Kahana, M. J., Lohnas, L. J., Healey, M. K., Aka, A., Broitman, A. W., Crutchley, P., ... Weidemann, C. T. (2024). The penn electrophysiology of encoding and retrieval study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: 10.1037/xlm0001319

Kintsch, W., & Buschke, H. (1969). Homophones and synonyms in short-term memory. *Journal of Experimental Psychology*, *80*(3, Pt.1), 403–407. doi: 10.1037/h0027477

Li, P., & MacWhinney, B. (2002). Patpho: A phonological pattern generator for neural networks. *Behavior Research Methods, Instruments, & Computers*, *34*(3), 408–415.

Lim, L. C., & Goh, W. D. (2019). False recognition modality effects in short-term memory: Reversing the auditory advantage. *Cognition*, *193*, 104008. doi: https://doi.org/10.1016/j.cognition.2019.104008

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490. doi: 10.3758/BF03210951

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

Luo, C. R., Johnson, R. A., & Gallo, D. A. (1998). Automatic activation of phonological information in reading: Evidence from the semantic relatedness decision task. *Memory & Cognition*, *26*(4), 833–843. doi: 10.3758/BF03211402

McDermott, K. B., & Watson, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language*, *45*(1), 160-176. doi: https://doi.org/10.1006/jmla.2000.2771

Miller, J. F., Lazarus, E., Polyn, S. M., & Kahana, M. J. (2013). Spatial clustering during memory search. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *39*(3), 773–781.

Mueller, S. T., Seymour, T. L., Kieras, D. E., & Meyer, D. E. (2003). Theoretical implications of articulatory duration, phonological similarity, and phonological complexity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1353–1380. doi: 10.1037/0278-7393.29.6.1353

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

Murdock, B. B. (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., & Okada, R. (1970). Interresponse times in single- trial free recall. *Journal of Verbal Learning and Verbal Behavior*, *86*(2), 263–267. doi: 10.1037/h0029993

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

Nairne, J. S., & Kelley, M. R. (1999). Reversing the phonological similarity effect. *Memory & Cognition*, *27*(1), 45–53. doi: 10.3758/BF03201212

Pazdera, J. K., & Kahana, M. J. (2023). Modality effects in free recall: A retrieved-context account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(6), 866-888. doi: https://psycnet.apa.org/doi/10.1037/xlm0001140

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

Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition*, *22*(5), 511–524. doi: 10.3758/BF03198390

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6 (2)*, 461-464.

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

Sommers, M. S., & Lewis, B. R. (1999). Who really lives next door: creating false memories with phonological neighbors. *Journal of Memory and Language*, *40*, 83-108.

Spurgeon, J., Ward, G., & Matthews, W. J. (2014). Examining the relationship between immediate serial recall and immediate free recall: Common effects of phonological loop variables but only limited evidence for the phonological loop. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1110.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer.* (pp. 237–249). Washington, DC: American Psychological Association. doi: 10.1037/10895-018

Vitz, P. C., & Winkler, B. S. (1973). Predicting the judged "similarity of sound" of english words. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 373-388. doi: https://doi.org/10.1016/S0022-5371(73)80016-7

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

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

Table 1

Phonological similarity metrics. Phoneme breakdowns and phonological similarity scores for the binary metric that classifies alliterative (same first phoneme) and rhyming (same last two phonemes) words as phonologically similar and the continuous metric that calculates phonological similarity as the ratio of the size of the intersection and union of the sets of phonemes. Bold indicates shared first or last phonemes, italics indicates intersection of phoneme sets.

Word	Shark	Shoulder	Palace	Waitress	Doughnut	Notebook
Phonemes	{**SH**, AA1, R, K}	{**SH**, OW1, L, D, ER0}	{P, AE1, L, **AH0**, **S**}	{W, EY1, T, R, **AH0**, **S**}	{D, *OW1*, *N*, AH2, *T*}	{*N*, *OW1*, *T*, B, UH2, K}
Alliteration/Rhyme	1		1		0	
Jaccard	0.125		0.222		0.375	

Table 2

Recall probability as a function of recall clustering. Slope estimates and 95% confidence intervals from linear mixed effects models indicate that clustering recalls by alliteration and rhyme (binary) relates positively with recall performance, but organization based on all phonemes (continuous) of words does not.

	Binary		Continuous	
Effect	Slope	95% CI	Slope	95% CI
pcs	0.080	[0.033, 0.126]	0.012	[-0.044, 0.068]
scs	0.116	[0.073, 0.159]	0.093	[0.043, 0.143]
tcs	0.217	[0.152, 0.282]	0.226	[0.156, 0.295]
pcs:scs	-0.118	[-0.170, -0.065]	0.045	[-0.017, 0.106]
pcs:tcs	-0.026	[-0.074, 0.022]	-0.027	[-0.084, 0.029]
scs:tcs	-0.100	[-0.154, -0.045]	-0.096	[-0.148, -0.045]

Table 3

Inter-response time (IRT) modeling. Slope estimates and 95% confidence intervals from linear mixed effects models do not discern a main effect of phonological similarity on IRTs by the binary (alliteration/rhyme) nor continuous (all phonemes) metrics. However, the models do find significant interactions of phonological similarity and absolute temporal lag using the binary and continuous metrics and phonological similarity and semantic similarity using the continuous metric.

	Binary		Continuous	
Effect	Slope	95% CI	Slope	95% CI
psim	-0.018	[-0.036, 0.000]	-0.004	[-0.047, 0.040]
ssim	-0.139	[-0.153, -0.126]	-0.129	[-0.145, -0.113]
abs_lag	0.053	[0.053, 0.054]	0.055	[0.054, 0.056]
outpos	0.040	[0.039, 0.040]	0.040	[0.039, 0.040]
ncr	-0.025	[-0.025, -0.024]	-0.025	[-0.025, -0.024]
len_ph	0.028	[0.026, 0.029]	0.029	[0.027, 0.031]
psim:ssim	-0.022	[-0.056, 0.013]	-0.122	[-0.207, -0.038]
psim:abs_lag	-0.008	[-0.010, -0.007]	-0.017	[-0.020, -0.013]
ssim:abs_lag	-0.016	[-0.018, -0.014]	-0.017	[-0.019, -0.015]

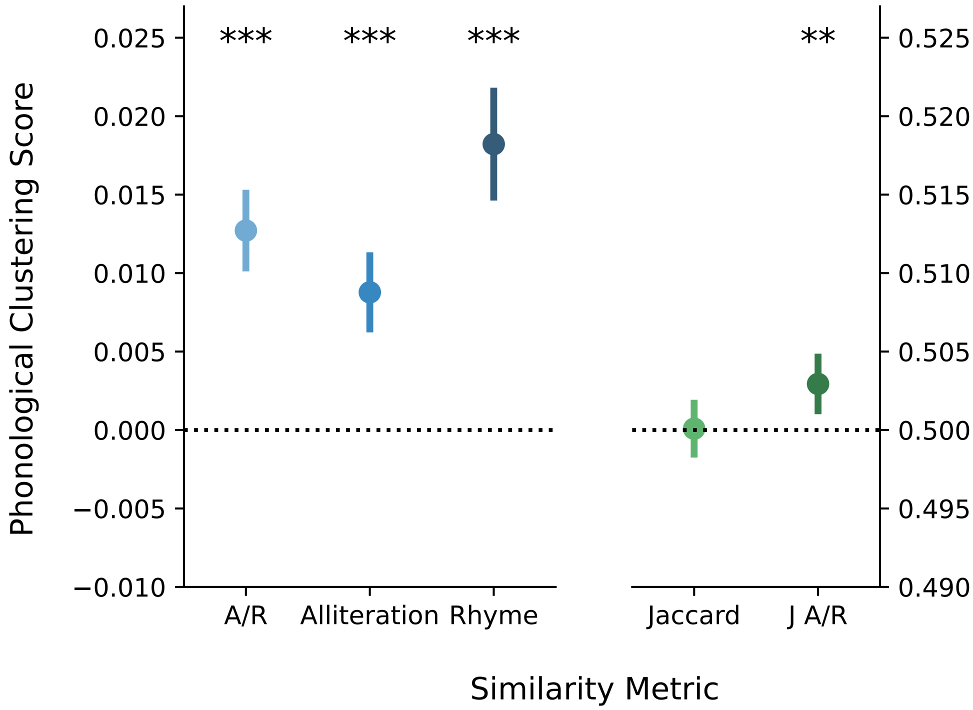

Figure 1

*Significant phonological clustering by alliteration and rhyme. Phonological clustering scores calculated via the binary metric of phonological similarity, based on alliteration or rhyme (A/R), demonstrate that participants organized recall using phonological features of the studied words. The effect remains when considering just the first (Alliteration) or just the last two (Rhyme) phonemes. Using a continuous metric that considers all phonemes of words via the ratio of the intersection and union of phoneme sets (Jaccard), the phonological clustering effect disappeared. However, using the Jaccard index only on the first and last two phonemes of words (J A/R), significant phonological organization reemerged. Error bars represent 95% confidence intervals. *** represents $p < 0.001$, ** represents $p < 0.01$, * represents $p < 0.05$.*

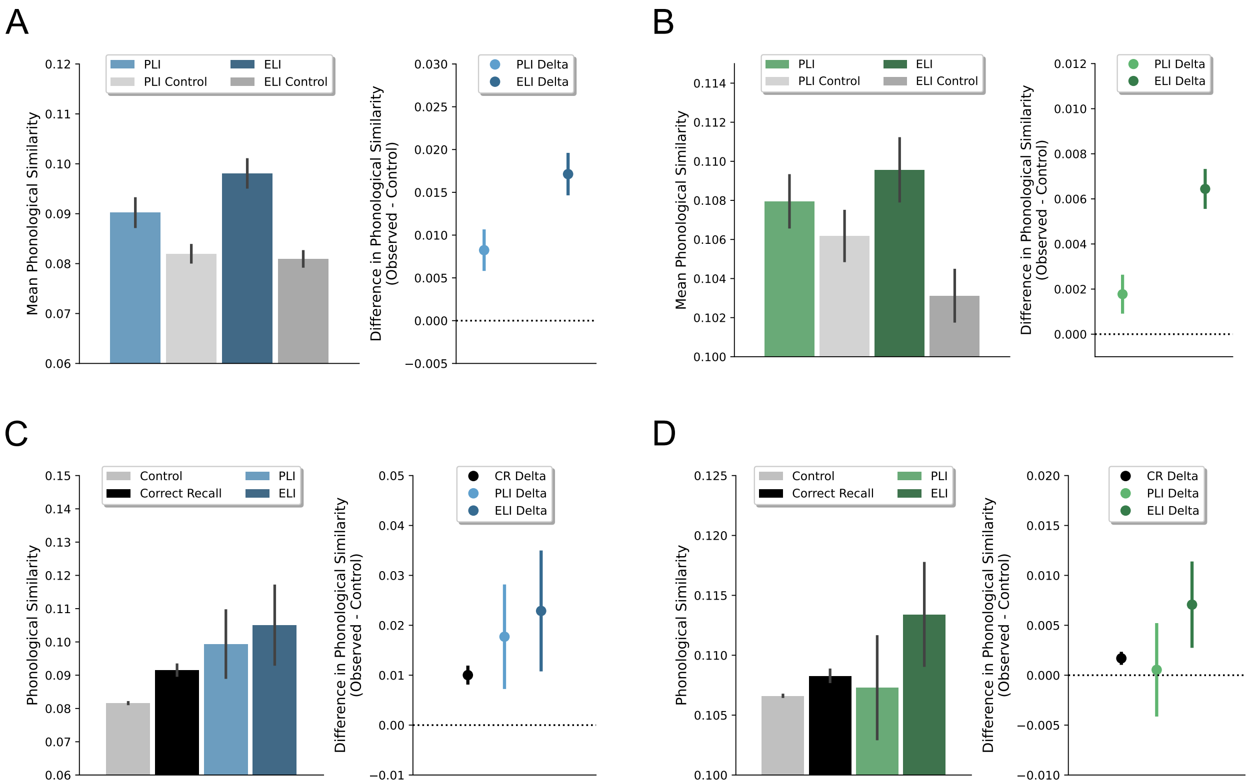

Figure 2

Phonological features induce intrusions. Prior-list (PLI) and extra-list (ELI) intrusions were more phonologically similar to words on the current list than other words in the wordpool using both the binary (alliteration/rhyme) (A) and continuous (all phonemes) (B) measures of phonological similarity. Phonological similarity between correct recalls, PLIs, and ELIs and the preceding correct recall for the binary (C) and continuous (D) measures of phonological similarity. Error bars represent 95% confidence intervals.

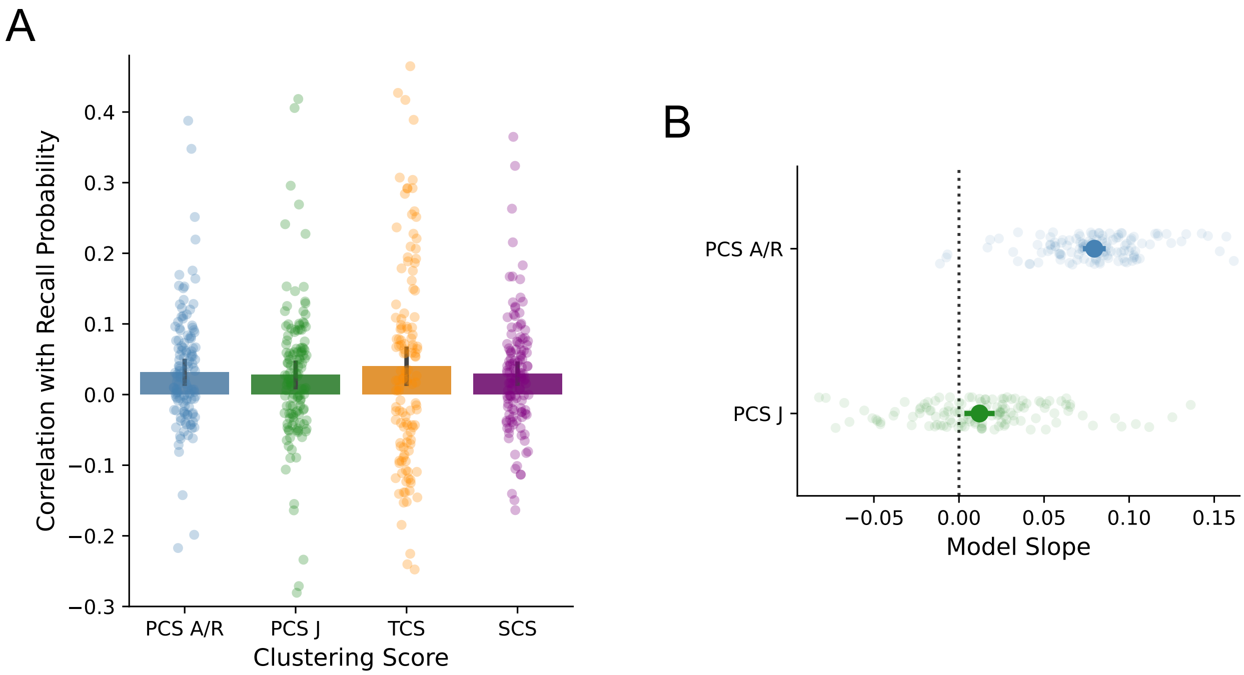

Figure 3

Phonological clustering promotes memory performance. (A) Participant-average list-level correlations between phonological (alliteration/rhyme = PCS A/R, all phonemes = PCS J), temporal (TCS), and semantic (SCS) clustering and recall rates. Dots represent individual participant scores. (B) Participant-level slopes from linear mixed-effects models illustrate that phonological organization by alliteration or rhyme (first or last two phonemes of words) related positively with recall probability. Error bars represent 95% confidence intervals.

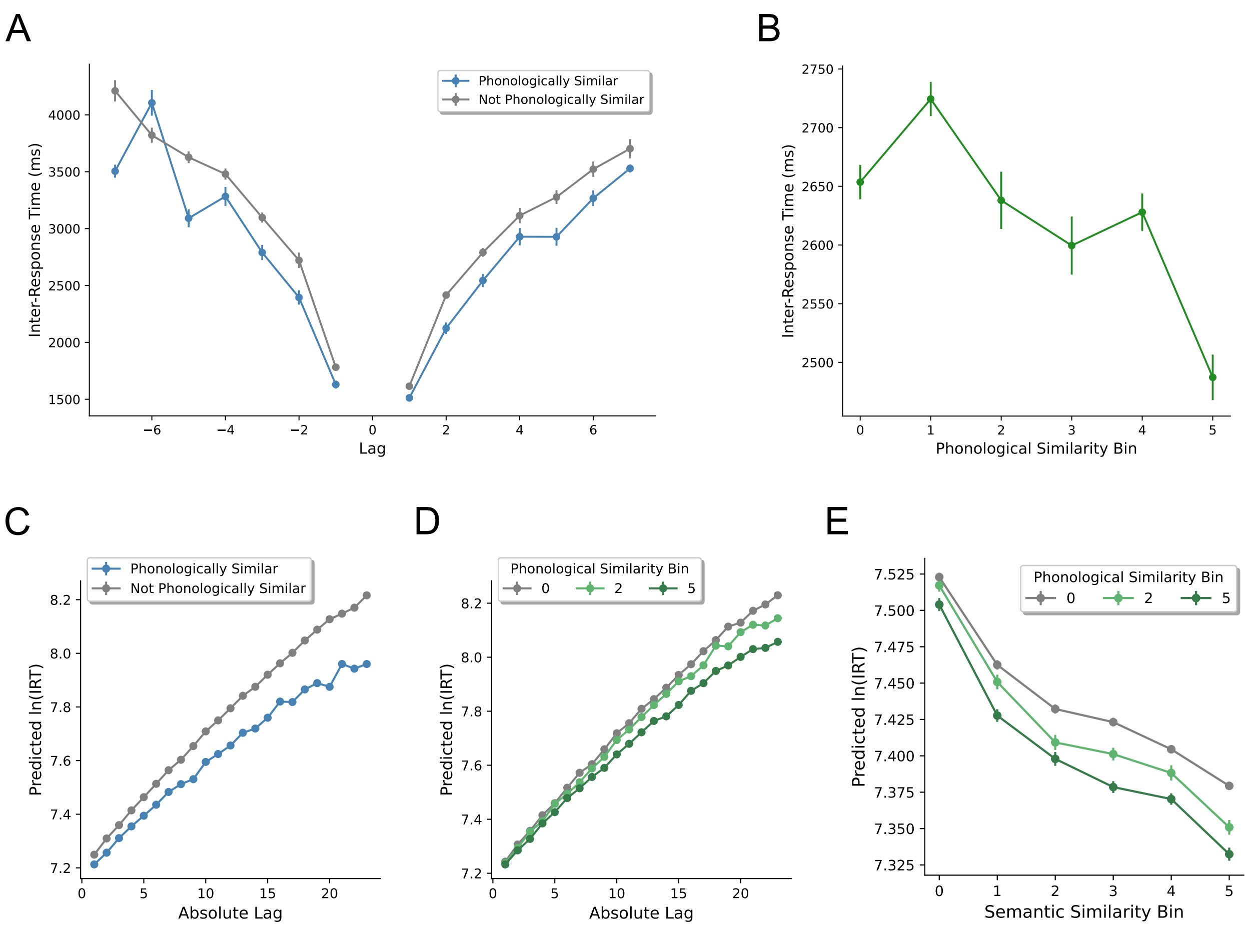

Figure 4

Phonological similarity interacts with temporal lag and semantic similarity to accelerate inter-response times (IRTs). (A) IRT as a function of temporal lag at encoding and phonological similarity using the binary metric that classifies alliterative or rhyming words as phonologically similar. (B) IRT as a function of phonological similarity using the continuous metric that calculates the ratio of the intersection and union of all phonemes of words. The zero bin contains all transitions between words with no shared phonemes, and bins one to five represent quintiles of the remaining transitions. Linear mixed effects models predicted that phonological similarity accelerated transitions between temporally distant (C, D) items for both the binary and continuous metrics of phonological similarity. Using the continuous metric, a linear mixed effects model also predicted that phonological similarity accelerated transitions between semantically similar (E) items. (D) and (E) only depict three phonological similarity bins for clarity. Error bars represent Loftus-Masson confidence intervals (Loftus & Masson, 1994).

Supplemental Materials for "Phonological Similarity and Recall Dynamics"

Haydn G. Herrema and Michael J. Kahana

University of Pennsylvania

Supplementary Tables and Figures

Table S1

AIC and BIC for linear mixed effects models predicting recall probability as a function of phonological, temporal, and semantic clustering. Both models also included all pairwise interactions between phonological, temporal, and semantic clustering.

		Binary		Continuous	
Model Random Effects	Degrees of Freedom	AIC	BIC	AIC	BIC
(1 | participant) + (1 | session) + (1 | list)	11	-42432.64	-42334.73	-42055.31	-41957.29
(1 + pcs + scs + tcs | participant) + (1 | session) + (1 | list)	20	-44736.40	-44558.38	-44283.15	-44104.93

Table S2

AIC and BIC for linear mixed effects models predicting inter-response times as a function of phonological similarity, absolute temporal lag, semantic similarity, output position, number of correct recalls, and length of the first word in phonemes. Both models also included all pairwise interactions between phonological similarity, absolute temporal lag, and semantic similarity.

		Binary		Continuous	
Model Random Effects	Degrees of Freedom	AIC	BIC	AIC	BIC
(1 | participant) + (1 | session) + (1 | list)	14	1296761.38	1296919.52	1296934.88	1297093.02
(1 + psim | participant) + (1 | session) + (1 | list)	16	1296703.53	1296884.26	1296899.79	1297080.52

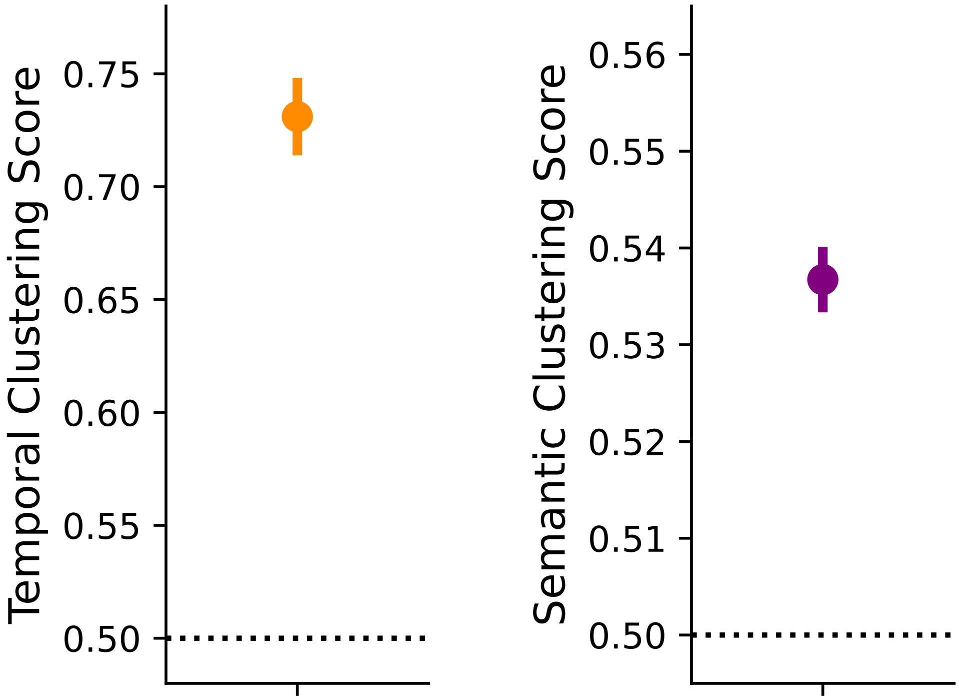

Figure S1

Participant-average temporal and semantic clustering scores illustrate that temporal contiguity and semantic similarity organize recall sequences. Error bars represent 95% confidence intervals.

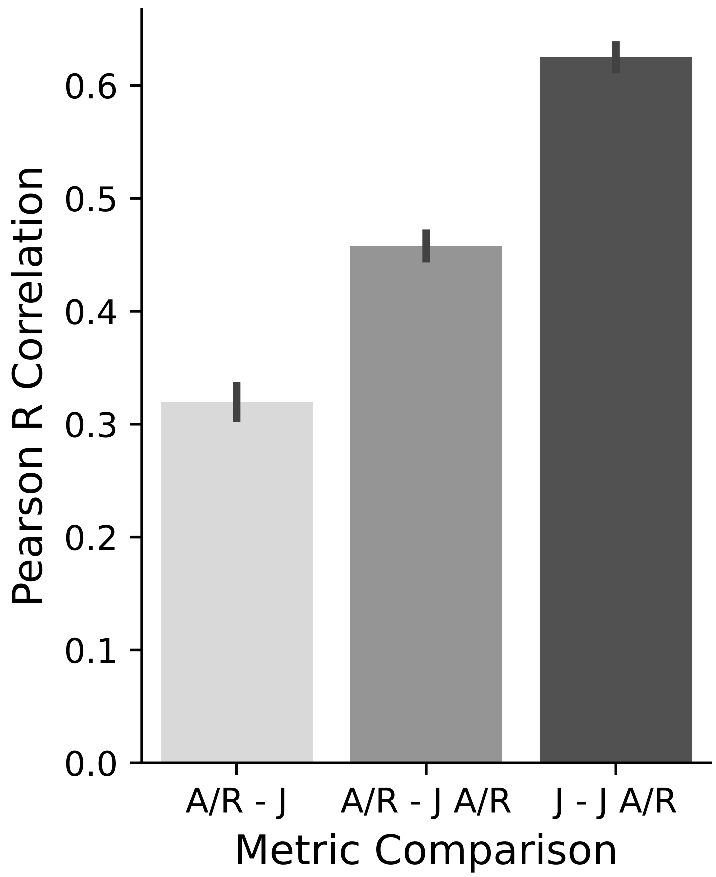

Figure S2

Participant-average correlations of phonological clustering scores calculated via the binary metric of phonological similarity, based on alliteration and rhyme (A/R), the continuous metric that considered all phonemes of words via the ratio of the intersection and union of phoneme sets (J), and the Jaccard index on only the first and last two phonemes of words (J A/R). Error bars represent 95% confidence intervals.