

# Does length or neighborhood size cause the word length effect?

Annie Jalbert · Ian Neath · Aimée M. Surprenant

Published online: 2 April 2011  
© Psychonomic Society, Inc. 2011

**Abstract** Jalbert, Neath, Bireta, and Surprenant (2011) suggested that past demonstrations of the word length effect, the finding that words with fewer syllables are recalled better than words with more syllables, included a confound: The short words had more orthographic neighbors than the long words. The experiments reported here test two predictions that would follow if neighborhood size is a more important factor than word length. In Experiment 1, we found that concurrent articulation removed the effect of neighborhood size, just as it removes the effect of word length. Experiment 2 demonstrated that this pattern is also found with nonwords. For Experiment 3, we factorially manipulated length and neighborhood size, and found only effects of the latter. These results are problematic for any theory of memory that includes decay offset by rehearsal, but they are consistent with accounts that include a redintegrative stage that is susceptible to disruption by noise. The results also confirm the importance of lexical and linguistic factors on memory tasks thought to tap short-term memory.

**Keywords** Memory · Recognition · Serial position effects · Short term memory · Working memory

The word length effect, the finding that short words (e.g., *lead*, *pig*, *grape*) are recalled better than long words (e.g., *aluminum*, *elephant*, *banana*), has played a significant role

in the development of theories of memory. This effect is the basis of the phonological loop component of working memory (Baddeley, 1992); it has been described as “the best remaining solid evidence” for such a temporary memory subsystem, in which decay is offset by rehearsal (Cowan, 1995, p. 42); and it has been termed a “benchmark finding” that computational models of memory must account for (Lewandowsky & Farrell, 2008). We (Jalbert et al., 2011) recently suggested that this effect may not be due to length per se, but rather to the effects of neighborhood size, because previous demonstrations of the word length effect confounded length and neighborhood size. In this article, we test two predictions that arise out of an account that attributes word length effects to neighborhood size rather than to length per se: (1) The neighborhood size effect, like the word length effect, should be eliminated if subjects engage in concurrent articulation.<sup>1</sup> (2) Long items with a large neighborhood size should be recalled better than short items with a small neighborhood size.

## Word length and working memory

In the first systematic exploration of the effects of word length, Baddeley, Thomson, and Buchanan (1975) reported three key results. First, a set of words was created in which the short and long items differed in pronunciation

---

A. Jalbert (✉) · I. Neath · A. M. Surprenant  
Psychology Department, Memorial University of Newfoundland,  
St. John's, NF A1B 3X9, Canada  
e-mail: annie.jalbert@mun.ca

---

<sup>1</sup> We use the term *concurrent articulation* rather than the more usual *articulatory suppression* because the former is a neutral description of what the subject is asked to do. In contrast, the latter term implies a specific effect of the manipulation, and below we argue for a different effect of this manipulation.

time but were equated for number of syllables, number of phonemes, and frequency. More short words than long words were recalled on an immediate spoken serial recall test. This is now referred to as the *time-based word length effect*, since the key difference between the short and long words is the time necessary to pronounce the words. Second, a different set of words was created that differed in both pronunciation time and number of syllables (and phonemes). Words from one to five syllables long and from the same category were used (e.g., *Maine, Utah, Wyoming, Alabama, Louisiana*). Again, recall was related to the length of the words in the list. The third key finding was that the word length effect was removed if subjects engaged in concurrent articulation, repeatedly saying the digits 1–8 out loud at an approximate rate of three digits per second, during list presentation.<sup>2</sup>

According to the working memory framework (Baddeley, 1986, 1992, 2000), this pattern reflects the operation of the phonological loop. The to-be-remembered words enter the phonological store and decay after about 2 s if they are not refreshed by an articulatory control process. Forgetting occurs when the time necessary to rehearse the items is longer than the decay rate. Assuming that there is a positive relationship between the rate of rehearsal and pronunciation time, it will take longer to refresh a list of long words than a list of short words, and therefore fewer long words will be available to be recalled, as compared to short words. Concurrent articulation is assumed under this account to prevent the use of the articulatory control process (hence the common term *articulatory suppression*), so both short and long items decay at the same rate.

The time-based word length effect has been replicated many times using the original stimuli (e.g., Cowan et al., 1992; Longoni, Richardson, & Aiello 1993; Lovatt, Avons, & Masterson 2000; Nairne, Neath, & Serra 1997). However, there have been no demonstrations of a time-based word length effect using other sets of words. Indeed, there are five sets of stimuli, in which short and long words differ only in pronunciation time, that fail to produce a time-based word length effect (Lovatt et al., 2000; Neath, Bireta, & Surprenant 2003; see also Service, 1998).

The lack of a time-based word length effect when using words other than the original stimuli poses a problem for memory models incorporating a subsystem like the phonological loop. Despite this, proponents of the decay-offset-by-rehearsal view emphasize the robustness of the syllable-

based word effect as evidence for their view. Unlike the time-based word length effect, the syllable-based word length effect has been found with a large number of different stimulus sets. Moreover, it is observable on a variety of tasks, including reconstruction of order (Bireta, Neath, & Surprenant 2006), serial recognition (Baddeley, Chincotta, Stafford, & Turk 2002), free recall (Watkins, 1972), single-item probe recall (Avons, Wright, & Pammer 1994), and both simple (LaPointe & Engle, 1990) and complex (Tehan, Hendry, & Kocinski 2001) span tasks.

### Alternative accounts and stimulus set specificity

Several alternative explanations of the word length effect have been proposed. Brown and Hulme (1995) proposed a model in which short and long words have a different decay rates, with short words decaying more slowly than long words, making short words easier to recall than long words. Unlike the phonological loop model, this one has no role for rehearsal. Nonetheless, simulations show that the model accounts for many aspects of the word length effect. The feature model (Neath & Nairne, 1995) also posits an item-based explanation of the word length effect, but with no role for decay. According to the model, long words have more segments than short words. If one assumes a fixed probability of making an error while assembling the segments for recall, there are more opportunities for a mistake with long words; thus, a word length effect will result. Again, simulations have shown that the model accounts for the major effects observed, including the removal of the word length effect by concurrent articulation.

Even though the phonological loop framework and the two simulation models all account for the word length effect, they do so by appealing to different underlying mechanisms. Because of this, the models make different predictions for situations in which short and long words are mixed in the same list. According to the phonological loop framework, the list that can be rehearsed most quickly will be recalled best, and the list that requires the most time to be rehearsed will have the lowest level of recall. The mixed lists as a whole will take less time to rehearse than the pure long lists, but more time to rehearse than the pure short lists, so recall of short and long items in mixed lists should be intermediate between that of pure short and pure long lists. Importantly, the recall of short and long items from mixed lists should be equivalent. In contrast, because they focus on item-based processes, both the Brown and Hulme (1995) model and the feature model predict that

<sup>2</sup> Baddeley, Lewis, and Vallar (1984) showed that it is necessary to have concurrent articulation continue during recall in order to eliminate the word length effect if auditory presentation is used. In this article, we focus just on visual presentation.

short words in mixed lists will be recalled better than long words in mixed lists.

Despite the clear-cut predictions, the results are ambiguous. For example, Cowan, Baddeley, Elliott, and Norris (2003) reported one experiment in which they included pure lists of six short (one-syllable) or six long (five-syllable) words and mixed lists of alternating short and long words. They found that recall performance was best for pure short lists, worst for pure long lists, and intermediate for mixed lists, just as predicted by the phonological loop framework. They also found that recall of short words from mixed lists was better than recall of long words from mixed lists, just as predicted by the two simulation models. In contrast, Hulme, Surprenant, Bireta, Stuart, and Neath (2004) found that recall of short items in mixed lists was equivalent to recall of long items in mixed lists, just as the phonological loop framework predicts. They also found that recall of short items in mixed lists was equivalent to recall of short items in pure lists, just as the two simulation models predict.

Bireta et al. (2006) examined whether the differing results were due to methodological differences or stimulus set differences. Regardless of which methodology they followed, they replicated the results reported by Cowan et al. (2003) when using Cowan et al.'s stimuli, and replicated the results reported by Hulme et al. (2004) when using Hulme et al.'s stimuli. Bireta et al. concluded that the different results are due to differences in the stimulus sets, and further noted that the majority of the stimulus sets resulted in no word length effect in mixed lists.

### Neighborhood size and word length

Given that it is increasingly apparent that the results of a study can vary substantially depending on the particular words included, Jalbert et al. (2011) looked at other ways in which short and long words might generally differ. Although researchers try to equate short and long words on those dimensions that could lead to performance differences, many possibly relevant dimensions are not usually considered. One such dimension is neighborhood size.

A *neighbor* of a word can be defined as a word that differs from the target word by only one letter (for orthographic neighbors) or only one phoneme (for phonological neighbors) (see Coltheart, Davelaar, Jonasson, & Besner 1977; Luce & Pisoni, 1998). Following Jalbert et al. (2011), we focus on orthographic rather than phonological neighbors, as that eliminates the difficulty of differences in pronunciation, and therefore phonological composition. We

also follow the definition that an orthographic neighbor is one that differs from the target word by the substitution of a single letter at any position. For example, orthographic neighbors of the word *cat* would include *bat*, *fat*, *cot*, *cut*, *cab*, *can*, and so on.

Three published studies have shown better recall of words with large neighborhoods than of words with smaller neighborhoods. Roodenrys, Hulme, Lethbridge, Hinton, and Nimmo (2002) used a span task, with auditory presentation and spoken recall, and found larger memory spans for words with a larger neighborhood size, as compared to words with a smaller neighborhood size. Importantly, they tested three different sets of consonant–vowel–consonant (CVC) words. Allen and Hulme (2006, Exp. 2) replicated this neighborhood size effect, but did so with spoken serial recall rather than by using a span task. Finally, Jalbert et al. (2011) also used stimuli from Roodenrys et al., but this time demonstrated a neighborhood size effect with visual presentation and a strict serial reconstruction-of-order test.

Nonwords can also have neighbors. An orthographic neighbor of a nonword is a word that differs from the nonword by just one letter; for example, neighbors of the nonword *rin* include *bin*, *ran*, and *rip*. Roodenrys and Hinton (2002) found that recall of lists of nonwords with a large neighborhood was better than recall of lists of nonwords with a small neighborhood. Roodenrys (2009) argued that the effects of neighborhood size on serial recall occur at retrieval by facilitating the reconstruction of a degraded trace. This process is called “redintegration.” Roodenrys argued that the effect should be placed at output on the basis of the results of phonological neighborhood effects in language tasks. In particular, large phonological neighborhoods (and high-frequency neighbors) act to reduce the probability that a word will be correctly perceived in noise and increase the response time when identifying spoken words (Luce, Pisoni, & Goldinger 1990). In contrast, those same variables have a facilitative effect on speech production tasks (e.g., Vitevitch, 2002; Vitevitch & Sommers, 2003). This concept of redintegration is not necessarily tied to any particular model; for example, it can be readily implemented in both interactive activation (McClelland & Rumelhart, 1981) and language-based models of short-term memory (Martin, Lesch, & Bartha 1999). These will be considered more fully in the General Discussion.

Of relevance to the word length effect, short English words tend to have more neighbors than do long English words. Jalbert et al. (2011) examined published word length studies in which details of the stimulus sets were reported. Neighborhood size was calculated using the Medler and Binder (2005) database, which in turn is based on the

CELEX database. For 13 such studies, the short words had a mean neighborhood size of 8.61, as compared to 0.24 for the long words. Thus, previous demonstrations of a word length effect could be due to length, but could also be due to the confound of neighborhood size.

Jalbert et al. (2011) reported a series of studies that examined whether word length or neighborhood size was driving the word length effect. In one experiment (using both a serial reconstruction-of-order and a serial recall task), they compared recall of lists made up of words with large and small neighborhoods as well as of mixed lists made up of alternating large- and small-neighborhood words. Word length was held constant. Importantly, the full pattern of results resembled the most often-seen pattern when short and long words are presented in pure and mixed lists: A neighborhood size effect was observed with pure lists, but not with mixed lists. This pattern of results was identical with both of the recall methods. If neighborhood size is to be a plausible factor in causing the word length effect, it is critically important that the results of the pure-versus-mixed manipulation be the same as in word length experiments.

Next, they demonstrated that when short and long words were equated for neighborhood size, the word length effect disappeared. They created two sets of stimuli. In each, the short (one-syllable) and long (three-syllable) words were equated on the following dimensions: concreteness, familiarity, imageability, frequency (Kučera–Francis, Thorndike–Lorge, and CELEX), orthographic frequency, orthographic neighborhood size, bigram frequency, neighbor overlap, and PSIMETRICA dissimilarity (see Jalbert et al., 2011, for details). For both sets of stimuli, the serial reconstruction of order of short and long words was equivalent. This result was also replicated using spoken serial recall. In other words, the word length effect disappeared when short and long words were equated for neighborhood size.

Jalbert et al. (2011) concluded that the word length effect might be better explained by the differences in linguistic and lexical properties of short and long words rather than by length per se. If the effect is really due to neighborhood size, however, the variables that interact with word length should also interact with neighborhood size. Thus, concurrent articulation, which abolishes the word length effect, should abolish the neighborhood effect. Note that, although concurrent articulation eliminates a great many phenomena in immediate serial recall, it by no means quashes all of them; in particular, concurrent articulation does not abolish many so-called “long-term memory effects,” including the concreteness effect (Acheson, Postle, & MacDonald 2010), the frequency effect (Gregg, Freedman, & Smith 1989; Tehan &

Humphreys, 1988), and the word class and imageability effects (Bourassa & Besner, 1994). So, if neighborhood size really does mediate the word length effect, the following should be observable: (1) The neighborhood size effect should be abolished by concurrent articulation, just as the word length effect is abolished by concurrent articulation. (2) Long words with large neighborhoods should be better recalled than short words with small neighborhoods. The purpose of the present experiments was to test these predictions.

## Experiment 1

Concurrent articulation is known to abolish or greatly attenuate the word length effect (Baddeley, Lewis, & Vallar 1984; Baddeley et al., 1975; Bhatarah, Ward, Smith, & Hayes 2009; Longoni et al., 1993; Romani, McAlpine, Olsen, Tsouknida, & Martin 2005; Russo & Grammatopoulou, 2003). If the word length effect is really due to differences in neighborhood size between short and long words, then concurrent articulation should also remove the neighborhood size effect. In Experiment 1, subjects saw a list of one-syllable words, half with large neighborhoods and half with small neighborhoods. Half of the subjects engaged in concurrent articulation during list presentation, and half did not.

## Method

**Subjects** A total of 32 undergraduates from Memorial University of Newfoundland volunteered to participate in exchange for a small honorarium. All were native speakers of English.

**Stimuli** The stimuli were the 32 three-phoneme CVC words from the low-neighborhood-frequency set in Experiment 3 of Roodenrys et al. (2002). Although initially selected for a manipulation of phonological neighborhood size—half had a large and half a small neighborhood size—the words also differed in terms of orthographic neighborhood size. Orthographic neighborhood size was calculated using the MCWord database (Medler & Binder, 2005), and these values were 3.8 for the small-neighborhood words and 12.6 for the large-neighborhood words.

Pronunciation time was measured for the small- and large-neighborhood words to ensure that any effect of neighborhood size found could not be attributable to differential articulation fluency. Following the procedure of Neath et al. (2003), 10 additional native speakers of English, who did not take part in the main experiment, were asked to repeat the lists 10 times out loud as fast as they

could and were recorded digitally. (These 10 subjects also provided the individual word pronunciation times in Exps. 2 and 3.) Individual word pronunciation times from the tenth repetition were measured using the Audacity program, which enables precise selection of each word. The small-neighborhood words took a mean of 563.33 ms to pronounce ( $SD = 40.45$ ), as compared to 563.03 ms ( $SD = 71.07$ ) for the large-neighborhood words; these values did not differ,  $t < 1$ .

Pronunciation time was also measured for the list as a whole, following the procedure of Woodward, Macken, and Jones (2008). On each trial, six words were randomly drawn from either the small- or the large-neighborhood pool and were presented simultaneously on the computer screen. There were 10 small-neighborhood lists and 10 large-neighborhood lists, in random order. A further 10 native speakers of English were asked to read the six-word lists out loud as quickly and as accurately as possible. (These 10 subjects also provided the total list pronunciation times in Exps. 2 and 3.) Total pronunciation time for the 10 small-neighborhood and the 10 large-neighborhood lists was then computed. The small-neighborhood lists took a mean of 2,663.09 ms to pronounce ( $SD = 327.30$ ), as compared to 2,691.84 ms ( $SD = 394.71$ ) for the large-neighborhood lists; these values did not differ,  $t < 1$ .

*Design and procedure* There were four types of lists: pure lists that contained only small-neighborhood words, pure lists that contained only large-neighborhood words, mixed lists that alternated small- and large-neighborhood words (i.e., small large small large small large), and mixed lists that alternated large- and small-neighborhood words (i.e., large small large small large small). There were 15 trials for each type of list, randomly ordered for each subject. Concurrent articulation was manipulated between subjects, and neighborhood size and list type were manipulated within subjects.

On each trial, six words were randomly selected from the pool and were presented at a rate of one item per second on a computer screen. At the end of list presentation, the six words from the current trial appeared as labels, in alphabetical order, on buttons on the computer screen, and subjects were asked to reconstruct the order in which the words were presented by clicking on the appropriately labeled buttons with the mouse. Subjects were asked to click on the first word first, the second word second, and so on. There was no time limit for recall. Once the subject had finished recalling the words, he or she clicked on a button on the computer to begin the next list.

Half of the subjects were asked to engage in concurrent articulation during the presentation of the items. They were

asked to say the letters *A, B, C, D, E, F, G* out loud during the presentation of the list of to-be-recalled words. Subjects were tested individually, and an experimenter was present throughout to ensure compliance with the instructions.

## Results

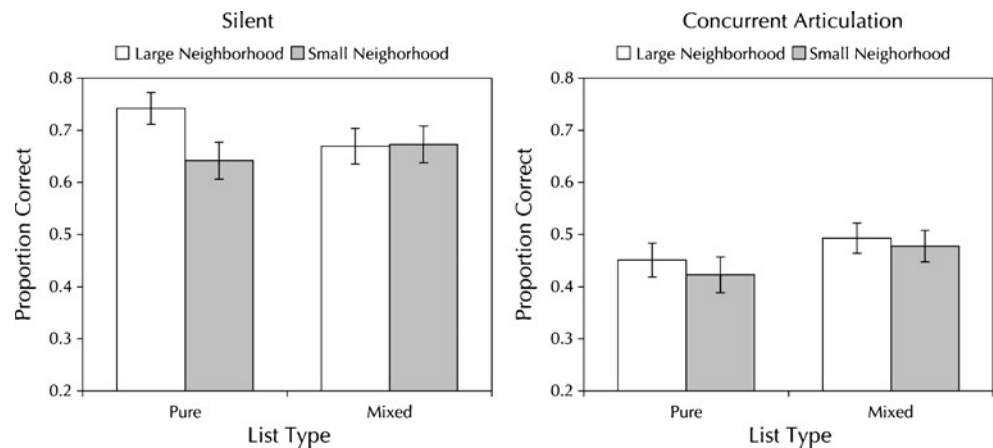
A word was considered correctly recalled if it was selected in the correct serial position. Following Hulme et al. (2004), derived lists for words from small and large neighborhoods presented in mixed lists were constructed. Thus, small-neighborhood words in mixed lists combined the first, third, and fifth words from the small large small large small large list and the second, fourth, and sixth words from the large small large small large small list. In this and all subsequent analyses, the .05 level of significance was adopted.

As can be seen in Fig. 1, in the silent condition large-neighborhood words in pure lists were recalled better than small-neighborhood words in pure lists, replicating the basic neighborhood size effect. Concurrent articulation eliminated this effect. For mixed lists, no neighborhood size effect was observed in either the silent or the concurrent articulation condition.

These trends were analyzed with a  $2 \times 2 \times 2$  mixed design ANOVA with Neighborhood Size (small vs. large) and List Type (pure vs. mixed) as within-subjects factors and Encoding Condition (silent vs. concurrent articulation) as a between-subjects factor. There was a significant main effect of neighborhood size,  $F(1, 30) = 12.665$ ,  $MSE = .003$ ,  $\eta_p^2 = .297$ , with words with large neighborhoods being better recalled than words with small neighborhoods (.590 vs. .554). The main effect of list type was not significant,  $F(1, 30) = 1.083$ ,  $MSE = .006$ ,  $\eta_p^2 = .035$ , with words from pure lists being recalled as well as words from mixed lists (.565 vs. .579). The main effect of encoding condition was significant,  $F(1, 30) = 26.378$ ,  $MSE = .059$ ,  $\eta_p^2 = .468$ , with recall being better in the silent condition than in the concurrent articulation condition (.682 vs. .461).

The interaction between neighborhood size and list type was significant,  $F(1, 30) = 24.014$ ,  $MSE = .001$ ,  $\eta_p^2 = .445$ , reflecting, in part, a difference in recall as a function of neighborhood size in pure but not in mixed lists. The interaction between list type and encoding condition was also significant,  $F(1, 30) = 6.636$ ,  $MSE = .006$ ,  $\eta_p^2 = .181$ , reflecting, in part, a difference between pure and mixed lists in the silent condition but no difference in the concurrent articulation condition. The interaction between neighborhood size and encoding condition failed to reach conven-

**Fig. 1** Proportions of words with large or small neighborhoods recalled from pure or mixed lists in the silent condition (left panel) and the concurrent articulation condition (right panel) for Experiment 1. Error bars show the standard errors of the means



tional levels of significance,  $F(1, 30) = 1.793$ ,  $MSE = .003$ ,  $\eta_p^2 = .056$ .

When interpreting the two-way interactions, it is important to keep in mind that the three-way interaction between neighborhood size, list type, and encoding condition was significant,  $F(1, 30) = 14.379$ ,  $MSE = .001$ ,  $\eta_p^2 = .324$ . This reflects the presence of a neighborhood size effect in pure but not mixed lists in the silent condition, which is then abolished by concurrent articulation. Consistent with this, Tukey HSD tests revealed a significant difference between recall of large- and small-neighborhood words in pure lists in the silent condition (.742 vs. .642), but no differences in any other condition (for mixed lists in the silent condition, .669 vs. .673; for pure lists in the concurrent articulation condition, .451 vs. .423; and for mixed lists in the concurrent articulation condition, .493 vs. .478, respectively).

## Discussion

If neighborhood size is an important factor in driving previous word length effects, then one should expect similar interactions between neighborhood size and the factors known to interact with word length. Experiment 1 found that a neighborhood size effect observed in pure lists was abolished by concurrent articulation, the same result seen with word length (e.g., Baddeley et al., 1975). This confirms the first prediction: Neighborhood size interacts with concurrent articulation in the same way that word length does. In addition, Experiment 1 replicated the finding that neighborhood size effects are observed only in pure lists, not in mixed lists. Again, the pattern resembles that most often seen with word length (Bireta et al., 2006).

These results are consistent with the claim that neighborhood size may have been the cause of previous demonstrations of the word length effect, because in those studies length and neighborhood size were confounded. If

the claim is accurate, the results previously attributed to differences in length should be observable with stimuli that do not differ in length, as long as the stimuli differ in neighborhood size.

It is difficult to explain these results from the perspective of the phonological loop framework, because concurrent articulation is thought to interfere with the articulatory control process. However, another way of thinking about concurrent articulation is as something that adds to the cognitive load by, for example, having to engage in a second activity and by adding noise to the to-be-remembered items (e.g., Murray, Rowan, & Smith 1988; Nairne, 1990; Neath, 2000). In pure lists, a large neighborhood may help recall by assisting with the redintegrative process (e.g., Jalbert et al., 2011; Roodenrys, 2009). For example, if one were to assume that the degraded cue serves as input to an interactive network, the slight activation in the network accruing from the commonalities of the neighbors—which by definition differ by only one letter—could readily lead to more successful redintegration of a target. In mixed lists, both small- and large-neighborhood items need identifying, which slightly helps the small-neighborhood items while slightly hurting the large-neighborhood items. The small-neighborhood items are helped by the removal (relative to the pure lists) of three additional harder to redintegrate items, whereas the large-neighborhood items benefit by the addition of three easier to redintegrate items. If concurrent articulation adds noise, the benefit conveyed by having a larger number of neighbors will be removed, thus lowering performance substantially for large-neighborhood items. However, small-neighborhood items never had much of a benefit from neighbors to begin with, so interfering with this process has little effect.

Regardless of the explanation, the confirmation of the first prediction supports the view that length may not be the

cause of the word length effect. We now turn to the second prediction: If we reverse the usual confounding of length and neighborhood size, such that the long words have large neighborhoods and the short words have small neighborhoods, would we still observe a neighborhood size effect? Unfortunately, one cannot use real words to test this, since there are not enough long words with large neighborhoods in the English language. Thus, we needed to use nonwords. Before doing so, however, it is necessary to demonstrate that the neighborhood size effect observed with nonwords is eliminated by concurrent articulation, just like the neighborhood size effect with words. After we have done that in Experiment 2, in Experiment 3 we will use nonwords to examine whether length or neighborhood size has the greater effect on recall.

## Experiment 2

The goal of Experiment 2 was to replicate the results from Experiment 1 with nonwords. Roodenrys and Hinton (2002) have already demonstrated a neighborhood size effect with nonwords, but we need to verify that, just as in Experiment 1, this effect is eliminated by concurrent articulation. Therefore, Experiment 2 was just like Experiment 1, except that the stimuli were a set of one-syllable nonwords, half with large neighborhoods and half with small neighborhoods.

### Method

**Subjects** Another 32 undergraduate students from Memorial University of Newfoundland volunteered to participate in exchange for a small honorarium. All subjects were native speakers of English, and none had participated in the previous experiment.

**Stimuli** A set of 24 nonwords (see Appendix A) was created using the orthographic word form database of Medler and Binder (2005). All of the nonwords consisted of one syllable and all contained five letters. Half of the nonwords had a large neighborhood size and half had a small neighborhood size (26.25 vs. 6.58). Pronunciation time was again measured for the small- and the large-neighborhood nonwords using the same extraexperimental subjects as in Experiment 1. The mean pronunciation time for small-neighborhood words was 553.32 ms ( $SD = 54.57$ ), as compared to 544.16 ms ( $SD = 37.66$ ) for large-neighborhood words; these values did not differ,  $t < 1$ . Pronunciation time was also computed for the entire lists using the same procedure and subjects as for Experiment 1. There was no difference in pronunciation time between large-neighborhood and small-neighborhood non-

words: 3,838.48 ms ( $SD = 737.60$ ) versus 3,865.96 ms ( $SD = 754.91$ ), respectively,  $t < 1$ .

**Design and procedure** The design and procedure were the same as in Experiment 1, except for the use of nonwords instead of words.

### Results and discussion

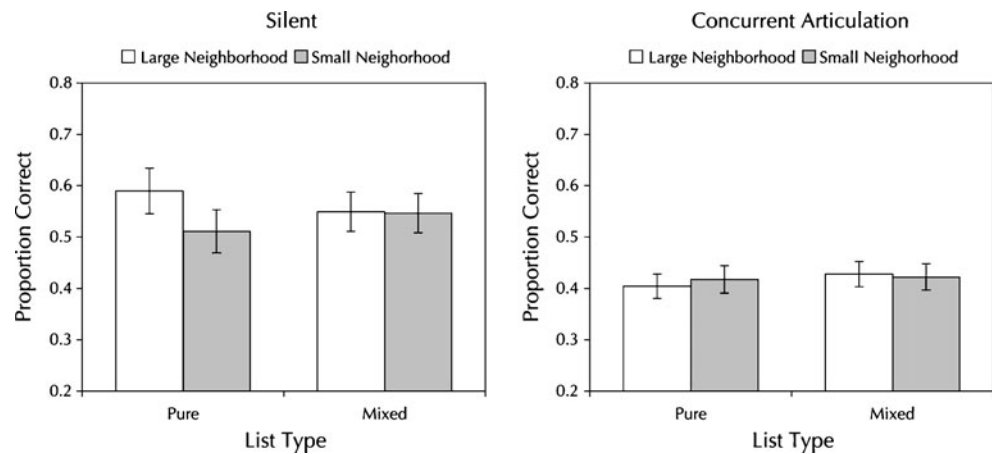
As can readily be seen, Fig. 2 looks just like Fig. 1, despite the change from words to nonwords. In the silent condition, large-neighborhood nonwords in pure lists were recalled better than small-neighborhood nonwords, replicating the basic neighborhood size effect. Concurrent articulation eliminated this effect, just as it did for words. In the mixed lists, no neighborhood size effect was observed in either the silent or the concurrent articulation condition.

These trends were analyzed with a  $2 \times 2 \times 2$  mixed design ANOVA with Neighborhood Size (small vs. large) and List Type (pure vs. mixed) as within-subject factors and Encoding Condition (silent vs. concurrent articulation) as a between-subjects factor. Unlike in Experiment 1, the main effect of neighborhood size did not reach the adopted significance level,  $F(1, 30) = 3.410$ ,  $MSE = .003$ ,  $p = .075$ ,  $\eta_p^2 = .102$ . The proportion of nonwords with large neighborhoods correctly recalled was .493, as compared to .474 for those with small neighborhoods. The main effect of list type was not significant,  $F < 1$ , with approximately equivalent recall in pure and mixed lists (.481 vs. .486, respectively). There was a significant main effect of encoding condition,  $F(1, 30) = 8.786$ ,  $MSE = .063$ ,  $\eta_p^2 = .227$ , with better recall performance in the silent condition than in the concurrent articulation condition (.549 vs. .418, respectively).

Neither the interaction between neighborhood size and list type,  $F(1, 30) = 2.245$ ,  $MSE = .018$ ,  $\eta_p^2 = .070$ , nor the interaction between list type and encoding condition,  $F < 1$ , was significant. However, the interaction between neighborhood size and encoding condition did reach conventional levels of significance,  $F(1, 30) = 4.973$ ,  $MSE = .003$ ,  $\eta_p^2 = .142$ . This reflects a difference in recall of nonwords from large and small neighborhoods in the silent condition (.569 vs. .529) but no difference in the concurrent articulation condition (.416 vs. .420).

When interpreting the two-way interactions, it is important to keep in mind that the three-way interaction between neighborhood size, list type, and encoding condition was significant,  $F(1, 30) = 6.175$ ,  $MSE = .003$ ,  $\eta_p^2 = .171$ . This reflects the presence of a neighborhood size effect in pure but not mixed lists in the silent condition, which is then abolished by concurrent articulation. Consistent with this,

**Fig. 2** Proportions of nonwords with large or small neighborhoods recalled from pure or mixed lists in the silent condition (left panel) and the concurrent articulation condition (right panel) for Experiment 2. Error bars show the standard errors of the means



and just as was observed in Experiment 1, Tukey HSD tests revealed a significant difference between recall of large- and small-neighborhood nonwords in pure lists in the silent condition (.590 vs. .511), but no differences in any other condition (for mixed lists in the silent condition, .549 vs. .547; for pure lists in the concurrent articulation condition, .404 vs. .417; and for mixed lists in the concurrent articulation condition, .428 vs. .422, respectively).

There were some slight differences in the particular patterns of significant interactions between Experiments 1 and 2, but nonwords do sometimes result in a slightly different pattern than do words (see, e.g., Romani et al., 2005). The major results of both experiments, however, are the same: (1) A neighborhood size effect is seen in pure lists but not mixed lists in the silent condition, and (2) this effect is removed by concurrent articulation. Once again, the results—this time with nonwords—parallel those observed with manipulations of word length.

### Experiment 3

The purpose of Experiment 3 was to use a completely factorial design to study word length and neighborhood size—that is, to compare short items with a small neighborhood, short items with a large neighborhood, long items with a small neighborhood, and long items with a large neighborhood. While the ideal experiment would use words, there are not enough suitable long words in the English language that have large neighborhoods. Given the similarity in the results found in Experiments 1 and 2, we therefore used nonwords. If neighborhood size drives the word length effect, there should be better recall of nonwords with large neighborhoods than with small

neighborhoods, regardless of their length. If length drives the word length effect, there should be better recall of short than of long nonwords.

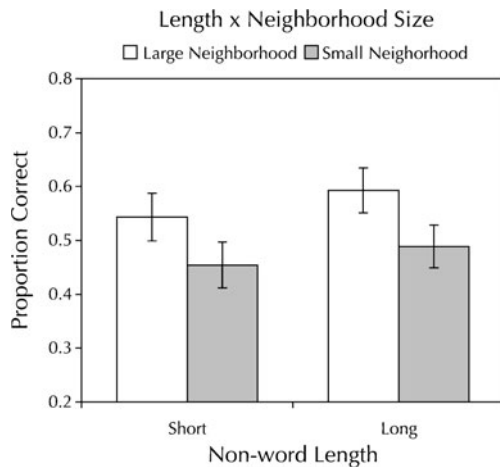
### Method

**Subjects** A group of 16 undergraduate students from Memorial University of Newfoundland participated in exchange for a small honorarium. All were native speakers of English and none had participated in Experiment 1 or 2.

**Stimuli** A set of 48 nonwords (see Appendix B) was created using the orthographic word form database of Medler and Binder (2005). Half were short (monosyllabic) and half were long (disyllabic). In addition, half had a small neighborhood size (0 neighbors) and half had a large neighborhood size. Pronunciation times were again measured using the same extraexperimental subjects as in Experiments 1 and 2. There was no difference in pronunciation times as a function of neighborhood size for short nonwords (488.48 ms,  $SD = 43.26$ , for small vs. 485.66 ms,  $SD = 58.96$ , for large;  $t < 1$ ) or for long nonwords (566.01 ms,  $SD = 63.95$ , for small vs. 562.12 ms,  $SD = 69.23$ , for large;  $t < 1$ ). However, when collapsed over neighborhood size, the short and long words differed significantly in pronunciation time: 487.07 ms ( $SD = 50.35$ ) versus 564.07 ms ( $SD = 62.35$ ),  $t(9) = 6.01$ ,  $p < .001$ . This difference is of the same order of magnitude as in studies investigating the time-based word length effect (e.g., Neath et al., 2003).

Pronunciation times were also computed for the entire lists using the same procedure and subjects as in Experiments 1 and 2. There was no difference in the pronunciation times for large-neighborhood and small-neighborhood nonwords for the short nonwords, 3,462.80 ms ( $SD = 827.02$ ) versus





**Fig. 3** Proportions of short and long nonwords with large or small neighborhoods correctly recalled for Experiment 3. Error bars show the standard errors of the means

3,609.27 ms ( $SD = 868.30$ ), respectively,  $t(9) = 1.25$ ,  $p = .24$ . There was also no difference in pronunciation times for large- and small-neighborhood nonwords for the long nonwords, 3,930.74 ms ( $SD = 1,146.47$ ) versus 3,861.47 ms ( $SD = 985.69$ ), respectively,  $t < 1$ . When collapsed over neighborhood size, the short and long nonwords differed significantly in pronunciation time, 3,536.04 ms ( $SD = 828.71$ ) versus 3,896.11 ms ( $SD = 1,041.20$ ),  $t(19) = 3.64$ ,  $p < .01$ .

**Design and procedure** The design and procedure were similar to those in Experiments 1 and 2, except for the following: (1) There was no concurrent articulation task and (2) there were no mixed lists. Length (short vs. long) and orthographic neighborhood size (small vs. large) were within-subjects variables. Each type of list (i.e., short length/small neighborhood, short length/large neighborhood, etc.) was presented 15 times; the order of the lists was randomized for each subject.

## Results and discussion

As a manipulation check, we first compared the recall of short nonwords with a large neighborhood to that of long nonwords with a small neighborhood. According to Jalbert et al. (2011), this corresponds to the stimuli used in typical word length studies. The short items should be better recalled than the long items, and indeed they were: The difference was .543 versus .490, significant by a Tukey HSD test.

As is shown in Fig. 3, recall was related to neighborhood size rather than length. The data were analyzed with a  $2 \times 2$

repeated measures ANOVA with neighborhood size (small vs. large) and length (short vs. long) as within-subjects factors. There was a main effect of neighborhood size,  $F(1, 15) = 25.371$ ,  $MSE = .006$ ,  $\eta_p^2 = .628$ , with better recall of nonwords with large neighborhoods than of nonwords with small neighborhoods (.568 vs. .472, respectively). The main effect of length was not significant,  $F(1, 15) = 3.209$ ,  $MSE = .009$ ,  $p > .09$ ,  $\eta_p^2 = .389$ . Although the difference was not significant, the trend was, if anything, for slightly better recall of the longer than of the shorter nonwords, .541 versus .499. The interaction between neighborhood size and length was not significant,  $F < 1$ .

Because there was no effect of length, it is possible that subjects used a strategy in which they focused only on the first letters of each nonword rather than on the entire nonword. If this strategy were adopted, a list with nonwords sharing the same first letter (i.e., *farnza*, *fidir*, *nublay*, *nusen*) would be harder to recall than a list in which all items began with a different first letter (i.e., *agald*, *fidir*, *nublay*, *riridy*). However, it is unlikely that this strategy was used here, because it would have also resulted in the absence of a neighborhood size effect. Focusing on just the first letters would remove both main effects, rather than selectively removing the effect of length.

As with words, short nonwords that follow the general rules of English have more neighbors than otherwise comparable long nonwords. Unlike words, however, there are a sufficient number of nonwords to make it possible to manipulate length and neighborhood size factorially. When this is done, two results stand out: (1) Only neighborhood size had a measurable effect on the proportion of items correctly recalled, and (2) short-length/large-neighborhood items are recalled better than long-length/small-neighborhood items. The latter finding corresponds to the typical manipulation of word length in the literature, in which length and neighborhood size are confounded.

## General discussion

The present experiments tested two predictions that arise from the claim that neighborhood size, rather than length per se, mediates the word length effect. If previous demonstrations of the word length effect were caused by comparing short items from large neighborhoods with long items from small neighborhoods, then (1) concurrent articulation should remove the neighborhood size effect, just as it removes the word length effect, and (2) long words with a larger neighborhood

should be recalled better than short words with a smaller neighborhood.

Experiment 1 showed that the neighborhood size effect observed in the silent condition was abolished in the concurrent articulation condition. Moreover, the neighborhood size effect was apparent only in the pure lists, not in the mixed lists. Both of these findings parallel those most often seen with word length. Experiment 2 replicated the major results of Experiment 1: Concurrent articulation also abolishes the neighborhood size effect for nonwords. Finally, Experiment 3 used a completely factorial design to assess length and neighborhood size, and found a main effect of neighborhood size and no effect of length.

Given these results, and those of Jalbert et al. (2011), the most plausible explanation of the word length effect is that it is not caused by length per se, but rather by some property correlated with length, such as neighborhood size. Neighborhood size is a better predictor of performance than is word length, but it is likely that other lexical or linguistic factors may be important as well. Consideration of such factors may also explain why so many of the results involving word length critically depend on the particular stimulus set used.

One possible concern is that the word length effect was attenuated by the recall methodology. More specifically, a proponent of the phonological loop might argue that visual presentation and reconstruction of order could diminish the size of the word length effect, because it is explained by articulation time. Jalbert et al. (2011) tested this possibility by comparing the recall patterns with short and long words using written recall and reconstruction of order. There was no difference in the recall pattern as a function of the test (Jalbert et al., 2011, Exp. 1). Furthermore, the absence of a word length effect when short and long words were equated for neighborhood size had previously been demonstrated using a spoken serial recall task (Jalbert et al., 2011, Exp. 5). In addition, in the present Experiment 3, there was an effect of word length when neighborhood size was confounded with it, as is typically done in word length studies. Therefore, these factors do not appear to be critical.

A second concern may be that because part of our argument is correlational in nature (i.e., emphasizing the similar effect of concurrent articulation on both word length and neighborhood size manipulations), the tests of our thesis are not particularly strong. This concern is only partly warranted. We acknowledge that finding that concurrent articulation abolishes the neighborhood size effect does not necessarily mean that it is the same as word length effect. However, had we failed to find that

concurrent articulation abolishes the neighborhood size effect, then our thesis would have been falsified. It was a distinct possibility that neighborhood size might be like manipulations of concreteness, frequency, imageability, and word class, which are immune to concurrent articulation. Thus, this study is a strong test of the hypothesis.

The observation that having more neighbors helps recall performance may appear surprising, because having a large neighborhood can be detrimental for certain tasks, such as spoken word recognition (see, e.g., Luce & Pisoni, 1998). However, having a large neighborhood size can help on tasks that require the production of the words from memory. For example, Vitevitch (2002) showed that subjects made more errors for words with fewer similar-sounding words (i.e., small neighborhood) than for words with more similar-sounding words (i.e., large neighborhood) in a speech production task. Similarly, words from small neighborhoods are identified more slowly than large-neighborhood words in picture-naming tasks (see also Vitevitch & Sommers, 2003). This is the reasoning behind placing the facilitative effects of neighborhood size at output: Speech production but not perception is enhanced by increased the number of neighbors.

These results are problematic for any version of the so-called “standard model” (see Nairne, 2002, and Surprenant & Neath, 2009, for reviews) in which decay is offset by rehearsal and concurrent articulation has its effect by blocking rehearsal. Perhaps the most well-known of these models is the phonological loop component of working memory. According to this account, “memory traces decay over a period of a few seconds, unless revived by articulatory rehearsal” (Baddeley, 2000, p. 419). This view predicts both a time-based word length effect and a syllable-based time word length effect, both of which are abolished (or greatly attenuated) by concurrent articulation. Within this framework, concurrent articulation is seen as preventing or interfering with articulatory rehearsal, which prevents the decaying traces from being refreshed. The problem for this type of account is explaining why there are sometimes no effects of word length and why concurrent articulation eliminates the neighborhood size effect. As an interesting side note, these data suggest that perhaps the effects of concurrent articulation may not be at the level of rehearsal, but rather may interfere with output of the words, perhaps interfering with output production mechanisms. However, this theory is purely speculative at this point and will need empirical investigation to support it.

The results are less problematic for some of the item-based accounts. For example, within the context of the

feature model, concurrent articulation has always been viewed as adding noise to the memory trace (Nairne, 1990; see also Murray et al., 1988). If this is the case, it is easy to explain the present effects. Roodenrys (2009; see also Roodenrys & Miller, 2008) suggested that the locus of the neighborhood size effect is during redintegration. If the degraded items in memory serve as input to an interactive network, activation in the network from the item's neighbors could lead to more successful redintegration of the to-be remembered items. In other words, the more neighbors you have, the more activation you will get in an interactive memory network and the easier the items will be to recall. However, if noise is added by having the task performed along with concurrent articulation, this could remove the benefit for the large-neighborhood items by reducing activation levels.

In addition, these effects can easily be handled by language-based models of short-term memory, including a number of different types of interactive activation models (e.g., Martin et al., 1999; McClelland & Rumelhart, 1981). In particular, the Martin et al. model includes separate input and output buffers that are connected only through the long-term knowledge structure. In this model, the different representations can be affected by different variables, thus accounting for the opposing effects neighborhood size has on speech perception and production. Thus, the redintegration argument put forth here would predict that concurrent articulation has its effect at the level of the output process in that model. Much of the data supporting this model come from individuals with various forms of brain damage, resulting in perception and/or production difficulties. It remains to be seen whether neighborhood size is a variable that shows similar effects in patient populations.

The word length effect has been termed one of the “benchmark findings” that models of short-term memory must account for (Lewandowsky & Farrell, 2008) and has greatly influenced the development of many theories of memory. However, the time-based word length effect occurs only with one set of stimuli, and Jalbert et al. (2011) suggested that past demonstrations of the syllable-based word length effect included a confound: The short words had more neighbors than the long words. If neighborhood size was driving previous demonstrations of the word length effect, two predictions would follow. The results of Experiment 1 confirm the first prediction, that concurrent articulation should remove the neighborhood size effect, and the results of Experiment 3 confirm the second, that when length and neighborhood size are factorially manipulated, size will be a factor but length

will not. These results are problematic for any theory of memory that includes decay offset by rehearsal, but they are consistent with accounts that include a redintegrative stage that is susceptible to disruption by noise.

**Author note** Portions of this article were presented at the 20th Annual Meeting of the Canadian Society for Brain, Behaviour and Cognitive Science, Halifax, Nova Scotia, June 2010. This research was supported by grants from NSERC to all three coauthors.

## Appendix A

### Nonwords used in Experiment 2

	Orthographic Neighborhood Size	Orthographic Neighborhood Frequency
Small Neighborhood		
chush	2	4.55
googe	3	4.96
grair	2	14.34
joach	3	12.20
jorth	3	95.78
nadge	4	2.29
olled	2	1.67
rorch	3	9.76
tedge	4	6.95
touge	3	12.16
zarsh	2	13.98
zoule	1	0.36
Mean	2.67	14.92
SD	0.89	25.93
Large Neighborhood		
boarg	2	42.95
chone	8	17.26
coose	8	37.21
gares	18	6.51
ghoss	3	15.25
jjight	10	256.62
korch	2	11.24
lorse	8	28.76
petch	8	7.74
puice	2	56.55
sheed	9	20.92
wroke	2	98.73
Mean	6.67	49.98
SD	4.79	70.12

## Appendix B

### Short and long nonwords used in Experiment 3

	Orthographic Neighborhood Size	Orthographic Neighborhood Frequency		Orthographic Neighborhood Size	Orthographic Neighborhood Frequency
Short/Small Neighborhood			Long/Small Neighborhood		
clende	0	0	aftin	0	0
colmes	0	0	agald	0	0
greng	0	0	banays	0	0
gruld	0	0	blatis	0	0
keage	0	0	civor	0	0
kese	0	0	colut	0	0
sheng	0	0	farnza	0	0
smond	0	0	fidir	0	0
thech	0	0	nublay	0	0
thoule	0	0	nusen	0	0
varre	0	0	rirdy	0	0
vefe	0	0	romir	0	0
Mean	0	0	Mean	0	0
SD	0	0	SD	0	0
Short/Large Neighborhood			Long/Large Neighborhood		
beath	5	64.15	afted	3	404.09
coust	5	52.88	entes	2	23.92
feem	7	126.17	givet	5	84.78
feen	8	473.09	hully	8	11.33
filld	4	36.28	lany	9	154.48
kife	6	160.33	lother	5	96.57
plat	7	71.89	nevel	7	161.22
plur	6	12.57	piter	3	21.81
sach	6	271.09	rever	8	132.62
smate	6	52.45	staved	9	18.61
whone	6	122.18	stily	3	308.11
yoom	5	100.09	tiver	8	17.66
Mean	5.92	128.60	Mean	5.83	119.60
SD	0.51	128.53	SD	0.51	125.18

## References

- Acheson, D. J., Postle, B. R., & MacDonald, M. C. (2010). The interaction of concreteness and phonological similarity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 17–36.
- Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language*, *55*, 64–88.
- Avons, S. E., Wright, K. L., & Pammer, K. (1994). The word-length effect in probed and serial recall. *The Quarterly Journal of Experimental Psychology*, *47A*, 207–231.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
- Baddeley, A. (1992). Is working memory working? The Fifteenth Bartlett Lecture. *The Quarterly Journal of Experimental Psychology*, *44A*, 1–31.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.
- Baddeley, A. D., Chincotta, D., Stafford, L., & Turk, D. (2002). Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition. *The Quarterly Journal of Experimental Psychology*, *55A*, 353–369.
- Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology*, *36A*, 233–252.
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575–589.

- Bhatarah, P., Ward, G., Smith, J., & Hayes, L. (2009). Examining the relationship between free recall and immediate serial recall: Similar patterns of rehearsal, and similar effects of word length, presentation rate, and articulatory suppression. *Memory & Cognition*, *37*, 689–713.
- Bireta, T. J., Neath, I., & Surprenant, A. M. (2006). The syllable-based word length effect and stimulus set specificity. *Psychonomic Bulletin & Review*, *13*, 434–438.
- Bourassa, D. C., & Besner, D. (1994). Beyond the articulatory loop: A semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin & Review*, *1*, 122–125.
- Brown, G. D. A., & Hulme, C. (1995). Modeling item length effects in memory span: No rehearsal needed? *Journal of Memory and Language*, *34*, 594–621.
- Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale: Erlbaum.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.
- Cowan, N., Baddeley, A. D., Elliot, E. M., & Norris, J. (2003). List composition and the word length effect in immediate recall: A comparison of localist and globalist assumptions. *Psychonomic Bulletin & Review*, *10*, 74–79.
- Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*, *31*, 1–17.
- Gregg, V. H., Freedman, C. M., & Smith, D. K. (1989). Word frequency, articulatory suppression and memory span. *British Journal of Psychology*, *80*, 363–374.
- Hulme, C., Surprenant, A. M., Bireta, T. J., Stuart, G., & Neath, I. (2004). Abolishing the word-length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 98–106.
- Jalbert, A., Neath, I., Bireta, T. J., & Surprenant, A. M. (2011). When does length cause the word length effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 338–353.
- LaPointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 1118–1133.
- Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 1–48). San Diego: Academic Press.
- Longoni, A., Richardson, J. T. E., & Aiello, A. (1993). Articulatory rehearsal and phonological storage in working memory. *Memory & Cognition*, *21*, 11–22.
- Lovatt, P., Avons, S. E., & Masterson, J. (2000). The word-length effect and disyllabic words. *The Quarterly Journal of Experimental Psychology*, *53A*, 1–22.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1–36.
- Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 122–147). Cambridge: MIT Press.
- Martin, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, *41*, 3–29.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*, 375–407. doi:10.1037/0033-295X.88.5.375.
- Medler, D. A., & Binder, J. R. (2005). MCWord: An on-line orthographic database of the English language. Available at [www.neuro.mcw.edu/mcword/](http://www.neuro.mcw.edu/mcword/).
- Murray, D. J., Rowan, A. J., & Smith, K. H. (1988). The effect of articulatory suppression on short-term recognition. *Canadian Journal of Psychology*, *42*, 424–436.
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251–269.
- Naime, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, *53*, 53–81.
- Nairne, J. S., Neath, I., & Serra, M. (1997). Proactive interference plays a role in the word-length effect. *Psychonomic Bulletin & Review*, *4*, 541–545.
- Neath, I. (2000). Is “working memory” still a useful concept? *Contemporary Psychology*, *45*, 410–412.
- Neath, I., & Nairne, J. S. (1995). Word-Length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review*, *2*, 429–441.
- Neath, I., Bireta, T. J., & Surprenant, A. M. (2003). The time-based word length effect and stimulus set specificity. *Psychonomic Bulletin & Review*, *10*, 430–434.
- Romani, C., McAlpine, S., Olsen, A., Tsouknida, E., & Martin, R. C. (2005). Length, lexicality, and articulatory suppression in immediate recall: Evidence against the articulatory loop. *Journal of Memory and Language*, *52*, 398–415.
- Roodenrys, S. (2009). Explaining phonological neighbourhood effects in short-term memory. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 177–197). Hove, East Sussex: Psychology Press.
- Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 29–33.
- Roodenrys, S., & Miller, L. M. (2008). A constrained Rasch model of trace reintegration in serial recall. *Memory & Cognition*, *36*, 578–587.
- 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.
- Russo, R., & Grammatopoulou, N. (2003). Word length and articulatory suppression affect short-term and long-term recall tasks. *Memory & Cognition*, *31*, 728–737.
- Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *The Quarterly Journal of Experimental Psychology*, *51A*, 283–304.
- Surprenant, A. M., & Neath, I. (2009). The nine lives of short-term memory. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 16–43). Hove, East Sussex: Psychology Press.
- 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.
- Tehan, G., Hendry, L., & Kocinski, D. (2001). Word length and phonological similarity effects in simple, complex, and delayed serial recall tasks: Implications for working memory. *Memory*, *9*, 333–348.
- Vitevitch, M. S. (2002). The influence of phonological similarity neighborhood on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 735–747.
- Vitevitch, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition*, *31*, 491–504.
- Watkins, M. J. (1972). Locus of the modality effect in free recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 644–648.
- 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.