WHEN DOES LENGTH CAUSE THE WORD LENGTH EFFECT?

ANNIE JALBERT

001311

CENTRE FOR NFLD. STUDIES
JUN 18 1978
MEMORIAL UNIVERSITY
OF NEWFOUNDLAND

WHEN DOES LENGTH CAUSE THE WORD LENGTH EFFECT?

by

Annie Jalbert

A thesis submitted to the

School of Graduate Studies

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

Psychology Department

Memorial University of Newfoundland

August 22nd 2011

Abstract

The word length effect -- the finding that lists of short words are better recalled than lists of long words -- has been termed one of the benchmark findings that any theory of immediate memory must address. The effect is viewed as the best remaining evidence for time-based decay of information in short-term memory. However, previous studies investigating this effect have confounded word length with orthographic neighborhood size. I suggest here that the word length effect may be better explained by the differences in lexical properties of short and long words than by length. Experiments 1a and 1b revealed typical effects of length when short and long words were equated on all relevant dimensions except for neighborhood size. Experiments 2 and 3 showed that when short and long words were equated for neighborhood size, the word length effect disappeared. Experiment 4 replicated the disappearance of the word length effect with spoken recall. In Experiment 5, one-syllable words with a large neighborhood were recalled better than one-syllable words with a small neighborhood. Experiment 6 found that concurrent articulation removed the effect of neighborhood size, just as it removes the effect of word length. Experiment 7 demonstrated that this pattern is also found with nonwords. In Experiment 8, length and neighborhood size were manipulated and only effects of the latter were found. These results are problematic for any theory of memory that includes decay offset by rehearsal, but 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. These results add to the growing literature identifying problems for theories of memory that include decay offset by rehearsal as a central feature.

Acknowledgements

First, I want to thank my advisor, Dr. Ian Neath, for his support, advice, and guidance that has been of great value to me during my doctorate at Memorial University and made this thesis possible. My deepest gratitude is also due to both members of my supervisory committee, Dr. Catherine Penney and Dr. Aimée M. Surprenant, without whose knowledge, and assistance this series of experiments would not have been successful. A special thanks to past and present members of the Cognitive Aging and Memory Laboratory, Caroline Barnes, Chris Bishop, Mark Brown, Roberta DiDonato, Brittany Faux, Martina Greene, Marlena Hickey, Sophie Kenny, and James March, for your moral support throughout my doctorate, and in particular, to Caroline Barnes for your assistance with running participants. I would also like to thank the Natural Science and Engineering Research Council of Canada for their financial assistance for the duration of my doctorate. Last but not least, I want to thank my family and friends for all your much appreciated understanding and mental support during my doctorate.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

Introduction

1.1 The Word Length Effect

The word length effect -- the finding that lists of short words (e.g., lead, pig, grape) are recalled better than lists of long words (e.g., aluminum, elephant, banana) -- has played such a significant a role in the development of theories of memory that it is now regarded as a "benchmark finding" that current theories of short-term or working memory must address (Lewandowsky & Farrell, 2008). Indeed, the basic finding is one of the core phenomena that led directly to the development of the phonological loop component of working memory (Baddeley, 1992). It has been termed the "best remaining solid evidence" for the existence of such temporary memory systems (Cowan, 1995, p. 42), and is the focus of many computational models (e.g., Brown & Hulme, 1995; Burgess & Hitch, 1999; Neath & Nairne, 1995; Page & Norris, 1998; Hulme, Surprenant, Bireta, Stuart, & Neath, 2004). Here, I consider evidence that questions the idea that length *per se* is the critical factor underlying the word length effect.

1.2 Word Length and Working Memory

Although the basic finding was known earlier (e.g., Watkins, 1972), the first systematic exploration of the word length effect was reported by Baddeley, Thomson, and Buchanan (1975). They reported three key results. First, a set of words was created in which the short and long items differed in pronunciation time but were equated for number of syllables, number of phonemes, and frequency. More short words were recalled on an immediate spoken serial recall test than long words. This is now referred to as the *time-*

based word length effect as 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 which varied in both pronunciation time and in the number of phonemes and/or syllables. One to 5-syllable words from the same semantic category were used (e.g., Maine, Utah, Wyoming, Alabama, Louisiana). Again more short words were recalled than longer words. This finding is known as the *syllable-based word length effect*. The third key finding was that both types of word length effects were removed if participants engaged in concurrent articulation, repeatedly saying the digits 1 to 8 outloud at an approximate rate of three digits per second, during list presentation. I use the term *concurrent articulation* rather than the more usual articulatory suppression because the former is a neutral description of what the participant is asked to do. In contrast, the latter term implies a specific effect of the manipulation and I will argue for a different effect of this manipulation later in this thesis.

According to Baddeley's working memory framework (Baddeley, 1986, 1992, 2000), the time-based word length effect, the syllable-based word length effect, and the abolishment of both word length effects with concurrent articulation all reflect the operation of the phonological loop. The to-be-remembered words enter the phonological store and decay after about two seconds if the articulatory control process does not refresh them. The articulatory control process is a subvocal rehearsal loop that counteracts the decay of information in the phonological store. 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 are available to be recalled compared to short words. Concurrent articulation is assumed under this account to prevent the use of the articulatory control process so neither short nor long items can be refreshed, making recall performance for short words equivalent to recall performance for long words.

1.3 The Time-Based Word Length Effect

The time-based word length effect was established in two initial studies. In their Experiment 3, Baddeley et al. (1975) showed that lists of disyllabic words that could be said quickly (*bishop, pectin, ember, wicket, wiggle, pewter, tipple, hackle, decor, phallic*) were recalled better than lists of disyllabic words that took longer to pronounce (*Friday, coerce, humane, harpoon, nitrate, cyclone, morphine, tycoon, voodoo, zygote*). In Experiment 4, a subset of these words was used such that the short and long words were equated for the number of syllables, the number of phonemes (given Scottish pronunciation), and frequency. Once again, a word length effect was obtained: Words that took less time to say were recalled better than the words that took more time to say. Since it takes longer for the articulatory control process to refresh a list of words that takes longer to pronounce, they are more prone to forgetting than a list of words that takes less time to pronounce. These results were taken as support for the phonological loop component of working memory (Baddeley, 1986).

Many studies have since replicated this time-based word length effect using the original stimuli (e.g., Cowan, Day, Saults, Kellar, Johnson, & Flores, 1992; Longoni, Richardson, & Aiello, 1993; Lovatt, Avons, & Masterson, 2000; Nairne, Neath, & Serra, 1997). However, there are no other sets of stimuli that produce this result. For example, Neath, Bireta, and Surprenant (2003) tested four different sets of short and long words that were equated for the number of syllables and phonemes, but differed in pronunciation time: Only the original Baddeley et al. (1975) stimuli produced a word length effect. An additional set of English words (Lovatt et al., 2000) and a set of Finnish nonwords (Service, 1998) also failed to yield a time-based word length effect. Thus, whereas one set of words does consistently produce the effect, five other sets of stimuli do not. Neath et al. (2003) concluded that the time-based word length effect was due to some unknown property of the original stimuli. They noted that unless a large number of other stimulus sets were shown to result in a time-based word length effect, it was reasonable to conclude that the effect does not exist. As Neath et al. (2003) pointed out, the absence of a time-based word length effect when using any other words than those used by Baddeley et al. (1975) poses a problem for theories that incorporate something like the phonological loop. Proponents of the phonological loop hypothesize a positive correlation between pronunciation time and the rate of rehearsal. Words decay in the phonological store after two seconds unless they are rehearsed. A list of words that takes longer to pronounce should always be recalled worse than a list of words that takes less time to pronounce because more "long" words will have time to decay before they can be refreshed by the articulatory control process.

1.4 The Syllable-Based Word Length Effect

A syllable-based word length effect is observed when words differ on both the number of syllables and the time it takes to pronounce them. In contrast to the time-based word length effect, the syllable-based word length effect is robust and has been demonstrated with numerous different sets of stimuli and a large variety of tasks including reconstruction of order (Neath et al., 2003), serial recognition (Baddeley, Chincotta, Stafford, & Turk, 2002), free recall (Watkins, 1972), single-item probe recall (Avons, Wright, & Pammer, 1994), and complex span (Tehan, Hendry & Kocinski, 2001). However, there are still disagreements about the cause of this effect. The following sections will outline the different models that have been proposed to account for the word length effect.

1.4.1 Phonological Loop Models – List-Based Models

One class of theories, based on the phonological loop, invokes an explanation based on the trade-off between decay and pronunciation time (e.g., Burgess & Hitch, 1999, 2006; Page & Norris, 1998, 2003), the lack of a pure time-based effect notwithstanding. According to the phonological loop explanation, words are hypothesized to decay in the phonological loop after about two seconds if they are not rehearsed. Forgetting occurs when the time it takes to rehearse the words is longer than the decay rate. Since long words take longer to pronounce than short words, it takes longer to rehearse a list of long words and they are more susceptible to forgetting. Accordingly, long words will not be recalled as well as short words. Models based on the phonological loop predict both a time-based and a syllable-based word length effect. Concurrent articulation prevents the use of the articulatory control process to refresh the memory traces in the phonological store. Neither short nor long items can be refreshed. Recall performance for short words would then be equivalent to recall performance for long words since they do not have the rehearsal advantage anymore.

To generate evidence in support of this view, researchers began examining recall of short and long items in pure lists (i.e., those made up of only short or only long items) and mixed lists, in which equal numbers of short and long items occurred. Using a computational model that incorporates the assumptions of the phonological loop, Burgess and Hitch (1999; Figure 16) generated the prediction that recall of lists made up of a mixture of short and long words would fall in between that of pure short and pure long lists. The list that can be rehearsed most quickly, the pure short list, will be recalled best, and the list that takes the longest amount of time to rehearse, the pure long list, will be recalled worst. The mixed lists take less time to rehearse than the pure long lists, but more time than the pure short lists, and so recall level will be intermediate.

Of relevance to the current thesis, phonological loop models make four predictions. First, for pure lists of all short or all long words, a word length effect will be observed, with short words being better recalled than long words. Second, for mixed lists of alternating short and long words, recall performance for short words will be equivalent to recall performance for long words. Since mixed lists take more time to rehearse than lists of short words, but less time to rehearse than lists of long words, recall level for mixed lists will fall between recall performance for pure short lists and pure long lists. Third, concurrent articulation will abolish rehearsal for both short and long words, making recall of short and long words equivalent. Fourth, since phonological loop models explain the word length effect by the trade-off between decay and pronunciation time, the same pattern of results stated in predictions 1, 2, and 3 should also be observed with pronounceable nonwords

1.4.2 Item-Based Models

In contrast to a model based on the phonological loop, theories based on the properties of individual items make quite different predictions. In the following section, three item-based models will be described: The Feature Model, the Brown and Hulme (1995) Model, and the Scale Invariant Memory, Perception, and Learning Model (SIMPLE). Other item-based models exist that include an explanation of the syllable-based word length effect but the following three were selected because they make clear-cut predictions about the effect of length on recall and because they have been adapted into computational models.

1.4.2.1 Feature Model

The Feature Model (Nairne, 1988, 1990) assumes that items are represented as a set of features called vectors. After the presentation of a list of words, the mnemonic representation of those words resembles degraded vectors, or traces. In order to be recalled properly, these traces need to be reassembled using long-term memory information. The more segments there are, the more chances of committing a re-assembly error. Since long words have more segments that need to be reassembled than short words, there is a greater chance of committing an error for long words. Consequently, short words will always be better recalled than long words (Neath & Nairne, 1995). According to this account, list composition does not matter; short items in mixed lists should be recalled just as well as short items in pure lists. Because a word length effect arises due to assembly errors, the Feature Model predicts a word length effect only when long and short words vary on the number of syllables or phonemes. There should be no difference in recall performance between two lists of words that differ only in pronunciation time, not on the number of syllables, since the word length effect is believed to be caused by reassembly errors.

The Feature Model also makes a prediction about the interaction between the word-length effect and concurrent articulation. Concurrent articulation is seen as adding noise to the vectors of each individual word. This process is called feature adoption. Feature adoption decreases the similarity between the word vector and the corresponding word in long-term memory, making recall harder. Even though short words have fewer segments than long words and should be easier to reassemble for recall, the word length effect would be abolished with concurrent articulation because the noise created by concurrent articulation removes the advantage that short words had. The word vectors for both short and long words would differ greatly from the corresponding words in long-term memory.

Of relevance to the current thesis, the Feature Model makes four predictions. First, for pure lists of all short or all long words, a word length effect will be observed when the to-be-recalled words differ in the number of syllables, with short words being better recalled than long words. Second, for mixed lists of alternating short and long words, short words will always be better recalled than long words. Third, concurrent articulation will abolish the short word advantage, making recall of short and long words equivalent. Fourth, since the Feature Model explains the word length effect as being due to reassembly errors based on how many syllables the to-be remembered items have, the same pattern of results stated in prediction 1, 2, and 3 should also be observed with nonwords.

It is, however, important to note that if the word length effect is found to be caused by something other than the number of syllables the words have, it is not critical to the Feature

Model. Since the word length effect is explained by the fact that there is a greater chance of committing a reassembly error at recall for long words than for short words, the Feature Model can easily remove the process that accounts for the word length effect without removing its ability to account for other core memory phenomena. In fact, a rudimentary redintegrative process was included in early versions of the Feature Model. If this redintegrative process is reinstated in the model, the Feature Model has the ability to explain how item characteristics can affect short-term recall performance.

1.4.2.2 Brown and Hulme Model

Brown and Hulme (1995) proposed a model in which rehearsal plays no role at all, but rather, differential decay of individual items is what leads to the word length effect. In contrast to the Feature Model where interference accounts for forgetting in short-term memory, the Brown and Hulme (1995) model hypothesizes that each segment of an item decays over time. In the Brown and Hulme model, forgetting is caused by decay, not interference. Since long words have more segments, the probability of correctly recalling every individual segment of a long word is smaller than for short words. Since the memory store is assumed to be blind to the lexical status of items, a word length effect should be observed with words, as well as with nonwords. Furthermore, because items decay at their given rate regardless of list composition, this account also predicts that recall of short items will be the same whether presented in a pure list or mixed with long items.

Brown and Hulme (1995) account for the interaction between the word length effect and concurrent articulation by assuming that concurrent articulation causes degradation of the memory traces during the gaps between presentation and recall. Since there are more gaps for short words because they take less time to encode, short words would suffer more from concurrent articulation. Again, because the memory store does not take into account lexical properties of items, this pattern of results will also be observed for nonwords.

Of relevance to the current thesis, Brown and Hulme's (1995) model makes the following four predictions. First, for pure lists of all short or all long words, a word length effect will be observed when the to-be recalled words differ in the number of syllables, with short words being better recalled than long words. Second, for mixed lists of alternating short and long words, since items decay at their given rate, short words will always be better recalled than long words. Third, concurrent articulation will cause more degradation of the memory traces for short words than long words, making recall of short and long words equivalent. Fourth, since the Brown and Hulme model does not take into account the lexical properties of the to-be remembered items, the same pattern of results stated in prediction 1, 2, and 3 should also be observed with nonwords.

1.4.2.3 SIMPLE

The Scale Invariant Memory, Perception, and Learning model (SIMPLE) is a local distinctiveness model in which memory performance is better for items that are more distinct, relative to other near items, at the time of retrieval (Brown, Neath, & Chater, 2007; Neath & Brown, 2006). If the items are similar on one or more relevant dimensions, such as serial position, phonological similarity, or spatial location, recall performance is worse than if the items were more easily discriminable. In other words, items with fewer close neighbours on relevant underlying dimensions in psychological space will be better remembered than items with more close neighbours.

The word length effect is explained by noting that short words are typically more distinctive (i.e., easier to apprehend) than long items because short words are less complex phonologically than long words (Neath & Brown, 2006).

In mixed lists, long words benefit from emergent distinctiveness; that is, compared to the short items, they now "stand out" more than when presented in a pure list of long words since a mixed list is more heterogeneous. Accordingly, long words should be about as well recalled as short words in mixed lists.

SIMPLE accounts for the interaction between the word length effect and concurrent articulation by assuming, like the Feature Model, that concurrent articulation adds noise to the memory traces. The addition of noise would make the short word traces less distinctive, abolishing the recall advantage for short words.

Of relevance to the current thesis, SIMPLE makes the following four predictions. First, a word-length effect will be observed for pure lists of all short or all long words, short being better recalled than long word lists. Second, for mixed lists of alternating short and long words, recall performance will be equivalent for short and long words, since long items now "stand-out" more in mixed lists. Third, concurrent articulation will abolish the word length effect making recall performance for short words equivalent to recall performance for long words. Concurrent articulation adds noise, making the short words memory traces less distinctive. Fourth, since short words are more distinctive than long words on a perceptual level and not on a lexical level, predictions 1, 2, and 3 will also be true for nonwords.

1.4.3 Empirical Evidence for the Syllable-Based Word Length Effect in Mixed Lists

Although the predictions are clear-cut, the empirical results are not. Cowan, Baddeley, Elliott, and Norris (2003) reported one experiment in which they included pure lists of six short words (1 syllable), or six long words (5 syllables), 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. Although performance in the mixed lists was in between that of the pure lists, as predicted by the phonological loop account, recall of short words from mixed lists was still better than recall of long words from mixed lists, a result predicted by the item-based accounts.

Hulme et al. (2004) reported a different pattern of results. They found, in two experiments, that recall of short items in mixed lists was equivalent to recall of long items in mixed lists, a result predicted by the list-based view, but recall of these items was equivalent to recall of short items in pure lists. The item-based view predicts that only short items from mixed lists would be recalled as well as short items from pure lists.

Bireta, Neath, and Surprenant (2006) argued that the difference in the pattern of results was attributable to particular properties of the stimulus sets used. Bireta et al. (2006) replicated the results reported by Cowan et al. (2003) when using Cowan et al.'s stimuli, and also replicated the results reported by Hulme et al. (2004) when using Hulme et al.'s stimuli. Bireta et al. noted that neither the item-based accounts nor the list-based account (i.e., the phonological loop) can predict either pattern in its entirety. As is the case with the time-based word length effect, then, aspects of the syllable-based word length effect appear to vary depending on the particular stimuli used.

1.5 The Phonological Loop Model Revisited

As more and more results were being published that contradicted the central claims of the phonological loop hypothesis, Mueller, Seymour, Kieras and Meyer (2003, p. 1353) published a paper in which they argued that these earlier results may have been due to "less than ideal measurements of articulatory duration and phonological similarity". To address the issue of articulatory duration, they introduced a different way of measuring the pronunciation time of the to-be-remembered items. To replace the various methods that have been used in the literature, Mueller et al. developed a procedure in which participants memorize a sequence of words and then produce the sequence from memory at least twice both "rapidly and accurately" (p. 1362). This procedure is then repeated with different orderings of the words, and the subsequent times analyzed.

To address the measurement of phonological similarity, Mueller et al. (2003) developed a new measure of phonological dissimilarity called PSIMETRICA (Phonological Similarity Metric Analysis). According to this measure, phonological dissimilarity between words is multidimensional and based on relevant dimensions like stress patterns and syllable onset. In order to compare words for dissimilarity using PSIMETRICA, each word is first decomposed into phonemes. Each syllable of a word is assumed to be composed of three different phoneme clusters: the onset (first consonants), the nucleus (vowel), and the coda (last consonants). The next step is to align the phoneme clusters in pairs of words. After the clusters have been aligned, phonological dissimilarity is measured to obtain a dissimilarity profile. Two identical clusters have a dissimilarity value of 0 and two very different clusters have a dissimilarity value closer to 1. The dissimilarity values for different phonemes can be calculated using a table of phonological features based on

Chomsky and Halle's (1968) system. For a list of words, the dissimilarity measure is comprised of the average of the dissimilarity value of all possible word pairs from the set.

Mueller et al. (2003) reported two experiments, one of which they stated demonstrated a time-based word length effect, and the other of which demonstrated a syllable-based word length effect. They argued that these results "confirm and extend the predictions of the phonological-loop model" (p. 1353).

However, the results are not as unambiguous as they initially appear, for three reasons. First, their method of measuring pronunciation time has been criticized. For example, Lewandowsky and Oberauer (2008, p. 879) noted that by using the time to reproduce the lists from memory as their measure of duration, Mueller et al. (2003) are "predicting accuracy in immediate serial recall from speed in immediate serial recall." This makes it difficult to claim it as a true prediction, as both measures -- accuracy and latency -- are typically highly correlated.

A second issue is that by one measure, Mueller et al. (2003) did not, in fact, demonstrate a time-based word length effect. The experiment involved three sets of words, simple short (Set 7), simple long (Set 8), and complex long (Set 9). For a pure time-based word length effect, there needs to be a difference between simple short and simple long words, as the complex long differ from the simple short in at least two ways (i.e., length and complexity). Although memory span for Set 7 was 5.21 compared to 5.05 for Set 8, this difference was not reported as statistically significant (see Mueller et al., 2003, p. 1371).

The third issue involves the evidence for a syllable-based word length effect. Like other researchers, Mueller et al. (2003) used a set of short and long words that confounded length with orthographic neighbourhood size, and thus it is not clear which difference is driving the effect. Of importance, the confound is the same one prevalent in the literature. I now turn to consideration of this issue.

1.6 Stimulus Set Specificity and Neighbourhood Effects

Despite the empirical and theoretical disagreements in the word length effect literature, one aspect has become increasingly apparent: The particular stimulus set used can critically determine whether effects of length will be seen (e.g., Bireta et al., 2006; Lovatt et al., 2000; Neath et al., 2003; see also Lewandowsky & Oberauer, 2008). Researchers do attempt to equate the short and long words on as many dimensions as possible, but it is difficult, if not impossible, to control every dimension of importance.

One factor rarely considered in such studies concerns the lexical neighbours of the to-be-remembered items. Words that are similar to a target word are referred to as its *neighbours* and the set of these words is referred to as the target word's *neighbourhood* (Coltheart, Davelaar, Jonasson, & Besner, 1977). Similarity can be defined on the basis of a word's orthography (Coltheart et al., 1977) or by its phonology (Luce & Pisoni, 1998). An orthographic neighbour is a word of the same length as the target that differs by only one letter. For example, given the word 'cat', the words 'bat', 'fat', 'cot', 'cut', 'cab', 'can', etc., are all considered orthographic neighbours. A phonological neighbour is one that differs from the target word by the substitution of a single phoneme at any position (Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002). There is a subtle difference between the Luce and Pisoni (1998) definition of a phonological neighbour and the Coltheart et al. (1977) definition of an orthographic neighbour. The former also includes all words that differ from the target word by the addition or deletion of a single phoneme in any position. Thus, the Luce and Pisioni definition includes *scat* and *at* as (phonological) neighbours of *cat* whereas the Coltheart et al. definition does not include either as (orthographic) neighbours of *cat*. The work reported here focuses on orthographic rather than phonological neighbourhood, as the use of orthographic neighbourhood eliminates the difficulty of differences in pronunciation and therefore phonological composition. Furthermore, the available data suggest both phonological and orthographic neighborhoods are highly correlated, and indeed, the measures are often confounded (Yates, Locker, & Simpson, 2004).

Two published papers have demonstrated better recall of words with a large neighbourhood than otherwise comparable words with a small neighbourhood. In their Experiment 1, Roodenrys et al. (2002) used CVC words, manipulating both neighbourhood size (small vs. large) and frequency of the target word. The task was memory span that used spoken recall for auditory presented items. Memory span was higher for words with larger neighbourhoods than those with smaller neighbourhoods. In Experiment 3, Roodenrys et al. used a second set of CVC words, this time manipulating neighbourhood size and the frequency of items that comprised the neighbourhood. Again, memory span was better for words with larger neighbourhoods. Finally, in Experiment 4, a third set of CVC words were used in which word frequency, neighbourhood size, and neighbourhood frequency were manipulated. The beneficial effect of neighbourhood size was replicated.

Allen and Hulme (2006, Experiment 2) used the stimuli from Experiment 1 of Roodenrys et al. (2002), but with a slightly different task. Their participants heard a list of seven words, and then immediately recalled the items outloud in the correct serial order. Despite the change in test, memory was again better for words with a larger neighbourhood than those with a smaller neighbourhood.

The beneficial effect of neighbourhood size is not limited to words; it is also observed with pronounceable nonwords (for a review, see Roodenrys, 2009). The neighbourhood of a nonword can be defined as all of the valid words that can be produced by the substitution of a letter (for orthographic neighbourhood) or phoneme (for phonological neighbourhood). For example, neighbours of the nonword *rin* include *bin*, *ran*, and *rip*. Roodenrys and Hinton (2002, Experiment 2) asked participants to listen to lists of four nonwords and then immediately repeat them back in order. Performance was better for nonwords with large neighbourhoods than those with small neighbourhoods. Thus, three sets of English words and one set of nonwords produce a recall advantage for items with a large neighbourhood over those with a small neighbourhood.

In contrast, Goh and Pisoni (2003) found better recall of words with few neighbours than words with many neighbours. However, there are a number of differences in stimuli and experimental design between their study and those of Roodenrys et al. (2002) and Allen and Hulme (2006). That makes it difficult to reconcile the results. First, Goh and Pisoni's (2003) small and large neighbourhood words were equated only for frequency and intra-set sharing neighbours, not for other variables known to affect immediate recall, like concreteness, familiarity, imageability, and PSYMETRICA dissimilarity.

Second, Roodenrys (2009) notes that even though Goh and Pisoni's (2003) small and large neighbourhood words did not significantly differ on neighbourhood overlap (how many neighbours the words of a list share), the actual probability was .14, with large-neighbourhood words having more overlap than small-neighbourhood words, making the conditions not as well matched as they could be. Furthermore, the distribution of neighbourhood overlap was not equivalent for small and large neighbourhood words. The large neighbourhood condition had a median of three overlapping neighbours with a range of zero to five while the small neighbourhood words had a median of two and a range of zero to seven overlapping neighbours. When Roodenrys (2009) removed the two words with six and seven overlapping neighbours from the small neighbourhood condition, an independent sample *t*-test now revealed that the small and large neighbourhood words did differ significantly on the number of overlapping neighbours, $p < .03$

Roodenrys (2009) argued that the effects of neighbourhood size on serial recall occur at retrieval by facilitating the reconstruction of a degraded trace. This process is called "redintegration". Roodenrys argued that the neighbourhood effect should be placed at output on the basis of results of phonological neighbourhood effects in language tasks. In particular, large phonological neighbourhoods (and high frequency neighbours) 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; Vitevich & Sommers, 2003). Consequently, in a short-term recall task where one has to produce the to-be recalled words, having more neighbours helps with the redintegration of to-be recalled words and improves the chances of correct recall. This concept of redintegration is not necessarily tied to any particular model; for example, it can be readily implemented in both interactive activation in long-term memory (McClelland & Rumelhart, 1981) and language-based models of short-term memory (Martin, Lesch, & Bartha, 1999).

Of relevance to the word length effect, short English words tend to have more neighbours -- both orthographic and phonological -- than do long words, and so neighbourhood size is likely to be confounded in word length effect experiments. To assess this, published studies on the syllable-based word length effect that used English words were examined. For those studies that reported the stimuli used, measures of orthographic neighbourhood size were obtained using the Medler and Binder (2005) database, which is based on the CELEX database. Table 1 lists the results. In all studies examined, short words had a larger orthographic neighbourhood than long words.

Table 1.1

Orthographic neighbourhood size for short and long words in syllable-based word length studies and the current study.

	Word Length	
Study	Short	Long
Baddeley et al. (1975, Experiment 6)	2.88	0.00
Baddeley et al. (2002, Experiment 1)	7.20	0.30
Coltheart et al. (2004, Experiment 1)	7.80	0.48
Cowan et al. (1994)	10.00	0.17
Cowan et al. (1997, Experiment 2)	14.17	0.17
Cowan et al. (2003)	6.33	0.33
Hulme & Tordoff (1989)	9.83	0.00
LaPointe & Engle (1990, Experiment 5)	8.37	0.31
McNeil & Johnston (2004, Experiment 1)	8.63	0.25
Mueller et al. (2003, Experiment 1)	8.42	0.17
Romani et al. (2005, Experiment 1)	7.25	0.38
Russo & Grammatopoulou (2003, Experiment 6)	8.40	0.00
Tehan & Turcotte (2002, Experiment 1)	12.60	0.60
Mean	8.61	0.24

One study in particular is highly suggestive: Coltheart, Mondy, Dux, and Stephenson (2004, Experiment 1) had three sets of stimuli: short one-syllable words (4 letters), long one-syllable words (6 or 7 letters), and three-syllable words (6 or 7 letters). The task was immediate serial recall of five-item lists presented at a rate of 1 item per second. The orthographic neighbourhood size for the three types of items was 7.80, 1.03, and 0.48 respectively. Recall level was affected by both word length (defined by the number of letters and the number of syllables) and also orthographic neighbourhood size: 0.76 for the shortest words, 0.62 for the intermediate length words, and 0.56 for the longest words.

1.7 Goal of the Current Thesis

Given the confound between word length and orthographic neighbourhood (see Table 1) and given that words with a large neighbourhood are better recalled than words with a small neighbourhood (Allen & Hulme, 2006; Roodenrys et al., 2002), the present thesis was designed to assess the extent to which neighbourhood size affects the word length effect. Visual presentation was used in all experiments. The first experiment was designed to show that a syllable-based word length effect (Experiments 1a and 1b) is observable with strict serial written recall and reconstruction of order. Previous studies on the word length effect have used strict written serial recall but a confound arises with written recall: output time. It takes longer to write down long words than it takes to write down short words. Consequently, since more time elapses between presentation and recall for long words, they could be harder to recall not because of their length, but because they had more time to decay or be interfered with in memory before recall.

In Experiments 2 and 3, different sets of short and long words were used, but this time the short and long words were equated for orthographic neighbourhood size. In

Experiment 4, results from Experiment 3 were replicated using spoken recall instead of a strict reconstruction of order test to see if the results could be replicated with a different recall method. Experiment 5 was designed to show that a typical neighbourhood size effect can be replicated with strict reconstruction of order.

Experiment 6 was designed to show that long items with a large neighbourhood size are better recalled than short items with a small neighbourhood size. Nonwords were used in Experiment 6 as it is easier to manipulate length and orthographic neighbourhood size with nonwords than with words.

Experiment 7 was designed to examine if the neighbourhood size effect, like the word length effect, would be eliminated by concurrent articulation. If neighbourhood size mediates the word length effect, the neighbourhood size effect should be abolished by concurrent articulation. Finally, Experiment 8 was intended as a replication of Experiment 7 using nonwords.[1]

1.8 Predictions

The current thesis tested the main hypothesis that the word length effect is caused by lexical variables underlying to-be-recalled words, not by the length of the words per se. More precisely, the possibility that neighbourhood size is a better explanation than is word length of the poorer recall of long words compared to short words in a short-term memory task than length was tested. Three predictions can be derived from this hypothesis. First, a word length effect will not be observed when short and long words are equated for

[1] Experiments 1a, 1b, 2, 3, 4, and 5 of the current thesis have been published in Jalbert, Neath, Bireta, & Surprenant (2011) while Experiments 6, 7, and 8 have been published in Jalbert, Neath, & Surprenant (in press).

neighbourhood size. Second, short words with a small neighbourhood size will be recalled worse than long words with a large neighbourhood size. Third, concurrent articulation will abolish the neighbourhood size effect.

Furthermore, the effect of recall task on recall performance was tested for short and long words as well as for small and large neighbourhood words. Written recall, reconstruction of order and spoken recall were compared. If the type of output task does not affect the pattern of recall for short, long, small neighbourhood and large neighbourhood words, reconstruction of order should be used because it removes the possible confound between output time and word length.

Chapter 2

Experiments

2.1 Experiment 1a

2.1.1 Rationale

The purpose of Experiment 1a was to demonstrate that typical word length effects are observable with written recall and visual presentation. Since the stimulus set used seems to have a great impact on results obtained for recall of short and long words, the goal here was to ensure that a word length effect could be observed with the method to be used in subsequent experiments. Before trying to abolish the word length effect by manipulating neighbourhood size, it is important to demonstrate that the effect can be obtained under the same conditions when neighbourhood size is confounded with word length. A new set of short (one syllable) and long (three syllable) items was created. The short and the long words were equated for frequency, concreteness, imageability, and familiarity, as well as for phonological dissimilarity as measured by PSIMETRICA. The words were not equated for orthographic neighbourhood size or frequency. Second, mixed lists were included in addition to pure lists to provide additional data on the effects of word length. Third, written serial recall was used.

2.1.2 Predictions

2.1.2.1 Phonological Loop

According to the phonological loop model, short words in pure lists should be better recalled than long words in pure lists. Long words take longer to rehearse than short words and are more prone to forgetting. For mixed lists of alternating short and long words, recall performance should be intermediate between recall of pure short and pure long lists. Mixed lists take longer than short pure lists to rehearse but less time than long pure lists.

2.1.2.2 Feature Model

According to the Feature Model, words are represented as a set of features. Since long words contain more segments than short words, there is a greater chance of making an error while reassembling the segments for recall. Therefore, short words will be better recalled than long words. Since the probability of correctly assembling segments is not related to list composition, short words will be better recalled than long words in both pure and mixed lists.

2.1.2.3 Brown and Hulme Model

Brown and Hulme (1995) hypothesized that words are divided into segments and that each segment decays over time. Since long words contain more segments than short words, the probability of correctly recalling a long word is less than the probability of correctly recalling a short word. Again, since the probability of correctly recalling all segments of words is unrelated to list composition, short words will be better recalled than long words in pure lists and in mixed lists.

2.1.2.4 SIMPLE

According to SIMPLE, short words are easier to apprehend than long words, making them more distinctive. Accordingly, short words will be better recalled than long words in pure lists. However, in mixed lists, short words lose their distinctiveness advantage. Long words in mixed lists now stand out more than short words when presented in pure lists. So, for mixed lists, recall performance should be similar for short words and for long words.

2.1.3 Method

2.1.3.1 Participants

Sixteen undergraduate students (9 women and 7 men, mean age = 21.69 yrs) from Memorial University of Newfoundland participated in exchange for a small honorarium. All participants were native English speakers.

2.1.3.2 Stimuli

A set of 15 short words and 15 long words was created (see Appendix A). The words were equated for familiarity, frequency (both Kucera-Francis and Thorndike-Lorge), concreteness and imageability using the Medical Research Council Psycholinguistics database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm). In addition, the set of short and long words were equated for phonological dissimilarity using Mueller et al.'s (2003) PSIMETRICA. The short words had a dissimilarity measure of 0.31 compared to

0.30 for the long words. However, the short and long words differed in orthographic neighbourhood size, with values typical of those in previous studies (9.00 vs. 0.22 respectively).

2.1.3.3 Design and Procedure

There were four types of lists: Pure lists that contained only short words, pure lists that contained only long words and two mixed lists with alternating short and long words, one mixed list starting with a short word and one mixed list starting with a long word. List type and word length were within-subjects variables. There were 15 trials for each type of list, randomly ordered for each participant.

On each trial, six words were randomly selected from the pool, and were presented at a rate of 1 item per second on a computer screen. At the end of list presentation, the participants wrote the words they had just seen in their original order. Strict serial recall instructions were given, such that participants were instructed to write the items in their exact order of presentation, beginning with the first one. They were told to leave a blank line if they could not recall an item at a given serial position, and were instructed not to backtrack to fill a blank. There was no time limit for recall. Once the participant had finished recalling the words, he or she clicked on a button on the computer to begin the next list. Participants were tested individually, and the experimenter was present throughout to ensure compliance with the instructions.

2.1.4 Results and Discussion

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

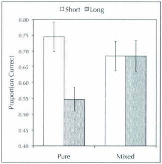

Figure 2.1: *Proportion of short and long words correctly recalled in Experiment 1a as a function of list type. Error bars show the standard error of the mean.*

As Figure 2.1 shows, a classic word length effect was observed in the pure lists, with substantially better recall of short than long words. However, recall of short and long words from mixed lists did not differ, with performance intermediate between that of short words in pure lists and long words in pure lists.

A 2 × 2 repeated measures ANOVA with word length (short and long) and list type (pure and mixed) as within-subject factors confirmed these trends. There was a main effect of word length, $F(1,15) = 45.05$, $MSE = 0.003$, partial $\eta^2 = 0.750$, with more short words correctly recalled in order than long words (0.715 vs. 0.616, respectively). There was also a main effect of list type, $F(1,15) = 6.12$, $MSE = 0.004$, partial $\eta^2 = 0.290$, with slightly more words correctly recalled in order in mixed lists than pure lists (0.685 vs. 0.646, respectively). These two factors also interacted, $F(1,15) = 46.14$, $MSE = 0.003$, partial $\eta^2 = 0.755$. This was due to a large difference between recall of short and long words in pure lists (0.745 vs. 0.547) and no difference short and long words in mixed lists (0.685 for both types of items). A Tukey HSD test confirmed that there was a reliable effect of word length in the pure lists but not in the mixed lists.

Another way of assessing the results is to see how many participants show a word length effect and how many do not. In pure lists, all 16 participants recalled more short than long words (significant by a sign test, $p < .0001$). For the mixed lists, 7 participants recalled more short than long words, with 8 showing the reverse and 1 tie, which is not significant by a sign test, $p > .90$.

The results of Experiment 1a showing a syllable-based word length effect confirm the predictions of the phonological loop hypothesis (see Burgess & Hitch, 1999). Pure lists of short words were recalled more accurately than lists of long words even though the words were equated for frequency, familiarity, concreteness, imageability, and phonological dissimilarity. In addition, recall of mixed lists was better than recall of pure long lists, but worse than recall of pure short lists. According to accounts based on the phonological loop hypothesis, it takes longer to refresh a list of long words than short words, and therefore, more long words will have decayed too far to be recallable at the time of test than short words. Similarly, it takes more time to rehearse a list consisting of both long and short words than it takes to rehearse lists of short words and consequently, pure lists of short words are recalled better than mixed-lists. Conversely, mixed-lists are rehearsed faster than pure lists of long words, making mixed-lists easier to recall than pure lists of long words.

The results of Experiment 1a also confirm the prediction of SIMPLE. Pure lists of short words were better recalled than pure lists of long words. Short words in pure lists are considered more distinctive than long words in pure lists, thus are easier to recall. Furthermore, according to SIMPLE, short words in mixed lists should lose their advantage when presented with long words, while long words would benefit from a mixed list presentation. Short and long words in mixed lists should be recalled equally well. Results of Experiment 1a showed exactly that pattern of results.

However, the results of Experiment 1a only partly confirm the prediction of the Brown and Hulme (1995) model and the Feature Model. Both models predict that short words should always be better recalled than long words, no matter how the list is composed. This pattern of results was observed only for pure lists. Short words were not better recalled than long words in mixed lists. That causes a problem for both the Feature Model and the Brown and Hulme model.

One possible problem with Experiment 1a is that written serial recall was used, which could cause a confound between word length and writing the words. Because it takes longer to write long words (*telegraph, sympathy, ...)* than short words (*sale, rose, ...*), output time is not equal in the two conditions. Experiment 1b removed this confound by using a strict serial reconstruction of order test rather than a strict written serial recall test. Strict serial reconstruction of order requires the participants to press on buttons labeled with the short and long words in the correct presentation order. Since it does not take more time to click on a button labeled with a long word that it takes to click on a button label with a short word, this recall method removes the confound of output time.

2.2 Experiment 1b

2.2.1 Rationale

Output time has been shown to be related to accuracy, with longer times associated with lower performance (e.g., Bireta et al., 2010; Dosher & Ma, 1998; Surprenant, Neath, & Brown, 2006). The purpose of Experiment 1b was to demonstrate that typical-looking word length effects are observable even when the confound of differential output time is removed. The same items as Experiment 1a were used, but a strict serial reconstruction of order test was used rather than written serial recall. This test yields results comparable to those observed with written serial recall, including not only word length effects (e.g., Neath et al., 2003), irrelevant speech and phonological similarity effects (e.g., Surprenant, Neath, & LeCompte, 1999), but also modality and suffix effects as well as effects of concurrent articulation (e.g., Surprenant, LeCompte, & Neath, 2000). More importantly, it permits output time to be equated. Unlike written or spoken recall, it takes the same amount of time to click on a button labeled with a long word as it does to click on a button labeled with a short word.

2.2.2 Predictions

Predictions of the Feature Model, the Brown and Hulme (1995) model and SIMPLE are the same as for Experiment 1a. The removal of the output time confound by using reconstruction of order instead of strict serial recall should not affect recall performance for short words or for long words because the word length effect is caused by intrinsic properties of the words. The phonological loop model may predict a slight decrease in the strength of the word length effect because the time confound at recall is removed. However, a word length effect should still be observed because of decay offset by rehearsal at encoding. The predictions of the other models remain unchanged.

2.2.3 Method

2.2.3.1 Participants

Sixteen undergraduate students (11 women and 5 men, mean age = 19.69 yrs) from Memorial University of Newfoundland participated in exchange for a small honorarium. All participants were native English speakers and none had participated in Experiment 1a.

2.2.3.2 Stimuli, Design and Procedure

The stimuli, design, and procedure were the same as in Experiment 1a except for the recall procedure. Following the presentation of the list, the six words from the current trial appeared in alphabetical order as labels on buttons on the computer screen and participants were asked to reconstruct the order in which the words were presented by clicking on the appropriately labeled buttons with the mouse. Participants were asked to click on the first word first, the second word second, and so on.

2.2.4 Results and Discussion

Despite the change in test, the results of Experiment 1b were almost identical to those of Experiment 1a. As Figure 2.2 shows, short words were better recalled than long words in the pure lists, but recall of short and long words in mixed lists was equivalent, and in between that of the short and long words from mixed lists. The results are exactly what the Burgess and Hitch (1999) model predicts.

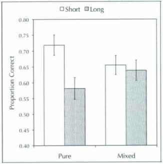

Figure 2.2: *Proportion of short and long words correctly recalled in Experiment 1b, as a function of list type. Error bars show the standard error of the mean.*

The data were analyzed with a 2 × 2 repeated measures ANOVA with word length (short vs. long) and list type (pure vs. mixed) as within-subject factors, which confirmed the observations noted above. There was a significant main effect of word length, with more short words correctly recalled than long words (0.687 vs. 0.610, respectively), $F(1, 15) = 44.871$, $MSE = 0.002$, partial $\eta^2 = 0.749$. There was no difference in recall of pure or mixed lists (0.650 vs. 0.647, respectively, $F < 1$).

Of importance, the interaction between word length and list type was significant, $F(1,15) = 19.110$, $MSE = 0.003$, partial $\eta^2 = 0.562$. This was due to finding a word length effect (i.e., better recall of short than long items) only in pure lists (0.719 vs. 0.582) and not in the mixed lists (0.656 vs. 0.639). A Tukey HSD test confirmed that there was a reliable effect of word length in the pure lists but not in the mixed lists.

In pure lists, 15 participants recalled more short than long words, 1 showed the reverse pattern, and there were no ties. The difference was significant by a sign test, $p <$

.05.) For the mixed lists, 10 participants recalled more short than long words, with 4 showing the reverse and 2 ties, which is not significant by a sign test, $p > 0.15$.

Experiment 1b demonstrated that a robust word length effect is observable with a strict reconstruction of order test. Short words were recalled better than long words in pure lists, but not in mixed lists; here, recall was in between that of pure short and pure long lists, and recall did not differ between mixed short and mixed long lists. This pattern is exactly what the Burgess and Hitch's (1999) model, which is based on the phonological loop, predicts. This pattern also differs subtly from previous patterns seen with pure vs. mixed lists. Unlike the results of Cowan et al. (2003), no word length effect was seen in mixed lists. Unlike the results using the stimuli of Hulme et al. (2004), recall of short and long items from mixed lists was worse than that of pure short lists.

There are several possible reasons for these differences. First, output time was equated for short and long words. Bireta, Fry, Jalbert, Neath, Surprenant, Tehan, and Tolan (2010) also measured output time, and also observed a word length effect with pure lists when output times did not differ. It is not known whether output times differed in the other studies, but this could easily be a factor. Second, it is possible that differences in the stimulus sets was the cause, particularly as the current set of stimuli were equated on more dimensions than either the Cowan et al. (2003) or Hulme et al. (2004) stimuli. Given that serial reconstruction of order removes the potential confound of output time and word length relative to written or spoken recall, Experiments 2, 3, 5, 6, 7 and 8 used a reconstruction of order task rather than strict serial recall.

2.3 Experiment 2

2.3.1 Rationale

Experiments 1a and 1b demonstrated that a word length effect is observed in pure but not mixed lists with both written recall and reconstruction of order tests. However, length and neighbourhood size were confounded in Experiments 1a and 1b, and it is not clear which factor is driving the effect. The purpose of Experiment 2 was to determine what happens when short and long items are equated for orthographic neighbourhood size and frequency, in addition to word frequency, concreteness, imageability, familiarity, phonological dissimilarity, and output time. If orthographic neighbourhood size plays no role in the word length effect and the effects observed in Experiment 1b are due to length *per se*, Experiment 2 should replicate Experiment 1b. If, on the other hand, the effects observed in Experiment 1b are due solely to neighbourhood characteristics, Experiment 2 should show no difference in recall of short and long words in either pure or mixed lists. Because a null result is being predicted, the number of participants in this experiment was doubled in size from Experiment 1.

2.3.2 Predictions

2.3.2.1 Phonological Loop

According to the phonological loop, the word length effect should still be observed when orthographic neighbourhood size of short and long words is controlled for. The word length effect arises because of decay offset by rehearsal, not because of intrinsic lexical properties of short and long words.

2.3.2.2 Feature Model and Brown and Hulme Model

According to the Feature Model and the Brown and Hulme (1995) model, controlling for neighbourhood size should not affect the word length effect. Long words are recalled worse than short words because the probability of correctly reassembling the segments for recall of long words is less than the probability of correctly reassembling a short word.

2.3.2.3 SIMPLE

According to the SIMPLE model, the word length effect arises from the enhanced predictability of short words caused by their phonological simplicity compared to long words (Neath & Brown, 2006). Consequently, short words are more distinctive than long words. Controlling for the number of neighbours should not affect the word length effect since it should not affect the predictability of short words overall.

2.3.3 Method

2.3.3.1 Participants

Thirty-two undergraduate students (24 women and 8 men, mean age = 18.84 yrs) from Memorial University of Newfoundland and The College of New Jersey participated in exchange for a small honorarium or course credit. All participants were native speakers of English, and none had been in previous experiments.

2.3.3.2 Stimuli

A set of 13 short and 13 long words was created (see Appendix B) in which the short and long words were equated on the same dimensions as in Experiments 1a and 1b, as well as being equated for orthographic neighbourhood size and frequency. The short words contained one syllable while the long words contained three syllables. For these two measures, the smallest p value associated with a t-test was $p = 0.48$. The measure of phonological dissimilarity was 0.33 for the short words compared to 0.28 for the long words.

2.3.3.3 Design and Procedure

With the exception of the stimuli used, the design and procedure were the same as in Experiment 1b.

2.3.4 Results and Discussion

The word length effect observed in Experiment 1b was not present in Experiment 2. As can be seen in Figure 2.3, recall of short words, whether in pure or mixed lists, did not differ from recall of long words, whether in pure or mixed lists. That is, there was no effect of word length when short and long words were equated for neighbourhood size.

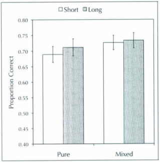

Figure 2.3: *Proportion of short and long words correctly recalled in Experiment 2 as a function of list type. Error bars show the standard error of the mean.*

Because no effect of length was observed, it is possible that participants had adopted a different strategy than in previous experiments. In particular, it is possible that participants were focusing on just the first letter of each word rather than on the whole word. If participants were memorizing only the first letter of each word, a list with words sharing the same first letter (i.e., *tree, table, soap, sack*) would be harder to recall than a list of words with a different first letter (i.e., *tree, chair, soap, bag*). This was not an issue for Experiment 1a since written serial recall was used as the recall methodology. Furthermore, Experiment 1b replicated almost perfectly the results from Experiment 1a, suggesting that participants did not adopt a different encoding strategy based on the first letter of each word. To assess the possibility that the first letter strategy was used in Experiment 2, shared first letter among the items was included as a covariate. The data were analyzed by a 2 × 2 repeated measures ANCOVA with word length (short vs. long) and list type (pure vs. mixed) as within-subject factors and shared first letter as a covariate. The covariate did not interact with word length or list type. There was no effect of word length, $F < 1$, with short and long words recalled equivalently (0.718 vs. 0.737, respectively). There was also no effect of list type, $F < 1$; the proportion of items recalled from mixed lists was 0.736 compared to 0.719 for pure lists. The interaction between length and list type was also not significant, $F(1,62) = 1.291$, $MSE = 0.010$, partial $\eta^2 = 0.020$, $p = .26$.

In pure lists, 15 of 32 participants recalled more short than long words, 17 showed the reverse pattern, and there were no ties. In the mixed list, the same pattern was observed. Neither are significant by a sign test, $p > 0.80$.

The only change between Experiment 1b and Experiment 2 was the set of words used, and the specific change was removing the confound of length and orthographic neighbourhood size. The short words in both experiments were all monosyllabic, and the long words were all trisyllabic. However, all the words in Experiment 2 had an orthographic neighbourhood of 1. Despite seeing robust effects of word length in Experiment 1b, no such effects were observed in Experiment 2.

The results from Experiment 2 are hard to explain from the perspective of models based on the phonological loop. Since long words take longer to rehearse than short words, no matter their neighbourhood size, they should be more prone to decay and be recalled worse than short words. However, the results of Experiment 2 clearly show that this is not the case: When short and long words are equated for neighbourhood size, the word length effect disappeared. These results critically compromise all models that have a decay offset by rehearsal component like the phonological loop.

The results of Experiment 2 also cause problems for the Feature Model, the Brown and Hulme (1995) model and SIMPLE. All three models predict that short words should be better recalled than long words because of their intrinsic item properties. Specifically, the Feature Model predicts that short words suffer less than long words from reassembly errors. Since neighbourhood size does not affect the number of segments the short and long words have, short words should still be better recalled than long words even when equated for neighbourhood size. The Brown and Hulme model also predicts that short words will be better recalled than long words since short words suffer less from decay than long words do. Since decay rate is not affected by neighbourhood size, short words should still be better recalled than long words even when equated for neighbourhood size. Finally, SIMPLE hypothesizes that short words are better recalled than long words because they are perceptually more distinctive. Therefore, SIMPLE cannot explain the result that short and long words are recalled equally well when neighbourhood size is controlled for.

It is not plausible to argue that Experiment 2 did not have a sufficiently powerful manipulation of length. First, the number of syllables in the short and long words was the same as in Experiment 1b. Second, although pronunciation time was not measured, an informal examination of pronunciation time showed that no matter what temporal measure was used (i.e., "normal" speaking, fast speaking, etc.), the long words were longer than the short. Third, a word length effect was observed in Experiment 1b with half the number of participants as in Experiment 2. Even so, null results may be obtained for a variety of reasons, and given the variability in results in word length effect experiments due to the particular stimulus set used, a replication was deemed necessary. To this end, Experiment 3 was designed as a replication of Experiment 2 but with a different set of stimuli.

2.4 Experiment 3

2.4.1 Rationale and Predictions

One possible concern with Experiment 2 is that the null results observed are due to some peculiarity of the particular stimulus set used. Experiment 3, therefore, was a replication of Experiment 3, but with a new set of short and long words that were also equated for orthographic neighbourhood size. Predictions of the models for Experiment 3 are the same as Experiment 2.

2.4.2 Method

2.4.2.1 Participants

Thirty-two undergraduate students (22 women and 10 men, mean age = 19.41 yrs) from Memorial University of Newfoundland and the College of New Jersey participated in exchange for a small honorarium or course credit. All participants were native speakers of English and none had participated in previous experiments.

2.4.2.2 Stimuli, Design, and Procedure

The only change from Experiment 2 was the set of stimuli. A new set of 14 short and 14 long words was created (see Appendix C) in which the short and long words were equated on the same dimensions as in Experiment 2. For concreteness, familiarity, frequency, imageability, PSYMETRICA dissimilarity measure, and neighbourhood size and frequency, the smallest p value associated with a t-test was $p = 0.56$. The measure of phonological dissimilarity was 0.28 for the short words compared to 0.28 for the long words. In addition, the short and long words were equated for orthographic neighbourhood size (this time 2.0 rather than 1.0) as well as orthographic frequency.

2.4.3 Results and Discussion

As can be seen in Figure 2.4, Experiment 3 replicated Experiment 2: With short and long words equated for orthographic neighbourhood size, there were no apparent effects of word length.

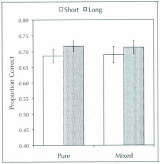

Figure 2.4: *Proportion of short and long words correctly recalled in Experiment 3 as a function of list type. Error bars show the standard error of the mean.*

A 2 × 2 repeated measures ANCOVA with word length (short vs. long) and list type (pure vs. mixed) as within-subject factors and shared first letter as a covariate found there was no effect of word length, $F < 1$, with recall approximately the same for short and long words (0.728 vs. 0.713, respectively). The main effect of list type failed to reach significance, $F < 1$, with mixed lists being recalled as well as pure lists (0.733 vs. 0.708, respectively). The interaction between length and list type was also not significant, $F < 1$. The covariate did not interact with word length or list type

In pure lists, 16 of 32 participants recalled more short than long words, 15 showed the reverse pattern, and there was 1 tie. This is not significant by a sign test, $p > 0.90$. In mixed lists, 9 participants recalled more short than long words, 19 showed the reverse, and there were 4 ties. Although the latter just fails to reach conventional levels of significance, $p > 0.08$, the direction of the difference is in favor of the long words, not the short words.

Experiment 3, with a different set of stimuli, replicated the null results from

Experiment 2: When short and long words are equated for orthographic neighbourhood size, there is no difference in recall of the short and long words.

The replication of the results from Experiment 2 with a new set of stimuli again poses a critical problem for models incorporating a phonological loop component. The replication shows that results from Experiment 2 was not caused by the stimulus set used but that word length seem to be caused by neighbourhood size, a factor not related to articulation time and decay offset by rehearsal. The replication also poses a problem for the Feature Model, the Brown and Hulme (1995) model, and SIMPLE. All three item-based models predicted a word-length effect when short and long words are equated for neighbourhood size.

2.5 Experiment 4

2.5.1 Rationale and Predictions

One possible concern is that the null results observed in Experiments 2 and 3 could be due to the recall method, even though there was no important difference between the results of Experiments 1a and 1b. With reconstruction of order, participants could possibly encode only the first letter of each word, even though this possibility has been statistically controlled for in Experiment 2 and 3. If participants were memorizing only the first letter of each word, a list with words sharing the same first letter would be harder to recall than a list of words with a different first letter. Therefore, Experiment 4 was a replication of Experiment 2, but with a spoken recall test. The use of spoken recall ensures the generalizability of the current results to another recall paradigm. Predictions of every model are the same as the predictions for Experiment 2.

2.5.2 Method

2.5.2.1 Participants

Sixteen undergraduate students (13 women and 3 men, mean age = 19.06 yrs) from Memorial University of Newfoundland volunteered to participate in exchange for a small honorarium. All participants were native speakers of English and none had participated in previous experiments.

2.5.2.2 Stimuli, design and procedure

The stimuli, design, and procedure were the same as in Experiment 2 except for the recall procedure. Following the presentation of the list, participants were ask to repeat out loud the words that were just presented. They were instructed to do so in the correct order of presentation. Participants' responses were taped using a digital recorder for later codification of the results. If participants were not sure what a word was, they were instructed to say *pass*.

2.5.3 Results and Discussion

Experiment 4 replicated Experiment 2. As can be seen in Figure 2.5, there is no word length effect apparent in either the pure or mixed lists conditions when the short and long words are equated for orthographic neighbourhood size.

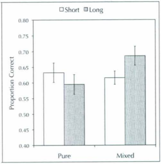

Figure 2.5: *Proportion of short and long words correctly recalled in Experiment 4 as a function of list type. Error bars show the standard error of the mean.*

Spoken recall performance was analyzed by a 2 × 2 repeated measures ANOVA with word length (short vs. long) and list type (pure vs. mixed) as within-subject factors. There was no difference in recall performance as a function of word length, $F(1, 15) = 1.666$, $MSE = 0.003$, partial $\eta^2 = 0.100$, $p = 0.22$, with a similar recall performance for short and long words (0.614 vs. 0.641). There was a significant main effect of list type, $F(1, 15) = 4.840$, $MSE = 0.005$, partial $\eta^2 = 0.244$, with better recall in mixed than pure lists (0.651 vs. 0.614). The interaction between word length and list type was significant, $F(1, 15) = 14.785$, $MSE = 0.003$, partial $\eta^2 = 0.496$. A Tukey HSD test confirmed that there was a reliable reversed word length effect in the mixed lists but no word length effect in the pure lists.

The results of Experiment 4 replicated results of Experiment 2 using a different recall methodology: When short and long words are equated for orthographic neighbourhood size, there is no difference in recall of the short and long words in pure lists.

In Experiment 4, it is unlikely that participants were adopting the strategy of memorizing the first letter of each word in the lists. Even when using spoken recall, the word length effect was abolished when the short and long words were equated for neighbourhood size. This is problematic for phonological loop models of the word length effect. The replication of Experiment 2 using spoken recall shows that results from Experiment 2 and 3 were not caused by the recall methodology or other mnemonic strategies encouraged by reconstruction of order but rather that the word length effect seems to be caused by neighbourhood size, a factor not related to articulation time and decay offset by rehearsal. The replication of Experiment 2 and 3 using spoken recall also poses a problem for the Feature Model, the Brown and Hulme (1995) model, and SIMPLE. All three item-based models predicted a word-length effect even when short and long words are equated for neighbourhood size.

2.6 Experiment 5

2.6.1 Rationale

In order to be able to attribute better recall of short words than long words to orthographic neighbourhood size, it is important to show a recall advantage of large neighbourhood words compared to small neighbourhood words. A recall advantage for words with a large orthographic neighbourhood has been demonstrated for three different sets of CVC words (Allen & Hulme, 2006; Roodenrys et al., 2002), as well as for nonwords (Roodenrys & Hinton, 2002). Each demonstration used auditory presentation and a memory span procedure or immediate serial recall and spoken recall. Thus, neighbourhood size effects have not been demonstrated in a reconstruction of order task nor have they been examined in mixed lists. The purpose of Experiment 5 was to determine whether the beneficial effect of a larger orthographic neighbourhood is observable with visual presentation and strict serial reconstruction of order.

2.6.2 Predictions

One general prediction that can be made for Experiment 5 is that if neighbourhood size is indeed driving the effect of word length, the same pattern of results as in Experiment 1b should be observed, even though word length is held constant. Experiment 1b used short and long words, which were not equated for neighbourhood size. In Experiment 1b, short words had a larger neighbourhood size than long words (9.00 vs. 0.22, respectively). Therefore, for Experiment 5, for pure lists of all large or small neighbourhood words, a neighbourhood size effect should be observed, with large neighbourhood words being better recalled than small neighbourhood words. For mixed lists of alternating large and small neighbourhood words, recall performance should be equivalent for large and small neighbourhood words and be intermediate between recall of large and small neighbourhood words from pure lists.

2.6.2.1 Phonological Loop

According to the phonological loop, forgetting occurs in working memory when the time it takes to rehearse words is longer than the time words take to decay. Since small and large neighbourhood words used in Experiment 5 have the same number of syllables, their decay rate should be approximately the same and small neighbourhood words should be as well recalled as large neighbourhood words.

2.6.2.2 Feature Model

The Feature Model does not make clear predictions about the effect of neighbourhood size on recall performance. However, a redintegration process was included in the early version of the model so it may be possible to add the beneficial effect of having a large number of neighbours for redintegration.

2.6.2.3 Brown and Hulme Model

The Brown and Hulme (1995) model also does not make clear predictions about the effect of neighbourhood size on recall performance. The model's purpose was to demonstrate that rehearsal was not necessary to explain immediate memory effects. If length is not the driving force in the word length effect, the Brown and Hulme model's assumption of differential decay rate for short and long words is questioned. However, it does not affect the model's ability to explain other memory phenomena.

2.6.2.4 SIMPLE

According to SIMPLE, words with fewer neighbours on the relevant underlying dimension are considered more distinctive and are consequently better recalled than words with more neighbours. SIMPLE would then predict better recall of small neighbourhood words compared to large neighbourhood words.

2.6.3 Method

2.6.3.1 Participants

Sixteen undergraduate students (12 women and 4 men, mean age = 22.81 yrs) from Memorial University of Newfoundland volunteered to participate in exchange for a small honorarium. All participants were native speakers of English and none had participated in previous experiments.

2.6.3.2 Stimuli

The stimuli were the 32 low neighbourhood frequency 3-phoneme CVC words from Experiment 3 of Roodenrys et al. (2002). Although initially selected for a manipulation of phonological neighbourhood size -- half of words had large phonological neighbourhoods and half had small phonological neighbourhoods -- the words also differ in terms of orthographic neighbourhood size. Orthographic neighbourhood size and frequency were calculated using the MCWord Database (Medler & Binder, 2005), and this value was 3.8 for the small neighbourhood words and 12.6 for the large neighbourhood words. The small and large neighbourhood words did not differ in terms of the PSIMETRICA measure of phonological dissimilarity: this value was 0.30 for the small neighbourhood set compared to 0.31 for the large neighbourhood set.

2.6.3.3 Design and Procedure

Except for the substitution of neighbourhood size for word length, the design and procedure were identical to that in Experiment 1b. That is, neighbourhood size (small vs. large) and list type (pure vs. mixed) were both within-subjects variables, and all lists contained six words. Pure small lists contained only words with small neighbourhoods and pure large lists contained only words with large neighbourhoods. Mixed lists alternated words with different neighbourhood sizes. Half of the mixed lists began with a small neighbourhood word (i.e., small, large, small, large, small, large) and the other half began with a large neighbourhood word (i.e., large, small, large, small, large, small). To construct each list, 6 words were drawn randomly from the appropriate pool. There were 15 trials for each of the 4 types of list, and these were randomly ordered for each participant.

2.6.4 Results and Discussion

As is shown in Figure 2.6, words with a large neighbourhood were recalled better than words with a small neighbourhood in pure lists, replicating the basic effect observed by Roodenrys et al. (2002) and Allen and Hulme (2006). Recall of large and small neighbourhood words did not differ in the mixed lists. This pattern is reminiscent of that observed in Experiments 1a and 1b, in which word length was manipulated except that here all the words were all 1-syllable words.

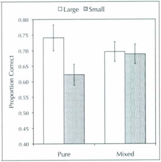

Figure 2.6: *Proportion of small and large neighbourhood words correctly recalled in Experiment 5 as a function of list type. Error bars show the standard error of the mean.*

The data were analyzed with a 2 × 2 repeated measures ANOVA with neighbourhood size (large vs. small) and list type (pure vs. mixed) as within-subject factors. The main effect of neighbourhood size was significant, $F(1, 15) = 17.566$, $MSE = 0.004$, partial $\eta^2 = 0.539$, with better recall of words with large neighbourhoods than those with smaller neighbourhoods (0.719 vs. 0.656, respectively). The main effect of list type was not significant, $F < 1$, with approximately equivalent recall of pure and mixed lists

(0.682 vs. 0.693, respectively).

The interaction was significant, $F(1, 15) = 13.801$, $MSE = 0.004$, partial $\eta^2 = 0.479$, due to an effect of neighbourhood size in pure lists (0.742 vs. 0.626) but no such effect in mixed lists (0.697 vs. 0.689). A Tukey HSD test confirmed that there was a reliable effect of neighbourhood size in the pure list but not in the mixed list.

Again, it was determined how many participants showed the orthographic neighbourhood effect and how many did not. For pure lists, 13 participants recalled more words from large than small neighbourhoods, 2 showed the reverse pattern, and 1 showed no difference. This is significant by a sign test, $p < .05$. For the mixed lists, 6 participants recalled more large than small neighbourhood words, with 10 showing the reverse and no ties; this is not significant by a sign test, $p > .40$.

With pure lists, Experiment 5 replicated the neighbourhood size effect reported by Roodenrys et al. (2002) and did so despite the many changes in design and procedure. Words with a large phonological or orthographic neighbourhood are better recalled on immediate serial recall tests than words with smaller neighbourhoods. It does not matter if presentation is auditory or visual, or if the test is memory span with written recall, immediate spoken serial recall, or strict reconstruction of order.

In mixed lists, however, there was no effect of neighbourhood size. Performance in these lists was in between that of the pure large and pure small conditions. That pattern is reminiscent of that predicted by the phonological loop models for word length effects with pure and mixed lists (e.g., Burgess & Hitch, 1999). These results provide some evidence that the confound between word length and neighbourhood size shown in Table 1 could be important. It is an indication that maybe neighbourhood size and not word length is the driving force in the word length effect.

The results of Experiment 5 are critical to the phonological loop model. Large and small neighbourhood words should have been equally well recalled if forgetting occurs when the time it takes to rehearse words is longer than the time words take to decay. Small and large neighbourhood words all have one syllable so their decay rate should be approximately the same.

The conclusion that length is not driving the word length effect is not critical to the Feature Model. If length is no longer a factor that needs to be explained by the model, removing the processes specific to word length does not reduce the model's ability to explain other memory phenomena. A redintegration process was included in the early version of the model, so it may be possible to add the beneficial effect of having a large number of neighbours for redintegration.

The results from Experiment 5 cause a problem for the Brown and Hulme model. If length is not the driving force in the word length effect, the Brown and Hulme model's assumption of differential decay rate for short and long words is challenged. However, it does not affect the model's ability to explain other memory phenomena.

The observation that large neighbourhood words are recalled better than small neighbourhood words poses a challenge for SIMPLE. According to this model, words with fewer neighbours on relevant dimensions are considered more distinctive and are recalled better. Results from Experiment 5 showed the opposite: Words with a larger neighbourhood were recalled better than words with a smaller neighbourhood.

2.7 Experiment 6

2.7.1 Rationale

Concurrent articulation is known to abolish or greatly attenuate the word length effect (Baddeley et al., 1975; Baddeley, Lewis, & Vallar, 1984; Bhatarah, Ward, Smith, & Hayes, 2009; Longoni et al., 1993; Romani et al., 2005; Russo & Grammatopoulou, 2003). If the word length effect is really due to differences in neighbourhood size between short and long words, then concurrent articulation should also remove the neighbourhood size effect. In Experiment 6, participants saw a list of one-syllable words, half with large neighbourhoods and half with small neighbourhoods. Half of the participants engaged in concurrent articulation during list presentation and half did not.

2.7.2 Predictions

2.7.2.1 Phonological Loop

Phonological loop models conceptualize concurrent articulation as preventing rehearsal in the phonological loop. If items cannot be rehearsed, they decay in the phonological store and cannot be properly recalled. Since the words used in Experiment 6 are all the same length, concurrent articulation should affect all the words the same way. Furthermore, there should not be a difference between recall of small neighbourhood words and large neighbourhood words.

2.7.2.2 Feature Model, Brown and Hulme Model, and SIMPLE

All three item-based models view concurrent articulation as adding noise in memory. The addition of noise makes everything harder to recall. Consequently, recall performance should be worse for small and large neighbourhood words with concurrent articulation than in the silent condition.

2.7.3 Method

2.7.3.1 Participants

Thirty-two undergraduate students (21 women and 11 men, mean age = 22.47 yrs) from Memorial University of Newfoundland volunteered to participate in exchange for a small honorarium. All participants were native speakers of English and none had participated in previous experiments.

2.7.3.2 Stimuli, Design, and Procedure

The stimuli were the same as those from Experiment 5. Concurrent articulation was manipulated between-subjects and neighbourhood size and list type were manipulated within-subjects. The procedure was similar to Experiment 5, except that half of the participants were instructed to perform a concurrent articulation task during the presentation of the items. They had to repeat the letters "A, B, C, D, E, F, G" as fast as they could during the presentation of the list of to-be recalled words.

2.7.4 Results and Discussion

As can be seen in Figure 2.7, large neighbourhood words in pure lists were recalled better in the silent condition than small neighbourhood words in pure lists, replicating the basic neighbourhood size effect. Concurrent articulation eliminated this effect. In the mixed lists, no neighbourhood size effect was observed.

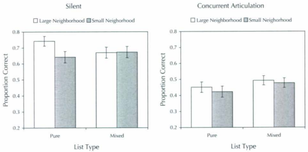

Figure 2.7: *Proportion of words with large or small neighbourhoods recalled from pure or mixed lists in the silent condition (left panel) and the concurrent articulation condition (right panel). Error bars show the standard error of the mean.*

These trends were analyzed with a 2 × 2 × 2 mixed design ANOVA with neighbourhood 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. There was a significant main effect of neighbourhood size, $F(1, 30) = 12.665$, $MSE = 0.003$, partial $\eta^2 = 0.297$, with words from large neighbourhoods being better recalled than words from small neighbourhoods (0.590 vs. 0.554). The main effect of list type was not significant, $F(1, 30) = 1.083$, $MSE = 0.006$, partial $\eta^2 = 0.035$, $p = .31$, with words from pure lists being recalled as well as words from mixed lists (0.565 vs. 0.579). The main effect of encoding condition was significant, $F(1, 30) = 26.378$, $MSE = 0.059$, partial $\eta^2 = 0.468$, with recall performance being better in the silent condition than in the concurrent articulation condition (0.682 vs. 0.461).

The interaction between neighbourhood size and list type was significant, $F(1, 30) =$

24.014, $MSE = 0.001$, partial $\eta^2 = 0.445$, reflecting, in part, a difference in neighbourhood size in pure, but not mixed lists. The interaction between list type and encoding condition was also significant, $F(1, 30) = 6.636$, $MSE = 0.006$, partial $\eta^2 = 0.181$, 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 neighbourhood size and encoding condition failed to reach conventional levels of significance, $F(1, 30) = 1.793$, $MSE = 0.003$, partial $\eta^2 = 0.056$, $p = .19$

When interpreting the significant two-way interactions, it is important to keep in mind that the three-way interaction between neighbourhood size, list type, and encoding condition was significant, $F(1,30) = 14.379$, $MSE = 0.001$, partial $\eta^2 = 0.324$. This reflects the presence of a neighbourhood 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 neighbourhood words in pure lists in the silent condition (0.742 vs. 0.642), but no differences in any other condition (for mixed lists in the silent condition, 0.669 vs. 0.673; for pure lists in the concurrent articulation condition, 0.451 vs. 0.423; and for mixed lists in the concurrent articulation condition, 0.493 vs. 0.478, respectively).

If neighbourhood size is an important factor in driving previous word length effects, then one should expect similar interactions between neighbourhood size and factors known to interact with word length. In Experiment 6, a neighbourhood size effect observed in pure lists was abolished by concurrent articulation, the same result seen with word length effects (e.g., Baddeley et al., 1975). This confirms the prediction that neighbourhood size interacts with concurrent articulation in the same way that word length does. In addition, Experiment 6 replicated the finding that neighbourhood 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., 2003).

Results from Experiment 6 are consistent with the claim that neighbourhood size may have been the cause of previous demonstrations of the word length effect, since in those studies length and neighbourhood size were confounded. If the claim is accurate, then 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 neighbourhood size. Concurrent articulation, which abolishes the word length effect, also abolishes 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), or the word class and imageability effects (Bourassa & Besner, 1994).

It is difficult to explain the 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 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). If a large neighbourhood helps recall by assisting with the redintegrative process (Roodenrys, 2009), then the result makes sense. For example, if one were to assume that the degraded cue serves as input to an interactive network, then the slight activation in the network accruing from the commonalities of the neighbours -- 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-neighbourhood items need identifying, which slightly helps the small neighbourhood items while slightly hurting the large neighbourhood items. The small-neighbourhood items are helped by the removal (relative to the pure lists) of three additional harder to redintegrate items whereas the large-neighbourhood items are hindered by the addition of three harder to redintegrate items. If concurrent articulation adds noise, then the benefit conveyed by having a larger number of neighbours will be removed, thus lowering performance substantially for large neighbourhood items. However, small neighbourhood items never had much of a benefit from neighbours to begin with, so interfering with this process has little effect.

Regardless of the explanation, the results from Experiment 6 support the view that length may not be the cause of the word length effect. The next question is whether reversing the usual confounding of length and neighbourhood size, such that long words have a large neighbourhood and short words have a small neighbourhood, will a neighbourhood size effect still be observable? Unfortunately, one cannot use real words to test this hypothesis, as there are not enough long words with large neighbourhoods in the English language. Thus, nonwords are needed. However, it is necessary to demonstrate that the neighbourhood size effect observed with nonwords is eliminated by concurrent articulation, just like the neighbourhood size effect with words. Having done that in

Experiment 7, Experiment 8 will then use nonwords to examine whether length or neighbourhood size has the greater effect on recall.

2.8 Experiment 7

2.8.1 Rationale and Predictions

The goal of Experiment 7 was to replicate the results from Experiment 6 with nonwords. Roodenrys and Hinton (2002) have already demonstrated a neighbourhood size effect with nonwords, but it is important to verify that just as in Experiment 6, this effect is eliminated by concurrent articulation. Therefore, Experiment 7 was just like Experiment 6 except that the stimuli were a set of one-syllable nonwords, half with large neighbourhoods and half with small neighbourhoods. Neighbours of nonwords are words that differ from the nonwords by only one letter. Since nonwords interact the same way as words with neighbourhood size (see, Roodenrys, 2009, for a review), predictions of the phonological loop models, the Feature Model, the Brown and Hulme model and SIMPLE are the same as for Experiment 6.

2.8.2 Method

2.8.2.1 Participants

Thirty-two undergraduate students (18 female and 14 male, mean age = 19.94 yrs) from Memorial University of Newfoundland volunteered to participate in exchange for a small honorarium. All participants were native speakers of English and none had participated in the previous experiments.

2.8.2.2 Stimuli

A set of 24 nonwords was created using the orthographic word form database of

Medler and Binder (2005) (see Appendix D). All of the nonwords were one-syllable and all contained five letters. Half of the nonwords had a large neighbourhood size and half had a small neighbourhood size (26.25 vs. 6.58).

2.8.2.3 Design and Procedure

The design and procedure was the same as Experiment 6 except for the use of nonwords instead of words.

2.8.3 Results and Discussion

As can readily be seen, Figure 2.8 looks just like Figure 2.7 despite the change from words to nonwords. Large neighbourhood nonwords in pure lists were recalled better in the silent condition than small neighbourhood nonwords in pure lists, replicating the basic neighbourhood size effect. Concurrent articulation eliminated this effect, just as it did for words. In the mixed lists, no neighbourhood size effect was observed.

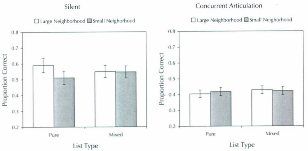

Figure 2.8: *Proportion of nonwords with large or small neighbourhoods recalled from pure or mixed lists in the silent condition (left panel) and the concurrent articulation condition (right panel). Error bars show the standard error of the mean.*

These trends were analyzed with a 2 × 2 × 2 mixed design ANOVA with neighbourhood 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 6, the main effect of neighbourhood size did not reach the adopted significance level, $F(1, 30) = 3.410$, $MSE = 0.003$, partial $\eta^2 = 0.102$, $p = 0.075$. The proportion of nonwords with large neighbourhoods correctly recalled was 0.493 compared to 0.474 for those with small neighbourhoods. The main effect of list type was not significant, $F < 1$, with approximately equivalent recall in pure and mixed lists (0.481 vs. 0.486, respectively). There was a significant main effect of encoding condition, $F(1, 30) = 8.786$, $MSE = 0.063$, partial $\eta^2 = 0.227$, with better recall performance in the silent condition than in the concurrent articulation condition (0.549 vs. 0.418, respectively).

Neither the interaction between neighbourhood size and list type, $F(1, 30) = 2.245$, $MSE = 0.018$, partial $\eta^2 = 0.070$, nor the interaction between list type and encoding condition, $F < 1$, were significant. However, the interaction between neighbourhood size and encoding condition did reach conventional levels of significance, $F(1, 30) = 4.973$, $MSE = 0.003$, partial $\eta^2 = 0.142$. This reflects a difference in recall of nonwords from large and small neighbourhoods in the silent condition (0.569 vs. 0.529), but no difference in the concurrent articulation condition (0.416 vs. 0.420).

It is important to keep in mind when interpreting the two-way interactions that the three-way interaction between neighbourhood size, list type, and encoding condition was significant, $F(1,30) = 6.175$, $MSE = 0.003$, partial $\eta^2 = 0.171$. This reflects the presence of a neighbourhood 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 neighbourhood words in pure lists in the silent condition (0.590 vs. 0.511), but no differences in any other condition (for mixed lists in the silent condition, 0.549 vs. 0.547; for pure lists in the concurrent articulation condition, 0.404 vs. 0.417; and for mixed lists in the concurrent articulation condition, 0.428 vs. 0.422, respectively).

There were some slight differences in the particular pattern of significant interactions between Experiments 6 and 7, but nonwords do sometimes result in a slightly different pattern than words (e.g., Romani et al., 2005). The major results of both experiments, however, are the same: (1) A neighbourhood 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.

2.9 Experiment 8

2.9.1 Rationale

Because long English words typically have far smaller neighbourhood sizes than short words, it is difficult to find long words with a large neighbourhood size. Turning to nonwords seemed practical. Short pronounceable nonwords are better recalled than long pronounceable nonwords (e.g., Romani et al., 2005). Furthermore, Roodenrys and Hinton (2002) showed that nonwords with larger neighbourhood sizes are recalled better than otherwise equivalent nonwords with smaller neighbourhood sizes. Similarly, Experiment 7 demonstrated that nonwords from large neighbourhoods are recalled better than words from small neighbourhood in the silent condition and that neighbourhood size interacts the same way as word length in the presence of concurrent articulation in that the neighbourhood size effect disappears in the presence of concurrent articulation. By using nonwords, length and neighbourhood size can be factorially manipulated. That is, one can compare short nonwords with a large neighbourhood, short nonwords with a small neighbourhood, long nonwords with a large neighbourhood, and long nonwords with a small neighbourhood. While an ideal experiment would use words, there are not enough suitable long words in the English language that have a large neighbourhood. As nonwords also show effects of length, Experiment 8 used nonwords. If neighbourhood size is driving the word length effect, there should be better recall of nonwords with large neighbourhoods than those with small neighbourhoods regardless of the length. If length is driving the word length effect, there should be better recall of short nonwords than long nonwords, regardless of neighbourhood size.

2.9.2 Predictions

2.9.2.1 Phonological Loop

Phonological loop models predict that a word length effect will be observed with short and long nonwords because the time it takes to pronounce a short nonword is less than the time it takes to pronounce a long nonword. Long nonwords would be more prone to forgetting because they decay before their memory traces have a chance to be refreshed. Neighbourhood size should not affect recall performance since it is not related to rehearsal rate. Consequently, recall performance for small neighbourhood nonwords should be the same as recall performance for large neighbourhood nonwords. The only significant factor in the factorial design of Experiment 8 should be the length of the nonwords.

2.9.2.2 Feature Model

According to the Feature Model, short words should always be better recalled than long words. Since the model explains the word length effect by the number of segments long and short words have and by the chance of committing errors while reassembling the segments, nonwords should produce the same pattern of results. Again, the Feature Model, as currently conceptualized, does not make a prediction about effects of neighbourhood size. However, if length is not the driving factor behind the word length effect, it is not critical to the Feature Model. The processes that are responsible for the word length effect can be removed from the model without removing its ability to explain other core memory phenomena. A redintegration process was included in the early version of the model so it may be possible to add the beneficial effect of having a large number of neighbours for redintegration.

2.9.2.3 Brown and Hulme Model

Similarly to the Feature Model, the Brown and Hulme (1995) model predicts a word length effect by assuming that long words are more prone to assembly errors in short-term memory. Since the phonological store is blind to lexical properties of items, the word length effect will also be observed with nonwords. Because of its limited scope (the Brown and Hulme model was intended as a demonstration that rehearsal is not necessary to explain short-term memory effects), the model as currently conceptualized does not make a prediction about the effect of neighbourhood size.

2.9.2.4 SIMPLE

SIMPLE makes the assumption that the word-length effect is caused by short words being more distinctive or easier to apprehend than long words. Since short words are more distinctive than long words, this prediction can also be applied to nonwords. SIMPLE also predicts that items with fewer neighbours on relevant dimensions are considered more distinctive and are recalled better than words with more neighbours. Thus, SIMPLE would predict better recall of nonwords from a small neighbourhood than nonwords with a large neighbourhood. It is not yet known which variable (length or neighbouhood size) would have a stronger influence on recall performance.

2.9.3 Method

2.9.3.1 Participants

Sixteen undergraduate students (12 women and 4 men, mean age = 23.63 yrs) from Memorial University of Newfoundland participated in exchange for a small honorarium. All participants were native speakers of English and none had participated in previous experiments.

2.9.3.2 Stimuli

A set of 48 nonwords was created using the orthographic word form database of Medler and Binder (2005) (see Appendix E). Half were short (monosyllabic) and half were long (disyllabic). In addition, half had a small neighbourhood size (0 neighbours) and half had a large neighbourhood size (5.92 for short and 5.83 for long items). Phonological dissimilarity was also equated: for the small neighbourhood, the PSIMETRICA measure was 0.30 for the short nonwords compared to 0.34 for the long nonwords, and for the large neighbourhood, the measure was 0.33 for the short nonwords and 0.33 for the long nonwords.

2.9.3.3 Design and Procedure

Length (short vs. long), and orthographic neighbourhood size (small vs. large) were within-subjects variables. The procedure was similar to Experiment 2, except that each type of list was tested 15 times. The order of the lists was randomized for each participant.

2.9.4 Results and Discussion

As a manipulation check, recall of short nonwords with a large neighbourhood and recall of long nonwords with a small neighbourhood were first compared. These correspond to the stimuli used in a typical word length study. The short items should be better recalled than the long items, and indeed, they were: 0.543 vs. 0.490, significant by a Tukey HSD test.

As can be seen in Figure 2.9, recall did not differ as a function of length but did differ as a function of neighbourhood size.

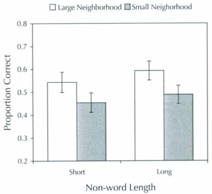

Figure 2.9: *Proportion of short and long nonwords with large or small neighbourhoods correctly recalled in Experiment 6. Error bars show the standard error of the mean.*

These trends were confirmed with a 2 × 2 repeated measures ANOVA with word length (short vs. long) and neighbourhood size (small vs. large) as within-subject factors. There was a main effect of neighbourhood size, $F(1,15) = 25.371$, $MSE = 0.006$, partial $\eta^2 = 0.628$. Nonwords with a large neighbourhood were better recalled than nonwords with a small neighbourhood, 0.568 vs. 0.472, respectively.

The main effect of length was not significant, $F(1,15) = 3.209$, MSE = 0.009, partial $\eta^2 = 0.389$, $p > .09$. Although the difference was not significant, the trend was for slightly better recall of the longer nonwords, 0.541 vs. 0.499. The interaction was not significant, F < 1.

As with words, short nonwords that follow the general rules of English have more neighbours than otherwise comparable long nonwords. However, there are a sufficient number of nonwords that it was possible to manipulate length and neighbourhood size factorially. When this was done, two results stood out: (1) only neighbourhood size had a measurable effect on the proportion of items correctly recalled, and (2) short-large neighbourhood items are recalled better than long-small neighbourhood items. This latter finding corresponds to the typical manipulation of word length in the literature, in which length and neighbourhood size are confounded.

Chapter 3

General Discussion

3.1 Review

3.1.1 Goals and Predictions

The goal of the current series of experiments was to test three predictions that arise from the claim that neighbourhood 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 neighbourhoods to long items from small neighbourhoods, then (1) a word length effect will not be observed when short and long words are equated for neighbourhood size, (2) long words with a larger neighbourhood should be better recalled than short words with a smaller neighbourhood, and (3) concurrent articulation should remove the neighbourhood size effect, just as it removes the word length effect. Furthermore, it was predicted that the type of recall task (reconstruction of order, written serial recall and spoken recall) would not interfere with the pattern of results.

3.1.2 Summary of the Main Findings

In Experiment 1, the set of stimuli used deliberately confounded length and neighbourhood size, such that the one-syllable short items had a larger neighbourhood size than the three-syllable long items. Experiment 1 demonstrated that the reconstruction of order task produces the same standard word length effect seen with strict written serial recall. In Experiment 2, a set of short (one syllable) and long (three syllables) words equated for orthographic neighbourhood size and frequency were used, and the word length effect observed in Experiment 1 disappeared. In Experiment 3, the null results of Experiment 2 were replicated when using a different set of short and long words equated for orthographic neighbourhood size and frequency. Experiment 4 extended the results of Experiments 2 and 3 by showing no word length effect with a spoken recall test when the items were equated for neighbourhood size. Experiment 5 replicated and extended the results of Roodenrys et al. (2002) and Allen and Hulme (2006) by showing that visually presented words with large orthographic neighbourhoods were better recalled than words with smaller orthographic neighbourhoods using a reconstruction of order task. Experiment 6 showed that the neighbourhood size effect observed in the silent condition was abolished by concurrent articulation. Experiment 7 replicated the results of Experiment 6 with nonwords, showing that the neighbourhood size effect was abolished by concurrent articulation. Finally, Experiment 8 used a complete factorial design to assess length and neighbourhood size, and found a main effect of neighbourhood size but no effect of length.

3.2 Neighbourhood Size and the Word Length Effect

Given these results, 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 neighbourhood size. Neighbourhood size is a better predictor of performance than 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.

The short and long words in Experiments 2, 3, and 4 were equated on all relevant dimensions thought to be important, but no effects of length were observed. Because it is possible that some other important dimension was overlooked, and because of the past history of differing word length results as a function of specific stimulus sets (e.g., Neath et al., 2003; Bireta et al., 2006), it is critically important that other researchers replicate these results using different stimulus sets. For any such sets, in addition to controlling for output time, researchers should also ensure that their short and long words are equated on at least the following dimensions: Concreteness, familiarity, imageability, frequency (Kucera-Francis, Thorndike-Lorge, and CELEX), orthographic frequency, orthographic neighbourhood size, and PSIMETRICA dissimilarity.

However, until more stimulus sets are tested, the following conclusion is warranted: neighbourhood size -- and possibly other related lexical and linguistic factors correlated with it -- rather than length *per se* is one of the critical factors underlying the syllable-based word length effect.

One possible concern is that the word length effect was attenuated by the recall methodology. More specifically, a proponent of the phonological loop hypothesis 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. This possibility was tested in

Experiment 1 by comparing recall patterns of short and long words using written recall and reconstruction of order. There was no difference in the recall pattern as a function of the type of test. Furthermore, the absence of a word length effect when short and long words were equated for neighbourhood size was demonstrated using a spoken serial recall task in Experiment 5. In addition, in the present Experiment 8, there was an effect of word length when neighbourhood size was confounded, 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 my argument is correlational in nature (i.e., emphasizing the similar effect of concurrent articulation on both word length and neighbourhood size manipulations), conclusions from Experiments 6 and 7 are not particularly strong. This concern is only partly warranted. I acknowledge that finding that concurrent articulation abolishes the neighbourhood effect does not necessarily mean that it is the same thing as the word length effect. However, had Experiments 6 and 7 failed to find that concurrent articulation abolishes the neighbourhood size effect, this prediction would have been falsified. It was a distinct possibility that neighbourhood size might be like manipulations of concreteness, frequency, imageability, and word class, and be immune to concurrent articulation (e.g., Acheson et al., 2010; Bourassa & Besner, 1994; Gregg et al., 1989; Tehan & Humphreys, 1988). Thus, the experiment is a strong test of the hypothesis.

3.3 Accounting for Neighbourhood Effects

Why does a large neighbourhood size benefit immediate recall? This result is surprising, as large neighbourhood size has previously been associated with some detrimental effects. In particular, there is a large literature that shows that spoken word recognition is facilitated for words with smaller neighbourhoods compared to those with larger neighbourhoods (e.g., Luce & Pisoni, 1998). However, facilitative effects for words with large neighbourhoods have been shown on certain production -- as opposed to perception -- tasks. For example, Vitevitch (2002) showed that more errors were elicited for words with fewer similar sounding words (i.e., small neighbourhood) than words with more similar sounding words (i.e., large neighbourhood). Similarly, in a picture-naming task, words from small neighbourhoods were identified more slowly than words from large neighbourhoods (see also Vitevitch & Sommers, 2003). This is the reasoning behind placing the facilitative effects of neighbourhood size at output: increasing the number of neighbours enhances speech production but not speech perception.

3.3.1 Redintegration and Associative Networks

Roodenrys (2009; see also Roodenrys & Miller, 2008) suggested one way in which both phonological and orthographical neighbourhood size could have a beneficial effect on recall. Many models of memory posit that at retrieval, one major task facing the rememberer is the interpretation of degraded items. Typically, a redintegrative process is invoked which recruits additional information to help interpret the ambiguous remnants of the to-be-remembered items. If one were to assume that the degraded string serves as input to an interactive network, such as might be encountered in speech production, then the slight activation in the network accruing from the commonalities of the neighbours could readily lead to more successful redintegration of a target. In other words, the more neighbours you have, the more activation you will get in the interactive network and the easier the items are going to be to recall. Such a process could also be extended to account for other beneficial effects of linguistic or lexical factors, and this could be added to those models that already include a redintegrative component. Roodenrys (2009; see also Roodenrys & Miller, 2008) has suggested that the locus of the neighbourhood size effect is during redintegration. If noise is added during presentation or retrieval (i.e., concurrent articulation) it could remove the benefit of the large neighbourhood items by reducing the activation level.

Roodenrys's (2009) redintegration may sound counterintuitive: If all the neighbours of a word are activated at recall, words with a large neighbourhood should suffer from the competition between the neighbours. However, Roodenrys (2009) bases his redintegration hypothesis not on whole item representations but on sublexical information. Each word is represented on two levels in the interactive network. The first level is the lexical level. It includes whole word representations. The second level is called the sublexical level and includes phonemic information. According to McClelland and Rummelhart's (1981) connectionist model of word perception, when a word is perceived, it is first perceived at the feature level, then at the letter or phoneme level and finally at the word level. In other words, the activation first passes through the feature level, followed by the phoneme/letter level before being perceived at the word level. The activation is not unidirectional; activation can also pass from the word level back down to the letter/phoneme level.

When a word needs to be redintegrated in order to be recalled, only certain phonemes/letters of the word are still available in the memory trace. Those phonemes/letters are used as input in the network. Consequently, having more neighbours helps redintegration by causing more activation in the network and consequently, increasing the chances of correctly filling-in the missing information with the correct phonemes/letters.

For example, consider a situation where the word *mink* has to be recalled in a short-term serial recall task but that only the last three letters remain in the memory trace and the first consonant is missing. The letters *i*, *n*, and *k* would be activated in the interactive network and these letters would activate all the words in long-term memory that contain them in that specific order. Words that contain more letters are activated to a greater degree than words that contain fewer letters. In this example, the trace of the word *mink* would strongly activate the word *mink*, but also the words *monk* or *mint* to a lesser degree. The activated words then feeds activation back to the letters they contain. Here, the three activated words, mink, monk, and mint would activate the missing letter *m*. Because the letter *m* is activated, it is now easier to recall *mink* from the memory trace containing *i*, *n*, and *k*.

The abolishment of the neighbourhood size effect in mixed list could be explained by differences in activation in mixed lists compared to pure lists. Both small and large neighbourhood words need identifying. The small neighbourhood words would be helped by the removal of three harder to redintegrate items, which are replaced by easier items (large neighbourhood items). The addition of large neighbourhood items would also create more activation, helping redintegration of small neighbourhood items. Large neighbourhood items would be hindered by the addition of items that are harder to redintegrate.

3.4 Implications for Theories

It was noted earlier that only one set of English words reliably produces a time-based word length effect, whereas all other sets tested so far do not (e.g., Lovatt et al., 2000; Neath et al., 2003). To this, I now add the evidence that previous demonstrations of the syllable-based word length effect may be due to a confound between word length and neighbourhood size in the stimulus sets, and when this confound is removed, so too are the effects of length.

3.4.1 Phonological Loop

To the extent that additional sets of stimuli can be found in which short and long words are appropriately equated and no word length effect emerges, models and theories based on the phonological loop (e.g., Baddeley, 1986, 1992; Burgess & Hitch, 1999; Page & Norris, 1998) are critically compromised. The basic architecture of these models requires that a word length effect be observed; if no such effects are observable, then the processes and architecture that predict the word length effect would need to be removed. Doing so, however, would also remove the model's ability to account for many other aspects of immediate memory. Furthermore, both a time-based word length effect and a syllable-based word length effect are abolished or greatly attenuated by concurrent articulation. Concurrent articulation is seen as preventing or interfering with articulatory rehearsal, which prevents the decaying traces from being refreshed. The problem for the phonological loop accounts is explaining why there are sometimes no effects of word length and why concurrent articulation affects the neighbourhood size effect.

3.4.2 Feature Model

The implications for accounts based on item properties are different. The account offered by the Feature Model (Neath & Nairne, 1995) does not require a time-based word length effect, and so the lack of one is not a fundamental problem, but it does make an incorrect prediction about the syllable-based word length effect in mixed lists (see Hulme et al., 2004; Hulme, Neath, Stuart, Shostak, Surprenant, & Brown, 2006), and the results of Experiment 1 compound this problem. According to the Feature Model, short words should always be better recalled than long words, no matter the composition of the list. However, if length *per se* is no longer a factor that needs to be explained, the processes that produce a word length effect can be removed. Unlike the case for models based on the phonological loop, removing these word-length specific processes does not affect the Feature Model's ability to account for the other core phenomena. Indeed, because a rudimentary redintegrative process was included in the original version of the model (Nairne, 1990), it may be possible to add the beneficial redintegrative effects of a large neighbourhood. Moreover, concurrent articulation has always been viewed as adding noise (Nairne, 1990; see also Murray et al., 1988). If this is the case, then it is easy to explain the abolition of the neighbourhood size effect by concurrent articulation.

3.4.3 Brown and Hulme Model

The Brown and Hulme (1995) model also explained the effects of length based on item-specific factors, and also made incorrect predictions about recall of short and long words in mixed lists. As the model was intended to demonstrate that rehearsal was not necessary for the word length effect, its scope and purpose was limited. With the demonstration that length effects are not always observed, the fundamental assumption of this account, differential decay rates, is also questioned. This does not, of course, make the model meaningless; rather, it continues to serve as an existence proof that rehearsal is not necessary in order to explain certain immediate memory effects.

3.4.4 SIMPLE

The final model considered (Hulme et al., 2004; see also Neath & Brown, 2006) is based on the framework of SIMPLE (Scale Invariant Memory and Perceptual Learning; Brown et al., 2007). SIMPLE is a relative distinctiveness model and assumes that items are represented on one or more dimensions. An item that "stands out" on its dimension (or position in multidimensional space) will be better recalled than one that has lots of neighbours. The word length effect was explained by noting that short words are typically more distinctive (i.e., easier to apprehend) than long items. In mixed lists, long words benefit from emergent distinctiveness, that is, compared to the short items, they now "stand out" more than when presented in a pure list. Indeed, when only one long item appears in a list of short items, it is in fact recalled better than the short items (Hulme et al., 2006). The challenge for SIMPLE is resolve the paradox that items with fewer close neighbours are seen as more distinct but items with more orthographic (or phonological) neighbours are recalled better. SIMPLE does not yet include a redintegration stage.

3.5 Time and Memory

As Nairne (2002) notes in his comprehensive review, the so-called "standard model" of short-term or working memory posits that items decay unless offset with rehearsal, and rehearsal speed is assumed to be related to pronunciation time. If items take longer to rehearse, fewer of them can be refreshed and so fewer can be recalled compared to shorter items. The syllable-based word length effect is a highly robust phenomenon demonstrated in numerous studies. However, those studies have confounded length with orthographic neighbourhood size (see Table 1). When short and long words are equated for neighbourhood size, no word length effect is observed. This result is devastating for any model that incorporates the idea of time-based decay offset by rehearsal. It is simply not possible to explain why three-syllable words are recalled as well as one-syllable words when three-syllable words take longer to rehearse and so should be more prone to decay.

Historically, decay as a cause of forgetting has been vigorously and repeatedly rejected (e.g., McGeoch, 1932; Osgood, 1953), and it was not until the so-called cognitive revolution that theories started including decay and de-emphasizing other causes of forgetting (see Neath, 1998, for a review). Now, it appears as though the tide is turning once again away from time-based decay as an explanatory construct. Indeed, there are an increasing number of empirical (e.g., Berman, Jonides, & Lewis, 2009) and theoretical (e.g., Lewandowsky, Oberauer, & Brown, 2009) papers which suggest that time-based decay simply does not exist; instead, forgetting is attributed to a number of different causes, including interference, changed cues, inappropriate processing, relative distinctiveness, and the like. The results of this thesis add to this growing consensus.

3.6 Conclusion

The word length effect has been termed one of the benchmark findings that any theory of short-term memory must account for. Indeed, the effect was one that led directly to the development of working memory and the phonological loop. Experiments 1a and 1b replicated the typical effects of length when short and long words were equated on all relevant dimensions previously identified in the literature. However, previous studies investigating the effect of word length have confounded length with orthographic neighbourhood size. In English, short words are more likely to have a larger neighbourhood size than long words, and Experiment 5 replicated the finding that words with a large neighbourhood are recalled better than words with a small neighbourhood. When a new set of short and long items were also equated for neighbourhood size, the word length effect disappeared. Furthermore, Experiment 6 and 7 showed that concurrent articulation abolished the neighbourhood effect, like it does the word length effect, for both words and nonwords. Finally, Experiment 8 showed that neighbourhood size is a better predictor of recall performance than word length. These findings add to the growing literature showing that performance in many memory tasks is affected by particular properties of the stimulus set used, and compounds the problems for theories of memory, such as working memory, that include decay offset by rehearsal as a central feature.

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. *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. *Quarterly Journal of Experimental Psychology, 44A*, 1-31.

Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science, 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. *Quarterly Journal of Experimental Psychology, 55A*, 353-369.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *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.

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 317-333.

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.

Bireta, T. J., Fry, S. E., Jalbert, A., Neath, I., Surprenant, A. M., Tehan, G., & Tolan, G. A. (2010). Backward recall and benchmark effects of working memory. *Memory & Cognition, 38*, 279-291.

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.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114*, 539-576.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*, 551-581.

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55*, 627-652.

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

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). San Diego, CA: Academic Press.

Coltheart, V., Mondy, S., Dux, P. E., & Stephenson, L. (2004). Effects of orthographic and phonological word length on memory for lists shown at RSVP and STM rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 815-826.

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., Kellar, 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 & Language, 31*, 1-17.

Cowan, N., Wood, N. L., & Borne, D. N. (1994). Reconfirmation of the short-term storage concept. *Psychological Science, 5*, 103-106.

Cowan, N., Wood, N. L., & Nugent, L. D., & Treisman, M. (1997). There are two word-length effects in verbal short-term memory: Opposed effects of duration and complexity. *Psychological Science, 8*, 290-295.

Dosher, B. A., & Ma, J.-J. (1998). Output loss or rehearsal loop? Output-time versus pronunciation-time limits in immediate recall for forgetting-matched materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 316-335.

Goh, W. D., & Pisoni, D. B. (2003). Effects of lexical competition on immediate memory span for spoken words. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *56*, 929-954.

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., Neath, I., Stuart, G., Shostak, L., Surprenant, A. M., & Brown, G. D. A. (2006). The distinctiveness of the word-length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 586-594.

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.

Hulme, C., & Tordoff, V. (1989). Working memory development: The effects of speech rate, word length, and acoustic similarity on serial recall. *Journal of Experimental Child Psychology*, *47*, 72-87.

Jalbert, A., Neath, I., & Surprenant, A. M. (in press). Does length or neighborhood size cause the word length effect? *Memory & Cognition*.

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, &*

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.

Lewandowsky, S., & Oberauer, K. (2008). The word-length effect provides no evidence for decay in short-term memory. *Psychonomic Bulletin & Review, 15*, 875-888.

Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences*, *13*, 120-126.

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. *Quarterly Journal of Experimental Psychology, 53A*, 1-22.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighbourhood activation model. *Ear & Hearing, 19*, 1-36.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighbourhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. Cambridge, MA: 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: Part 1. An account of basic findings. *Psychological Review*, *88*, 375-407.

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review*, *39*, 352-370.

McNeil, A. M., & Johnston, R. S. (2004). Word length, phonemic, and visual similarity effects in poor and normal readers. *Memory & Cognition*, *32*, 687-695.

Medler, D. A., & Binder, J. R. (2005). MCWord: An on-line orthographic database of the English language. http://www.neuro.mcw.edu/mcword/

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.

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. (1988). Familiarity and lexicality as determinants of the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 694-699.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*, 251-269.

Nairne, 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. (1998). *Human memory: An introduction to research, data, and theory*. Hove, UK: Psychology Press.

Neath, I. (2000). Is "working memory" still a useful concept? *Contemporary Psychology, 45*, 410-412.

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.

Neath, I., & Brown, G. D. A. (2006). SIMPLE: Further applications of a local distinctiveness model of memory. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 201-243). San Diego, CA: Academic Press.

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

Osgood, C. E. (1953). *Method and theory in experiment psychology*. New York, US: Oxford University Press.

Page, M. P. A., & Norris, D. (1998). The Primacy Model: A new model of immediate serial recall. *Psychological Review, 105*, 761-781.

Page, M. P. A., & Norris, D. (2003). The irrelevant sound effect: What needs modeling, and a tentative model. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 56*, 1289-1300.

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- and long-term memory in the verbal domain* (pp. 178-197). London: 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., 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.

Roodenrys, S., & Miller, L. M. (2008). A constrained Rasch model of trace redintegration in serial recall. *Memory & Cognition, 36*, 578-587

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. *Quarterly Journal of Experimental Psychology, 51A*, 283-304.

Surprenant, A. M., LeCompte, D. C., & Neath, I. (2000). Manipulations of irrelevant information: Suffix effects with articulatory suppression and irrelevant speech. *Quarterly Journal of Experimental Psychology, 53A*, 325-348.

Surprenant, A. M., Neath, I., & Brown, G. D. A. (2006). Modeling age-related differences in immediate memory using SIMPLE. *Journal of Memory and Language, 55*, 572-586.

Surprenant, A. M., Neath, I., & LeCompte, D. C. (1999). Irrelevant speech, phonological similarity, and presentation modality. *Memory, 7*, 405-420.

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., & Turcotte, J. (2002). Word length effects are not due to proactive interference. *Memory, 10*, 139-149.

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 sublexical and lexical representations on the processing of spoken words in English. *Clinical Linguistics & Phonetics, 17*, 487-499.

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 Behaviour, 11*, 644-648.

Yates, M., Locker, L. Jr., & Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review*, *11*, 452-457.

Appendix A

The short and long words used in Experiments 1a and 1b.

	Conc.	Fam.	Imag.	KF Freq.	TL Freq.	C Freq.	Orth.	Orth. Freq.
aisle	509	503	528	6	72	7.6	1	0.2
beam	502	476	539	21	127	9.2	9	23.7
draw	442	542	435	56	428	58.7	6	14.4
flood	553	523	598	19	325	15.6	2	158.3
howl	434	447	536	4	72	2.6	5	6.8
joke	388	580	483	22	230	34.6	6	6.2
lice	543	397	532	2	4	1.9	13	251.5
mink	589	524	604	5	27	3.3	16	42.2
pain	426	569	502	88	541	77.7	11	47.6
peal	402	451	433	1	13	1.1	15	39.5
pint	483	536	487	13	92	10.1	12	8.4
rose	608	556	623	86	801	80.1	16	30.5
sale	364	555	422	44	403	33.7	20	55.6
threat	335	524	408	42	108	64.3	2	28.2
wrath	304	466	377	9	51	7.5	1	0.1
Mean	458.8	509.9	500.5	27.9	219.6	27.2	9.0	47.5
Stdev.	92.1	52.2	75.1	29.2	233.8	29.1	6.2	68.5
abundant	351	524	443	9	50	9.8	0	0.0
accident	419	564	518	33	399	50.3	1	0.1
approval	267	526	375	51	108	29.7	0	0.0
article	479	533	421	68	550	41.0	0	0.0
avenue	539	529	564	46	320	24.5	1	1.7
foreigner	492	499	516	4	92	7.4	0	0.0
hexagon	559	387	527	1	4	0.8	0	0.0
musician	564	558	585	23	72	5.3	0	0.0
occasion	346	566	305	58	424	64.8	0	0.0
paragraph	493	559	482	12	72	10.0	0	0.0
recital	476	468	495	8	27	3.5	0	0.0
sedative	459	423	459	1	13	1.3	0	0.0
sympathy	278	501	402	36	228	31.8	0	0.0
telegraph	547	460	518	21	126	3.0	0	0.0
telephone	619	605	655	76	800	102.9	1	0.1
Mean	459.2	513.5	484.3	29.8	219.0	25.7	0.2	0.1

Stdev.	106.7	58.6	87.9	25.1	233.9	29.1	0.4	0.4

Note: Conc. = concreteness; Fam. = familiarity; Imag. = imageability; KF Freq. = Kucera-Francis frequency; TL Freq. = Thorndike-Lorge frequency; C. Freq. = CELEX frequency; Orth. = number of orthographic neighbours; Orth. F. = CELEX frequency of orthographic neighbours. The first four measures are from the Medical Research Council Psycholinguistics database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm) and the remaining measures are from the MCWord database of Medler and Binder (2005; http://www.neuro.mcw.edu/mcword/).

Appendix B

The short and long words used in Experiments 2 and 4.

	Conc.	Fam.	Imag.	KF Freq.	TL Freq.	C. Freq.	Orth.	Orth. Freq.
disc	553	466	575	6	5	8.0	1	23.08
grief	303	505	480	10	137	16.2	1	47.00
guess	247	585	330	56	933	60.3	1	25.46
numb	379	487	477	4	55	4.2	1	11.72
phase	360	516	319	72	91	31.2	1	9.16
rogue	424	378	478	1	11	2.6	1	3.45
shriek	481	458	515	5	101	3.9	1	6.19
sponge	597	538	577	7	51	7.5	1	1.43
square	516	576	610	143	573	92.0	1	3.69
squeak	461	506	492	1	22	2.9	1	1.84
teeth	618	593	611	103	405	82.1	1	10.53
throng	400	377	452	3	60	3.6	1	9.76
wheat	594	510	577	9	158	30.3	1	4.22
Mean	456.4	499.6	499.5	32.3	200.2	26.5	1.0	12.1
Stdev.	117.3	68.7	94.7	46.7	276.0	31.7	0.0	12.9
assemble	394	482	413	9	98	5.5	1	26.95
avenue	539	529	564	46	320	24.5	1	1.67
depression	303	541	453	24	244	24.9	1	11.36
destroyer	513	448	508	2	14	3.9	1	45.81
emission	397	446	416	32	1	1.5	1	3.39
fisherman	567	471	610	5	70	3.6	1	6.19
gentleman	516	537	559	28	580	24.4	1	29.09
insolent	311	388	357	2	25	2.3	1	1.43
minister	563	500	584	61	228	101.0	1	13.68
officer	550	549	593	101	585	79.3	1	33.43
photograph	590	551	618	18	342	28.7	1	1.49
primary	326	497	367	96	58	40.9	1	1.96
socialist	443	480	352	21	17	56.7	1	33.26
Mean	462.5	493.8	491.8	34.2	198.6	30.6	1.0	16.1
Stdev.	105.1	48.4	102.2	33.3	207.3	31.5	0.0	15.5

Appendix C

The short and long words used in Experiment 3.

	Conc.	Fam.	Imag.	KF Freq.	TL Freq.	C. Freq.	Orth.	Orth. Freq.
birch	620	518	561	2	34	2.5	2	38.4
broad	399	523	463	84	282	42.3	2	40.3
cloud	554	553	595	28	367	32.5	2	10.5
flask	595	401	614	5	16	4.3	2	14.1
gloom	399	475	429	14	74	11.3	2	5.5
itch	488	526	486	5	20	2.5	2	12.6
myth	334	514	359	35	22	19.9	2	2.0
pledge	360	442	408	3	70	5.5	2	0.6
prune	611	444	578	1	104	2.0	2	3.8
split	417	514	445	30	119	38.7	2	0.3
swarm	406	463	488	3	76	3.3	2	0.4
trend	328	503	373	46	75	22.7	2	3.2
tweed	570	429	540	5	76	5.1	2	0.2
vault	550	445	550	2	35	3.6	2	19.6
Mean	473.6	482.1	492.1	18.8	97.9	14.0	2.0	10.8
Stdev.	107.2	45.3	82.9	23.9	102.4	14.5	0.0	13.5
altitude	373	420	472	4	53	4.4	2	41.8
charity	373	518	445	8	158	14.3	2	5.6
convention	488	466	502	28	251	16.1	2	2.4
deduction	327	492	316	12	20	5.9	2	12.1
invader	485	402	419	1	15	1.0	2	3.6
lecturer	561	574	551	6	24	7.3	2	9.1
observer	505	469	489	16	82	12.7	2	20.5
opening	455	542	462	83	124	62.6	2	0.0
procession	500	462	534	5	89	12.8	2	11.9
radio	615	644	613	120	393	73.6	2	6.8
retailer	521	429	445	1	27	1.0	2	0.6
scavenger	486	474	501	1	10	0.6	2	0.2
treasurer	557	511	493	14	34	4.5	2	4.1
vocation	349	458	404	3	19	2.7	2	14.6
Mean	471.1	490.1	474.7	21.6	92.8	15.7	2.0	9.5
Stdev.	86.2	64.5	71.1	35.4	110.3	22.9	0.0	11.1

Appendix D

Nonwords used in Experiment 7.

	Phon.	Phon. Freq.
Small Neighbourhood		
chush	10	27
googe	6	3
grair	4	115
joach	5	57
jorth	6	370
nadge	10	7
olled	3	588
rorch	12	158
tedge	9	883
touge	3	175
zarsh	4	62
zoule	7	11
Mean	6.58	204.67
Stdev	3.03	275.76
Large Neighbourhood		
boarg	22	916
chone	21	478
coose	24	2591
gares	30	546
ghoss	20	4517
jight	26	4217
korch	23	2628
lorse	27	2895
petch	20	73
puice	20	427
sheed	43	3986
wroke	39	665
Mean	26.25	1994.92
Stdev	7.61	1660.08

Appendix E

The short and long nonwords used in Experiment 8.

	Orth.	Orth. Freq.			Orth.	Orth. Freq.
Short Small Neighbourhood				Long Small Neighbourhood		
clende	0	0		aftin	0	0
colmes	0	0		agarld	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
Stdev	0	0		Stdev	0	0
Short Large Neighbourhood				Long Large Neighbourhood		
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
Stdev	0.51	128.53		Stdev	0.51	125.18

CENTRE FOR NFLD. STUDIES
JUL 18 1978
MEMORIAL UNIVERSITY
OF NEWFOUNDLAND