

Conditional recall and the frequency effect in the serial recall task: an examination of item-to-item associativity

Leonie M. Miller · Steven Roodenrys

Published online: 13 June 2012
© Psychonomic Society, Inc. 2012

Abstract The frequency effect in short-term serial recall is influenced by the composition of lists. In pure lists, a robust advantage in the recall of high-frequency (HF) words is observed, yet in alternating mixed lists, HF and low-frequency (LF) words are recalled equally well. It has been argued that the preexisting associations between all list items determine a single, global level of supportive activation that assists item recall. Preexisting associations between items are assumed to be a function of language co-occurrence; HF–HF associations are high, LF–LF associations are low, and mixed associations are intermediate in activation strength. This account, however, is based on results when alternating lists with equal numbers of HF and LF words were used. It is possible that directional association between adjacent list items is responsible for the recall patterns reported. In the present experiment, the recall of three forms of mixed lists—those with equal numbers of HF and LF items and pure lists—was examined to test the extent to which item-to-item associations are present in serial recall. Furthermore, conditional probabilities were used to examine more closely the evidence for a contribution, since correct-in-position scoring may mask recall that is dependent on the recall of prior items. The results suggest that an item-to-item effect is clearly present for early but not late list items, and they implicate an additional factor, perhaps the availability of resources at output, in the recall of late list items.

Keywords Short term memory · Language production · Recall

L. M. Miller (✉) · S. Roodenrys
School of Psychology, University of Wollongong,
Wollongong, NSW 2522, Australia
e-mail: leoniem@uow.edu.au

In the context of immediate serial recall, the frequency effect is complex and difficult to explain in straightforward terms (e.g., Hulme, Stuart, Brown, & Morin, 2003; Morin, Poirier, Fortin, & Hulme, 2006). While tasks involving pure lists of either only high-frequency (HF) or only low-frequency (LF) words have shown a clear advantage for HF words across many replications (Allen & Hulme, 2006; Hulme et al., 1997; Hulme et al., 2003; Morin et al., 2006; Poirier & Saint-Aubin, 1996; Stuart & Hulme, 2000; Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002; Roodenrys & Quinlan, 2000; Saint-Aubin & LeBlanc, 2005; Saint-Aubin & Poirier, 2005; Watkins & Watkins, 1977), the results of experiments using mixed lists, where HF and LF words appear together on the same trial, indicate that under some conditions, LF items can be recalled as well as HF words (Hulme et al., 2003; Morin et al., 2006; Saint-Aubin & LeBlanc, 2005).

Early explanations of the frequency effect in serial recall, motivated by the results from experiments using pure lists, focused on item-specific differences between HF and LF words, consistent with the prevailing explanation of verbal short-term memory (STM) phenomena. For example, the phonological loop (Baddeley & Hitch, 1974) was characterized as a serially ordered, speech-based system that was responsible for performance on memory span and serial recall tasks (Baddeley, 1986). It comprised a phonological short-term store (STS) that retained the short-term traces of items and a subvocal rehearsal mechanism. Traces in the STS were subject to passive decay and consequent degradation unless refreshed by rehearsal in the loop. Wright (1979) demonstrated articulation rate differences between HF and LF words; HF items of the same length are articulated faster than LF words. Accordingly, the frequency effect will manifest from rehearsal rate differences between HF and LF words, since greater rehearsal efficiency leads to superior retention of the short-term traces on which recall is reliant.

However, when the effect in pure lists was tested under conditions preventing rehearsal (Gregg, Freedman, & Smith, 1989; Tehan & Humphreys, 1988) or when differences in articulation rate were statistically accounted for (Hulme et al., 1997; Hulme et al., 2003; Stuart & Hulme, 2000), a frequency effect persisted, implicating a second source of effect. Hulme et al. (1997) nominated phonological long-term memory (LTM) as this source; a second-stage process utilizing available long-term phonological traces, redintegration (Schweickert, 1993), would restore degraded short-term traces through pattern matching and completion. It was argued that such traces were more accessible for HF than for LF words and, consequently, HF items were more likely to be successfully redintegrated. Notably, in this account, differences in recall were attributed to another item-specific property—namely, access to phonological long-term representations.

More recently, the notion that item-specific properties form the basis of the frequency effect was challenged by a series of experiments demonstrating list-dependent influences on the recall of HF and LF words (Hulme et al., 2003; Morin et al., 2006; Saint-Aubin & LeBlanc, 2005; Stuart & Hulme, 2000). Stuart and Hulme investigated whether the manipulation of preexperimental association between HF and LF items affected recall performance. Their study did not “mix” HF and LF items within the same list but compared recall for lists of items that had been familiarized by pairwise association during a training period with that for lists of items familiarized individually. Stuart and Hulme observed a benefit to the recall of pure LF lists after familiarization that did not occur for pure HF lists. They argued that familiarization per se did not result in better recall, while LF lists constructed from the same subset of items—that is, lists containing pairwise familiarized items—were recalled better than the alternating lists where adjacent items were not from the same familiarization pool.¹ Consequently, the authors argued that the frequency effect could be explained entirely in terms of associative links between items. Instead of the frequency effect being driven by the accessibility of individual item representations in LTM, they suggested that it might be an outcome of the preexperimental associations formed from the co-occurrence of items in natural language (Deese, 1960). The representations of items would form a “mutually supporting network of item nodes” (p. 801) facilitating the accessibility of each item’s long-term trace in the retrieval process; that is, the associative links would mutually excite connected list members and

determine the accessibility of these LTM representations (Saint-Aubin & Poirier, 2005).

Hulme et al. (2003) showed that in alternating lists of HF and LF words, the frequency effect is eliminated (Experiment 1) or much reduced (Experiment 2). The pattern of item errors mirrored the serial recall performance across the list types; the pure list advantage for HF items was nullified (Experiment 1) or unreliably reversed (Experiment 2) in mixed lists. It was argued that such outcomes were further support for the influence of preexperimental associations between list items in redintegration (Stuart & Hulme, 2000). Furthermore, Morin et al. (2006) identified that the abolition of the frequency effect with alternating lists was not an outcome of task awareness, since it occurred under both incidental and intentional learning conditions. Accordingly, Morin et al. (2006) argued that the STM processes responsible for the nature of the frequency effect are not consciously mediated and nominated Hulme et al.’s (2003; Stuart & Hulme, 2000) redintegration hypothesis as the preferred explanation.

Nonetheless, Hulme et al. (2003) referred to a model of episodic memory, the temporal context model (TCM; Howard & Kahana, 2002a, 2002b), as an alternate conceptual framework suggesting how associative effects might underlie the frequency effect. In this model, item recall is dependent upon the reinstatement of the temporal context at encoding, and recall of an item then acts as a cue to establish the temporal context for the next item to be recalled, and so on. The encoded temporal contexts are determined by a slowly evolving context state that is combined with former contexts in which an item has been encoded. Hence, items presented close together in time will have some overlap in their contextual states, and historical co-occurrence of items, as in the case of strong semantic associates, will enhance the degree of contextual overlap. Accordingly, TCM states that interitem associations drive the forward order of recall and the stronger the semantic association between successive items, the greater the likelihood that the first item in a pair will facilitate the retrieval of the second. Furthermore, it suggests that effects of interitem semantic association should be localized in the associative strengths of consecutive pairs of list items.

This contrasts with the position taken by Hulme et al. (2003; Stuart & Hulme, 2000), where the interitem associations of all list items were seen to operate in a mutually supporting and nondirectional manner, facilitating their retrieval. The alternating lists used by Hulme et al. (2003) did not provide an adequate test of these positions, since frequency-wise alternation would yield intermediate levels of associative strength between consecutive pairs of items across the list (HF item to LF item or vice versa) and intermediate association strength at the list level (since each mixed list contained three HF and three LF items). The question of

¹ The manipulation of familiarity used by Stuart and Hulme (2000) has been called into question by Saint-Aubin and Poirier (2005), who proposed that familiarity was confounded with set size in this experiment. They found a comparable effect of familiarity when items are familiarized alone and argued against associativity as the mechanism responsible for the frequency effect.

whether associative effects are directional or nondirectional remained unresolved.

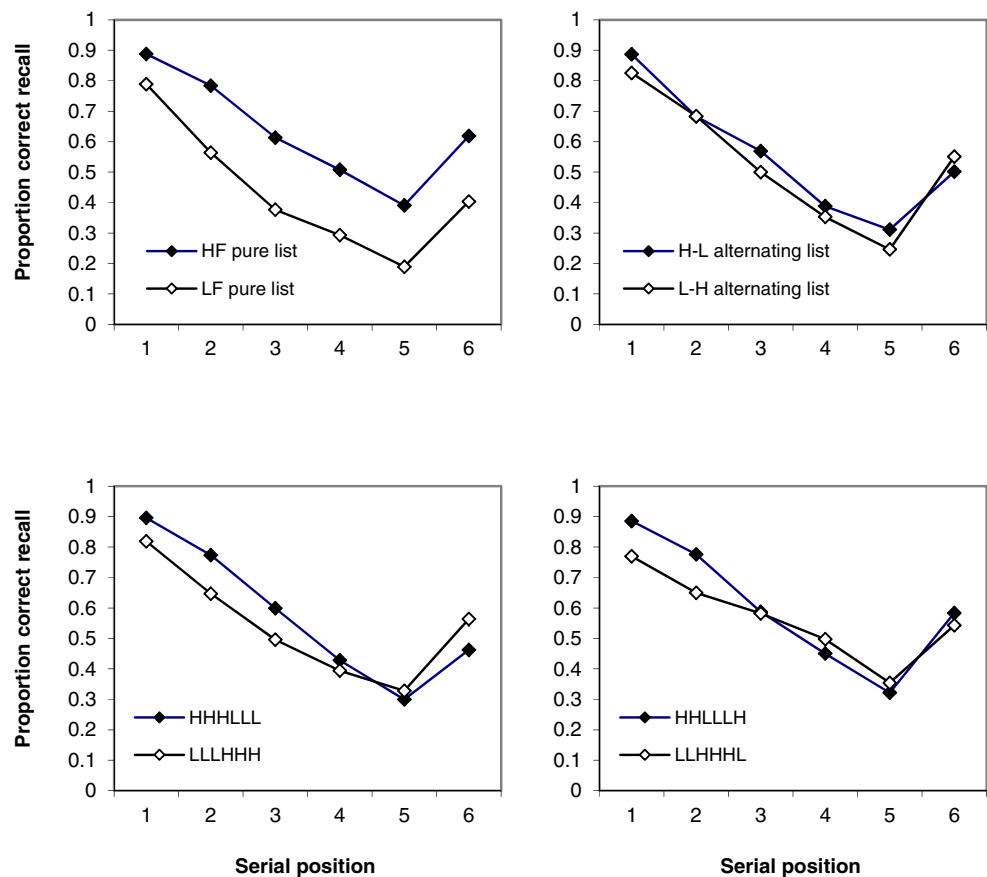
Recently, Miller and Roodenrys (2012) presented the results of experiments using a *half-list* manipulation originally proposed by Hulme et al. (2003) as a clear test of the directionality of associative effects in lists of mixed frequency. Half lists are constructed from the halves of pure lists so that HF and LF items are presented in sequence (HHHLLL and LLLHHH) and match the list composition of alternating lists with three HF and three LF items each. A sample was tested on the recall of pure and half lists; it was observed that half lists mimicked the performance with pure lists in the first half of the list, while recall for half lists in the second half did not differ statistically between list types and was intermediate with respect to performance with pure lists (except for a possible difference in the recency position). In descriptive terms, HHHLLL lists behaved like HF pure lists, and LLLHHH lists matched recall for pure LF lists for the first three serial positions. Recall performance for half lists intersected between positions 4 and 5, well after the change in list composition (see Fig. 1 for a replication of this feature). Clearly, these results are not consistent with a nondirectional associative redintegration explanation, since it predicts that serial position curves for half lists should be the same as those found with alternating lists (Hulme et al., 2003). In contrast, a

directional mechanism might predict an abrupt shift in recall at position 4 in the half lists, but this was not observed. Accordingly, it was proposed that recall was unlikely to be dictated by purely directional associative influences.

However, the lag in the crossover in half-list recall with the change in list composition is not necessarily inconsistent with the notion of interitem association that operates between successive items. While changes in interitem association would be expected to alter the rate of change in correct recall between items, the absolute difference between serial position curves prior to a change in list composition might dictate how far beyond a change in list composition the crossover actually occurs. Furthermore, when absolute recall levels are similar, as occurs at crossover points, the capacity of item-to-item associativity to generate significant effects between lists in subsequent positions might be limited.

The present work sought to further clarify the extent to which directional associations influence the recall of list items, in two ways. First, a between-subjects design comparing the within-subjects serial recall of two list types included pure, alternating, and half lists, and a new format—*sequence* lists (HHLLLH or LLHHHL)—was used to examine in greater detail the serial recall behavior in lists where a sequence of items of the same frequency type occurred after an early transition between HF and LF words. The sequence list format

Fig. 1 Serial recall of words as a function of frequency and list composition. Panels represent the between-subjects conditions



is a variation on the half-list composition, where the sequence of same-frequency items in the second half of the list is brought forward by one serial position and the displaced item type from the first half sequence is moved to the last serial position. This format maintains the same composition as the other mixed lists formats (three HF and three LF items) and has the added benefit of deconfounding directional associative effects with any other influences arising from the coincidence of the last item in sequence with the recency position, as occurs for half lists.

Second, this experiment sought to explore more closely recall at points of transition within mixed lists. If associativity is bidirectionally equivalent, and if the recall of an item is a function of the interitem associativity that it shares with the previous item in the list (Howard & Kahana, 2002b), then the likelihood it will be recalled, given recall of the previous item, should be the same for HF-to-LF and LF-to-HF transitions. The standard serial recall analysis masks these sensitivities, since all instances of successful serial recall at each serial position are reported. Accordingly, in this study, a novel approach to recall scoring was adopted where recall at each position was examined when conditionalized upon successful recall of the previous item.

Method

Participants

A total of 103 undergraduate University of Wollongong students participated in the experiment for course credit. The data from 7 participants were excluded from analysis because they were not native Australian English speakers (6), or were visually impaired (1). The remaining 96 participants were allocated to one of four conditions. The pure list participants (21 female, 3 male) had a mean age of 23.8 years ($SD = 8.6$ years), the alternating list participants (18 female, 6 male) had a mean age of 22.2 years ($SD = 8.0$ years), the half-list participants (22 female, 2 male) had a mean age of 22.8 years ($SD = 5.8$ years), and the sequence list participants (19 female, 5 male) had a mean age of 21.0 years ($SD = 3.8$ years).

Materials

The stimulus sets were those used by Miller and Roodenrys (2012). The two sets of 96 CVC words selected on the basis of frequency ratings (instances per million words of text) were drawn from the Celex database (Baayen, Piepenbrock, & Van Rijan, 1993). These ratings combined database entries for the same orthography and the same word identification number, so that, for example, the frequency counts for *bird* and *birds*

were tallied. The counts of any homophones were also included in the frequency ratings. The mean \log_{10} frequency ratings of the sets were, for LF, $M = 0.70$ ($SD = 0.37$) and for HF, $M = 2.21$ ($SD = 0.27$), and the words sets were found to differ significantly on raw frequency ratings, Mann–Whitney $U = 0.00$, $p < .001$. Sets were matched on concreteness (MRC database; Coltheart, 1981), using a weighted average concreteness value for homophone items calculated with the relative frequencies as weights, $U = 4,495.50$, $p = .770$, and the number of phonological neighbors (derived from the Celex database), $U = 4,560.50$, $p = .902$. Furthermore, the number of items that were phonological neighbors of other items in the set did not differ between sets, $U = 4,555.50$, $p = .890$.

Phonological similarity was measured using an Excel Visual Basic program that compares phoneme feature similarity for onset, vowel, and coda segments of stimuli (PSI-METRICA; Mueller, Seymour, Kieras, & Meyer, 2003). Pairwise comparisons are made for all stimuli of a set, and the dissimilarity profile is determined as the average of each cluster measure across all pairwise combinations. None of the resultant phonological similarity measures differed significantly between sets (onset, $U = 4,117.00$, $p = .202$; nucleus, $U = 4,398.00$, $p = .585$; coda, $U = 4,365.00$, $p = .527$).

Vowel quality (categorized as short vowels, long vowels, and diphthongs) in each set was measured using the DeCaro and Goswami (2002) database. A $3 \times 2 \chi^2$ test for independence identified that a marginal difference in vowel quality was present between HF and LF sets, $\chi^2(2) = 4.78$, $p = .091$. However, since there were more words with short vowels and fewer words with long vowels in the LF set, any effect favoring HF words in serial recall cannot be attributed to differences in vowel quality.

Recent work has highlighted the possibility that co-articulatory fluency—namely, the ease with which articulatory transitions between items are performed—can influence serial recall performance (Murray & Jones, 2002; Woodward, Macken, & Jones, 2008). More specifically, this account offers a rival explanation of the frequency effect—namely, that the transitions between HF items are more easily negotiated than transitions between LF items (Woodward et al., 2008). Accordingly, the stimulus sets were examined in relation to the coarticulation characteristics (manner and place of articulation and voicing) of all coda–onset combinations (9,120 boundaries per word set). This analysis suggested that differences in articulatory complexity between stimulus sets were unlikely to contribute markedly to a frequency effect in serial recall. The proportions of word boundaries with the same manner of articulation were similar for HF and LF sets (.264 vs. .279). Transitions involving the same place of articulation were similar for HF and LF sets (.325 vs. .319). Although boundaries likely to be easily assimilated (those

with easily coarticulated coda–onset consonants—e.g., *phone* → *ball*) were greater for HF than for LF words (.131 vs. .089), instances of minor place changes were fewer for HF than for LF words (.234 vs. .309). Furthermore, more instances of word boundaries with major place changes in articulation occurred in the HF set (.310 vs. .285). HF and LF sets possessed similar proportions of cases with boundaries involving no change of voicing (.509 vs. .516).

A final check of possible differences in coda–onset coarticulation in the word sets used transitional probabilities of biphone frequency in language. These measures quantify how often two phonemes occur in order in natural language (Toro, Nespors, Mehler, & Bonatti, 2008). It is argued that transitional probabilities can be used as a proxy to estimate coarticulation fluency between words. The biphone frequency database of Frankish (unpublished) that is sourced from Celex database information provided transitional probabilities of all coda–onset combinations in each set. Setwise distributions were found to be positively skewed, so these data were subjected to a square root transformation. No significant difference between HF and LF sets ($M_{Diff} = .002$), $t(18238) = 1.62$, $p = .105$, was found on the transformed data sets, suggesting that the frequency with which transitions were encountered in either set was similar in terms of biphone frequency.

The stimuli were digitally recorded by a female native Australian English speaker and were converted to sound files using the ProTools LE software on a G4 Macintosh computer. Experimental sessions were conducted on an IBM-compatible PC that ran prepared script files loaded into SuperLab v. 2.0.4. Sound files were amplified using an external speaker that was attached to the PC.

Script files contained 64 six-word trials and tested one of four list format conditions: pure frequency, alternating, half, or sequence lists. Each item of the HF and LF word sets was presented twice within script files, once within each of the list types of the list format. Allocation of items to serial positions was random within the constraints of each list format, and the presentation of individual trial types was random.

Procedure

Participants were assigned to each of the list format conditions on a rotating basis, according to the order of testing. All participants were tested individually. The total time to complete the experiments was approximately 45 min. Initiation of each trial occurred when the participant pressed the space bar of the keyboard. The program would then play the sound files of each word in the trial at a rate of one word per second. After the presentation of the sixth word, a recall prompt (“?????”) appeared on the screen, indicating that the participant should commence recall. Spoken recall was

according to strict serial recall criteria—namely, (1) words were recalled in order of presentation; (2) if a word could not be recalled, the participant would indicate by saying “blank”; and (3) previous items were not to be recalled after participants moved on to successive items in the list. The experimental program presented four practice trials that were used to familiarize participants with the task requirements before the commencement of the experimental phase.

Results

The data were scored according to a strict serial recall criterion; that is, a recalled item was considered to be correct if it was recalled in its presentation serial position. The mean number of correctly recalled items by serial position and condition is shown in Fig. 1, where the distinct effects of list composition are readily observed. The patterns for the pure, alternating, and half-list formats replicated those obtained in previous experiments (Hulme et al., 2003; Miller & Roodenrys, 2012), while the novel sequence format was consistent with the half-list condition, in that the convergence of the curves occurred at the point of change of item type. The means for formats collapsed across list types suggested that the variation in the overall level of performance for each condition was small; descriptively, the items in pure lists were recalled the least well ($M = .535$), followed by those in alternating lists ($M = .541$) and then half lists ($M = .559$), while the items in the sequence lists were recalled the most ($M = .583$). However, since list format was the between-groups variable in the experiment, this variation may be participant related. The overall means for the list types of each format revealed that lists beginning with HF words were recalled better than lists beginning with LF words: pure lists, HF ($M = .634$) versus LF ($M = .436$); alternating lists, HLHLHL ($M = .556$) versus LHLHLH ($M = .526$); half lists, HHHLLL ($M = .576$) versus LLLHHH ($M = .541$); and sequence lists, HHLLLH ($M = .601$) versus LLHHHL ($M = .566$).

In summary, it would appear that HF items in mixed lists beginning with LF words were recalled better than their counterparts in only a handful of serial positions. Furthermore, the differences where performance was superior in these lists did not compensate for the advantage that lists beginning with HF words realized in the other serial positions.

Serial recall

An alpha level of .05 was applied to the statistical tests performed. The data were subjected to a $4 \times 2 \times 6$ mixed analysis of variance where list format (pure, alternating, half, and sequence) was the between-subjects factor and list type (lists commencing with HF or LF words) and serial

position (1–6) were the within-subjects factors. This analysis revealed a main effect of list type, $F(1, 92) = 138.30$, $p < .001$, $\eta_p^2 = .601$, confirming that lists beginning with HF words were recalled better than lists beginning with LF words, and a main effect of serial position, $F(5, 460) = 363.07$, $p < .001$, $\eta_p^2 = .798$, but the effect of format was nonsignificant, $F(3, 92) = 0.82$, $p = .480$, $\eta_p^2 = .026$. Thus, differences in mean performances between formats were not reliable. The list type \times list format interaction was significant, $F(3, 92) = 42.80$, $p < .001$, $\eta_p^2 = .583$, reflecting primarily the greater difference in recall of pure list types in comparison with differences between list types in the other formats. The list type \times serial position interaction was also significant, $F(5, 460) = 9.96$, $p < .001$, $\eta_p^2 = .098$, driven by the asymmetries in the list types for the half and sequence formats. However, the interaction of format and serial position was nonsignificant, $F(15, 460) = 1.07$, $p = .382$, $\eta_p^2 = .034$, identifying that there was no difference across serial position in the average recall between formats. This result was qualified by a significant three-way interaction (format \times list type \times serial position), $F(15, 460) = 9.70$, $p < .001$, $\eta_p^2 = .240$, which demonstrated that when performance was considered by list type, differences in serial recall curves between lists beginning with HF items and those beginning with LF items varied across the four formats.

This interaction was explored further by analyzing the data for each format separately as four 2×6 repeated measures analyses of variance. Pure lists contained a significant main effect of list type, $F(1, 23) = 199.19$, $p < .001$, $\eta_p^2 = .896$, replicating the standard frequency effect, and a significant main effect of serial position, $F(5, 115) = 111.23$, $p < .001$, $\eta_p^2 = .829$. The list type \times serial position interaction was also significant, $F(5, 115) = 5.52$, $p < .001$, $\eta_p^2 = .194$, indicating that the effect increased over the first serial positions. To resolve whether the interaction was a result of a ceiling effect operating on the first position for HF lists, the last five serial positions were reanalyzed. The main effects were once again significant [list type, $F(1, 23) = 219.10$, $p < .001$, $\eta_p^2 = .905$; serial position, $F(4, 92) = 61.80$, $p < .001$, $\eta_p^2 = .729$], but the interaction was nonsignificant, $F(4, 92) = 0.13$, $p = .971$, $\eta_p^2 = .005$, supporting a ceiling effect interpretation of the interaction.

A 2×6 repeated measures analysis of variance on the alternating list data produced a significant main effect of list type, $F(1, 23) = 5.63$, $p = .026$, $\eta_p^2 = .197$. Thus, lists that began with an HF item were recalled modestly but reliably, better than those beginning with an LF item. The main effect of serial position was significant, $F(5, 115) = 82.82$, $p < .001$, $\eta_p^2 = .948$, as was the list type \times serial position interaction, $F(5, 115) = 6.30$, $p < .001$, $\eta_p^2 = .612$. The presence of an interaction was due presumably to the subtle sawtooth pattern present in both list types, indicating the

possibility of small item-specific effects (Saint-Aubin & LeBlanc, 2005).

The equivalent analysis of the half-list format data resulted in main effects of list type, $F(1, 23) = 11.79$, $p = .002$, $\eta_p^2 = .339$, and serial position, $F(5, 115) = 127.56$, $p < .001$, $\eta_p^2 = .847$. Furthermore, the list type \times serial position interaction was significant, $F(5, 115) = 16.91$, $p < .001$, $\eta_p^2 = .424$, highlighting the superiority of lists beginning with HF words. Bonferroni-adjusted simple effects identified a significant frequency effect for positions 1, 2, 3, and 6. Therefore, at the point of change in item frequency (the fourth serial position), performance converged, was similar for the fifth position, and then diverged at the recency position, where HF items were recalled better than LF items.

Lastly, a within-subjects analysis of the sequence list data revealed main effects of list type, $F(1, 23) = 6.51$, $p = .018$, $\eta_p^2 = .221$, and serial position, $F(5, 115) = 63.13$, $p < .001$, $\eta_p^2 = .733$. The list type \times serial position interaction was also significant, $F(5, 115) = 9.73$, $p < .001$, $\eta_p^2 = .297$. The interaction adopted the pattern for half lists over the initial serial positions, confirming the better recall of the HHLLLH than of the LLHHHL list type. Simple effects analysis using Bonferroni adjustment revealed that significant frequency effects were present in only the first two serial positions. At the first change in list composition (position 3), performance converged; however, in this condition, the difference in recall between list types was not reliable for the remainder of the serial positions (3–6).

Using conditional probabilities to examine interitem effects

To explore the possibility that serial recall is driven, at least in part, by some item-to-item associative mechanism, the data in the experiment were rescored according to a conditional recall criterion. Should an interitem associative mechanism influence recall, it would be expected that the conditional likelihood that an item would be recalled would be the same for transitions between HF and LF items in mixed lists, regardless of the order of items, since these have been argued to possess interitem association of intermediate strength (Hulme et al., 2003).

Despite concerns regarding the accuracy of dependency measures (see Henson, Norris, Page, & Baddeley, 1996), it was thought, given the use of open stimulus sets in this instance, that the use of transitional shift probabilities, (i.e., conditionalized probabilities based on the recall status of the previous item only) should be sufficient to indicate dependency on the relationship between adjacent items. However, it is acknowledged that effects of disruption on recall of later list items may mask dependencies due to item-to-item association. Accordingly, it is appropriate to remain mindful that measures of this sort for later serial positions might contain influences from a number of sources.

A strict position interpretation was adopted for these data, due to the short length of the supraspan lists. That is, the proportion of instances where an item presented in serial position j was recalled in position j on the condition that the previous items in position $j - 1$ had also been recalled correctly was recorded. These data would identify any instances where interitem associations between list items could explicitly operate. To account for the changing size of the sample space when calculating the conditional probabilities associated with these proportions, the following formula was used:

$$P(j|j-1) = \frac{P(j \cap j-1)}{P(j-1)},$$

where $P(j|j-1)$ is the probability of recall of an item in position j subject to the correct recall of the previous item $j - 1$, $P(j-1)$ is the probability of correct recall of the prior item, and $P(j \cap j-1)$ is the probability of the correct recall of items j and $j - 1$ (Larsen & Marx, 1981). The term $P(j-1)$ becomes an estimate for the adjusted sample space. These probabilities are presented for each mixed list condition in Fig. 2a. The data for the first serial position are the proportions recalled as per the original serial recall scoring, since there is no previous event for this position. The equivalent data for the pure lists are presented for each mixed list condition and act as an envelope for the mixed list values.

The identification of recall events in this way allows the data to be fractionated into *continuous* recall, as described above, and those instances where recall occurred despite recall of the previous item being in error, termed *recovery* recall. In this situation, recall might reflect an item-specific influence and indicate how well recall can recover from disruption at output for serial positions 2–6. The conditional probabilities for the recovery data are given in Fig. 2b. The continuous and recovery recall data sets were examined separately.

Simple effects on the conditionalized data were conducted to determine whether they conformed to the patterns that directional interitem associativity would anticipate. With respect to continuous recall, frequency effects would be expected in serial positions where corresponding sequences of HF and LF items were presented in a condition, while no difference in effect would occur at transition points in these lists. Specifically, sequences of HF items should be recalled better than sequences of LF words because HF items have stronger preexperimental association, and HF-to-LF and LF-to-HF transitions in lists should result in the same level of recall. Therefore, it would be predicted that positions 2, 3, 5, and 6 in half lists and positions 2, 4, and 5 in sequence lists would produce differences. In contrast, no difference between the conditional probabilities of continuous recall should exist for all positions in alternating lists

(2–6), while position 4 in the half lists and positions 3 and 6 in the sequence lists should also be equivalent. Additionally, under the assumption that the fractionation of continuous and recovery recall accurately separates interitem and item-specific effects, it would be expected that if item-specific effects do not influence recall, no differences in the recovery data should exist for any of the conditions.

The analysis found that all points of transition between HF and LF items, except for position 5 in the alternating lists, produced nonsignificant differences in the conditional probabilities for continuous recall; the exception was found to be a marginal result. Furthermore, significant differences were found for positions 2, 3, and 6 in half lists and position 2 in sequence lists. In summary, one out of the eight positions where no difference was predicted produced a marginal effect, while four out of the seven positions predicted to produce a frequency effect did so. Therefore, the continuous data, particularly across the first four serial positions, were consistent with a directional interitem associativity explanation of serial recall.

The recovery recall identified that a significant difference occurred in position 3 of the sequence lists. That is, at the point of transition in the list, HF words were recovered better than LF words after the previous item was not correctly recalled. None of the other comparisons for these data reached significance, although position 3 for alternating lists was marginal. In general then, according to this analysis, recovery episodes were free of the influence of frequency.

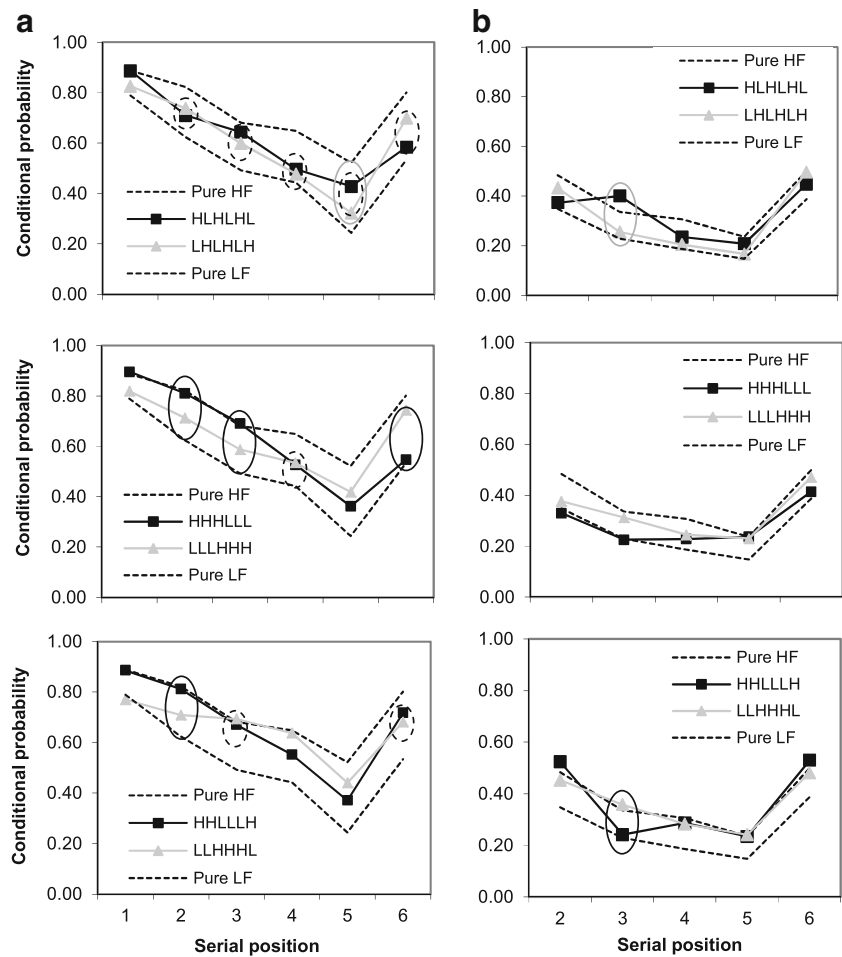
Item analysis

Responses in each trial were classified as correct (recall of the item occurred in the correct serial position), an order error (the item belonged to another serial position in the trial), or an item error (intrusions, omissions, and repetitions). The proportion of errors of each classification, for each item type in each list type, collapsed across serial position and participants is given in Table 1, together with the proportion of items correct.

Order errors were conditionalized to indicate the extent to which recalled items were positioned incorrectly, relative to presentation order (Murdoch, 1976; Poirier & Saint-Aubin, 1996; Saint-Aubin & Poirier, 1999). This process identified that the adjusted error rate for HF items in pure lists was .127, for LF items in pure lists was .149, for HF items in alternating lists was .119, for LF items in alternating lists was .115, for HF words in half lists was .110, for LF words in half lists was .122, for HF words in sequence lists was .100, and finally, for LF words in sequence lists was .099.

A 4×2 (list format \times item type) mixed ANOVA was performed on the conditionalized order data and identified nonsignificant effects of frequency $F(1, 92) = 2.35, p = .129, \eta_p^2 = .025$, and list format, $F(3, 92) = 1.86, p = .142, \eta_p^2 = .057$,

Fig. 2 The conditional probabilities of continuous (a) and recovery (b) recall between consecutive items for list types of mixed list conditions. Dashed lines form the envelope of the conditional probabilities of pure lists. Dashed circles in black identify whether interitem associativity predicts no difference in likelihood of recall (serial positions 2–6). Solid circles in black indicate where effects exist with Bonferroni adjustment (serial positions 2–6). Solid circles in gray indicate marginal effects (serial positions 2–6)



and a nonsignificant frequency × format interaction, $F(3, 92) = 1.82, p = .147, \eta_p^2 = .056$. Therefore, there was no difference between HF and LF stimuli in the proportion of all items remembered but recalled in the wrong position.

The analysis of the total item error data revealed an effect of frequency, $F(1, 92) = 254.46, p < .001, \eta_p^2 = .734$, and a significant frequency × list format interaction, $F(3, 92) = 50.08, p < .001, \eta_p^2 = .620$, but the main effect of list format was not significant, $F(3, 92) = 0.50, p = .683, \eta_p^2 = .016$. Across conditions, HF words incurred fewer item errors than did LF words. The interaction reflected the greater frequency effect for pure than for mixed lists. Bonferroni-adjusted tests on each list format identified, however, that all conditions produced significant frequency effects [pure lists, $t(23) = 16.66, p < .001$; half lists, $t(23) = 5.88, p < .001$; sequence lists, $t(23) = 4.51, p < .001$; alternating lists, $t(23) = 3.52, p = .002$].

Discussion

An objective of this experiment was to explore the impact of shifting a three-item sequence, as found in the second half of

half lists, one position forward. This arrangement served to avoid the coincidence of the third item in sequence with the recency position and to test whether frequency effects between list items were possible in the second half of the list after a transition between HF and LF items.

The analysis of correct recall for sequence lists failed to find a significant frequency effect for the third HF and LF items in sequence (the serial position 5). Therefore, according to these data, influences of preexperimental interitem associativity between list items are insufficient to support differences in the recall of mixed lists in the second half of the list. While nonsignificant trends indicate better recall for HF than for LF words, it is clear that directional associativity alone does not determine the relative success of recall in late serial positions.

The correct serial recall data for pure lists revealed a frequency effect replicating previously reports (e.g., Hulme et al., 2003, Experiment 2). The results for alternating lists indicated that lists beginning with HF items experience a small advantage in correct recall, possibly as a consequence of an item-specific contribution that operates at the start-of-list position (Hulme et al., 2003, Experiment 2). Evidence for a directionally sensitive contribution of interitem associativity

Table 1 Mean proportion of correct recall, order errors, and item errors as a function of item type and list format

	Correct	Errors	
		Order	Item
HF in pure lists	.634 (.125)	.090 (.044)	.277 (.115)
LF in pure lists	.434 (.101)	.073 (.035)	.494 (.095)
HF in alt. lists	.559 (.128)	.071 (.029)	.370 (.119)
LF in alt. lists	.524 (.126)	.063 (.030)	.413 (.110)
HF in half lists	.592 (.113)	.071 (.033)	.337 (.101)
LF in half lists	.525 (.136)	.069 (.034)	.406 (.124)
HF in seq. lists	.613 (.107)	.065 (.033)	.322 (.087)
LF in seq. lists	.554 (.135)	.054 (.024)	.391 (.120)

Note. Standard deviations are given in parentheses. HF, high frequency; LF, low frequency; alt., alternating; seq., sequence

was produced in the recall of half lists, where recall mimicked pure lists across the first three serial positions. However, recall beyond the halfway point of the list did not produce a reliable frequency effect, except for the recency position, which might also respond to the item-specific properties of items.²

The possibility that interitem associativity between consecutive list items had been obscured by the absolute levels of recall of the previous items and masked in the second half of the lists by instances of recovery after a failure to recall the previous item was considered by fractionating the correct recall data into continuous and recovery recall events and expressing these as conditional probabilities. Tests on these measures, gauging how well the continuous data conformed to a directional interitem associativity explanation, indicated that the first three to four list positions were well accounted for, although a significant difference in the recovery data occurred for sequence lists, where recovery of HF words in position 3 was greater than for LF words. Accordingly, the results of a second analysis reinforces the proposition that while interitem effects of a directional nature operate for the primacy portion of the list, these effects are less influential for late list positions. Only one out of four positions where sequences of HF and LF items occurred in the second half of the lists produced a significant effect, and this instance coincided with the recency position.

This experiment did not find any order effects related to word frequency. If it is assumed that the proportion of order errors due to the loss of item information is the same for HF and LF words, this places the locus of frequency effect with differences in the retention of item information (Hulme et al.,

2003; Poirier & Saint-Aubin, 1996; Stuart & Hulme, 2000). The pattern of total item errors revealed that list composition influenced the degree to which item recall was superior for HF words; the effect was greater for pure than for mixed lists. The inclusion of HF and LF words in the same list reduces errors for LF words and increases them for HF words, relative to pure lists (Hulme et al., 2003). However, all formats produced frequency effects in the item error data. Therefore, despite differences in item arrangement, LF words were associated with greater item error rates than were HF words across mixed list conditions. These results contrast with those of Hulme et al. (2003, Experiment 2), who found a reversed frequency effect for item errors in alternating lists.

Accordingly, the first serial positions in recall are influenced by the preexperimental associations between adjacent list items. Lists containing HF-to-LF transitions in the first serial positions (e.g., alternating lists) will produce recall performance consistent with the moderate strength of association between items. Mixed lists that contain sequences of HF items at the start of the list will be recalled reliably better than those containing sequences of LF words (half and sequence lists) in these serial positions. The similar levels of recall observed for later list items of mixed lists might be a product of directional associativity in combination with the level of available resources that are determined by the efficiency of output for earlier items. That is, for half and sequence lists, the absence of frequency effect might arise because lists beginning with HF items require fewer resources to output these early items than do pure LF lists and so, relatively speaking, the recall of LF items in half and sequence lists enjoy a benefit that buffers the effects of directional associativity. Conversely, relative to pure HF lists, half and sequence lists commencing with LF items might have fewer resources available for the recall of HF items later in the lists, with this cost countering the facilitative effect of directional associativity. Furthermore, if item-to-item associativity was maintained throughout and obscured for final items, relationships between learning contexts and preexisting associations of items in semantic memory, of the kind promoted by the TCM (Howard & Kahana, 2002a, 2002b), might be responsible.

While coarticulatory fluency is thought to be a minor feature in the present investigation, it is worth considering whether the presumed effects of this variable (Woodward et al., 2008) are consistent with the patterns of results observed across the conditions tested. A simple interpretation of this approach would predict that recall would be a function of the difficulty of coarticulating phonemes at word boundaries (as determined by frequency-based transitions) during rehearsal. Consistent with this interpretation, the portion of the list observed to reflect putative item-to-item associative effects does correspond to the subspan of items that are

² While an effect at recency was consistent with item type in this experiment, Hulme et al. (2003) found a marginally significant negative frequency effect—that is, better recall for LF than for HF words—in the recency position with alternating lists in their Experiment 2.

cumulatively rehearsable within the interstimulus interval at presentation (Page & Norris, 1998; Tan & Ward, 2008).³ Therefore, early sequences of HF items, argued to be more reproducible in terms of speech programming than sequences of LF words, should be recalled better. However, the absence of a frequency effect between sequences of HF and LF words in the second half of the list suggests that influences wider than the coarticulation at word boundaries determine recall levels for these serial positions. It could be argued, for example, in the case of half lists, that the reduction in recall of HF items in the second half of the list, when compared with the recall of pure lists, would be due to the impairment to speech planning brought about by the relative difficulty in sequencing three LF words in the first half of the list. Similarly, the recall for LF items in the second half of the list, observed to be greater than the recall for pure LF lists, could be explained as the benefit afforded to these items by the more efficient speech programming for the first half of the list. One test of this proposal would be to conduct replications of these studies using articulatory suppression; however, Miller and Roodenrys (2012) found that the pattern of the frequency effect in half lists was not altered by suppression. Therefore, if coarticulatory factors play a role in the formation of the effect, they must act at a point prior to subvocal rehearsal. The perceptual–gestural account of STM (Hughes, Marsh, & Jones, 2009; Jones, Hughes, & Macken, 2006; Murray & Jones, 2002), from which the coarticulation account derives, argues that serial recall does not involve mnemonic storage and processes. Instead, stimuli are organized into a perceptual stream that is used to sequence speech motor processes and produce an utterance for output. Accordingly, the coarticulation hypothesis would need to argue that some level of articulatory planning is the locus of the frequency effect.

The presence of preexperimental, item-to-item associative effects in serial recall, as determined by word frequency, highlights yet another way in which the organization of language knowledge constrains what is remembered over the short term. The complexity of the relationship between list composition, item arrangement, and the frequency effect is heightened by the possibility that directional associativity exists throughout recall but is masked for late serial positions by the relative efficiencies or costs that arise from previous output. The findings reinforce the claim that the serial recall task is not as simple as it may seem on the surface (Watkins & Watkins, 1977), and teasing apart the processes that contribute to serial recall at different points in the list poses a considerable challenge.

³ Although the presentation rate used in this experiment was one word per second, all items had a CVC structure and so could be rehearsed rapidly.

Author Note Leonie M. Miller, School of Psychology, University of Wollongong; Steven Roodenrys, School of Psychology, University of Wollongong.

This research was conducted as part of the first author's doctoral thesis and was funded by an Australian Postgraduate Award.

References

- Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language*, 55, 64–88.
- Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The Celex Lexical Database*. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.
- Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.
- Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, 33A, 497–505. Retrieved from http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm
- DeCara, B., & Goswami, U. (2002). Similarity relations among spoken words: The special status of rimes in English. *Behavior Research Methods, Instruments, & Computers*, 34, 416–423.
- Deese, J. (1960). Frequency of usage and number of words in free recall: The role of association. *Psychological Reports*, 7, 337–344.
- Gregg, V., Freedman, C. M., & Smith, D. K. (1989). Word frequency, articulatory suppression and memory span. *British Journal of Psychology*, 80, 363–374.
- Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *The Quarterly Journal of Experimental Psychology*, 49A, 80–115.
- Howard, M. W., & Kahana, M. J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46, 269–299.
- Howard, M. W., & Kahana, M. J. (2002b). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, 46, 85–98.
- Hughes, R. W., Marsh, J. E., & Jones, D. M. (2009). Perceptual–gestural (mis)mapping in serial short-term memory: The impact of talker variability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1411–1425.
- Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a reintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 1217–1232.
- Hulme, C., Stuart, G., Brown, G. D. A., & Morin, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language*, 49, 500–518.
- Jones, D. M., Hughes, R. W., & Macken, W. J. (2006). Perceptual organization masquerading as phonological storage. *Journal of Memory and Language*, 54, 265–281.
- Larsen, R. J., & Marx, M. L., (1981). *An introduction to mathematical statistics and its applications*. Englewood Cliffs, NJ: Prentice-Hall.
- Miller, L. M., & Roodenrys, S. (2012). Serial recall, word frequency and mixed lists: The influence of item arrangement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/a0028470.

- Morin, C., Poirier, M., Fortin, C., & Hulme, C. (2006). Word frequency and the mixed-list paradox in immediate and delayed serial recall. *Psychonomic Bulletin & Review*, *13*, 724–729.
- 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*, 1353–1380.
- Murdock, B. B., Jr. (1976). Item and order information in short-term serial memory. *Journal of Experimental Psychology: General*, *105*, 191–216.
- Murray, A., & Jones, D. M. (2002). Articulatory complexity at item boundaries in serial recall: The case of Welsh and English digit span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 594–598.
- Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*, 761–781.
- Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology*, *50*, 408–412.
- Roodenrys, S., & Quinlan, P. T. (2000). The effects of stimulus set size and word frequency on verbal serial recall. *Memory*, *8*, 71–78.
- Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1019–1034.
- Saint-Aubin, J., & LeBlanc, J. (2005). Word frequency effects in immediate serial recall of pure and mixed lists: Tests of the associative link hypothesis. *Canadian Journal of Experimental Psychology*, *59*, 219–227.
- Saint-Aubin, J., & Poirier, M. (1999). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *Quarterly Journal of Experimental Psychology*, *52A*, 367–394.
- Saint-Aubin, J., & Poirier, M. (2005). Word frequency effects in immediate serial recall: Item familiarity and item co-occurrence have the same effect. *Memory*, *13*, 325–332.
- Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, *21*, 168–175.
- Stuart, G., & Hulme, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 796–802.
- Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review*, *15*, 535–542.
- Tehan, G., & Humphreys, M. S. (1988). Articulatory loop explanations of memory span and pronunciation rate correspondences: A cautionary note. *Bulletin of the Psychonomic Society*, *26*, 293–296.
- Toro, J. M., Nespore, M., Mehler, J., & Bonatti, L. L. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science*, *19*, 137–144.
- Watkins, O. C., & Watkins, M. J. (1977). Serial recall and the modality effect: Effects of word frequency. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 712–718.
- Woodward, A. J., Macken, W. J., & Jones, D. M. (2008). Linguistic familiarity in short-term memory: A role for (co-)articulatory fluency? *Journal of Memory and Language*, *58*, 48–65.
- Wright, C. E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory & Cognition*, *7*, 411–419.