Journal of Experimental Psychology:
Learning, Memory, and Cognition
2002, Vol. 28, No. 6, 1019–1034

0278-7393/02/$5.00 DOI: 10.1037//0278-7393.28.6.1019

Word-Frequency and Phonological-Neighborhood Effects on Verbal Short-Term Memory

Steven Roodenrys
University of Wollongong

Charles Hulme
University of York

Alistair Lethbridge, Melinda Hinton, and Lisa M. Nimmo
University of Wollongong

Immediate memory span and maximal articulation rate were assessed for word sets differing in frequency, word-neighborhood size, and average word-neighborhood frequency. Memory span was greater for high- than low-frequency words, greater for words from large than small phonological neighborhoods, and greater for words from high- than low-frequency phonological neighborhoods. Maximal articulation rate was also facilitated by word frequency, phonological-neighborhood size, and neighborhood frequency. In a final study all 3 lexical variables were found to influence the recall outcome for individual words. These effects of phonological-word neighborhood on memory performance suggest that phonological information in long-term memory plays an active role in recall in short-term-memory tasks, and they present a challenge to current theories of short-term memory.

The present experiments explored the mechanisms underlying verbal memory span. We demonstrated that lexical factors, particularly word frequency and phonological neighborhood properties, affect how well words are recalled in short-term-memory tasks.

It is well established that performance in memory span tasks makes use of a system that relies heavily on some form of phonological coding (for reviews, see Baddeley, 1986; Penney, 1989). One of the most detailed and widely accepted accounts of the role of phonological coding mechanisms in memory span tasks comes from trace decay with rehearsal models (e.g., Baddeley, 1986; Brown & Hulme, 1995; Schweickert & Boruff, 1986). According to such models, items are encoded into traces in a phonological short-term store, and the memory traces are subject to passive decay. Such decay results in the traces of items fading within a short period of time. However, decay can be overcome by rehearsal, which involves subvocal articulation.

The most influential model of this type is the phonological loop model of Baddeley and his colleagues (e.g., Baddeley, 1986; Baddeley, Gathercole, & Papagno, 1998; Baddeley & Hitch, 1974). In this model, traces in the phonological store have a limited useful duration estimated to be between 1.8 and 2.2 s (Baddeley, Thomson, & Buchanan, 1975; Schweickert & Boruff, 1986; Standing, Bond, Smith, & Isely, 1980). These traces can be maintained for longer periods by subvocal articulation or rehearsal. Effectively, rehearsal involves a process of reencoding the traces of items into a phonological store before the traces have decayed to the point at which they can no longer be identified.

It has been argued that trace decay with rehearsal models are inadequate for both theoretical and empirical reasons (Brown & Hulme, 1995). One critical issue is that speech rate, which has typically been interpreted as a measure of rehearsal rate, does not provide an adequate explanation for variations in memory span across materials. So, for example, it has been shown that memory span differences between words and nonwords (Hulme, Maughan, & Brown, 1991) and between high- and low-frequency words (Hulme et al., 1997; Roodenrys, Hulme, Alban, Ellis, & Brown, 1994) cannot be explained by differences in speech rate. It has been argued that differences in memory span between high- and low-frequency words and between words and nonwords are due to the differential effectiveness of a process that acts to redintegrate degraded memory traces at retrieval.

According to this revision of the trace-decay model, many memory traces that are retrieved from the short-term store will be degraded. The fact that words erroneously substituted in short-term serial recall tasks are often phonologically similar to the words they have replaced (e.g., Drewnowski & Murdock, 1980) shows that the loss of information occurs at a level lower than that of the whole word. It has been argued that decayed memory traces retrieved from the short-term store may be redintegrated or cleaned up using phonological information retrieved from long-term memory (e.g., Hulme et al., 1997; Schweickert, 1993). According to this explanation, memory performance is better for words than

Steven Roodenrys, Alistair Lethbridge, Melinda Hinton, and Lisa M. Nimmo, Department of Psychology, University of Wollongong, Wollongong, New South Wales, Australia; Charles Hulme, Department of Psychology, University of York, York, United Kingdom.

This research was supported by Australian Research Council Grant A79803822 to Steven Roodenrys and by Economic and Social Research Council of the United Kingdom Grant R000232 567. We thank Barbara Church, Steve Goldinger, and an anonymous reviewer for helpful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Steven Roodenrys, Department of Psychology, University of Wollongong, Northfields Avenue, Wollongong, New South Wales 2522, Australia. E-mail: steven_roodenrys@uow.edu.au

nonwords because nonwords lack a representation in long-term memory that can be used in the redintegration process. Similarly, performance is better for high-frequency than low-frequency words because the representations of high-frequency words in long-term memory are either more accessible or better specified than those of low-frequency words (Mandler, 1980; Scarborough, Cortese, & Scarborough, 1977).

A number of researchers have proposed models of immediate serial recall that incorporate a trace redintegration process at retrieval (Hulme et al., 1991; Nairne, 1990; Schweickert, 1993; Sperling & Speelman, 1970). An idea that is shared by these models is that redintegration involves comparison of degraded traces in short-term memory to a set of long-term-memory traces. From this it seems likely that both the size of the activated search set in long-term memory and the distinctiveness of these items from others in long-term memory will be important in determining the effectiveness of the redintegration process. Evidence that the size of the long-term memory search set is important comes from the finding that immediate serial recall of lists of words drawn from a single semantic category (which will enable participants to restrict the candidate search set in long-term memory) is better than that of lists drawn from more than one category (Poirier & Saint-Aubin, 1995; Wetherick, 1975). It has also been demonstrated that serial recall performance is better when the items on each trial are drawn from a small pool and presented repeatedly in different lists in the experiment (V. Coltheart, 1993; Conrad, 1963; Roodenrys & Quinlan, 2000). Under such conditions, it is plausible that participants are able to use knowledge of the item set to restrict the comparison set in long-term memory.

The distinctiveness of words used in a short-term-memory task in relation to other words in long-term memory is also likely to influence redintegration and, hence, immediate serial recall performance. Drewnowski and Murdock (1980) found that intrusion errors in a serial recall task tended to be phonologically related to the target item. This would be expected if redintegration relies on a comparison of information in the degraded phonological short-term-memory trace with a set of long-term-memory traces of words that is not restricted to the words presented in that particular list. This, in turn, suggests that the likelihood of correctly reconstructing a degraded short-term-memory trace will be influenced by how phonologically distinct the word is from other items in the lexicon. According to this view, a word that is relatively unusual or distinctive in its phonological form might be expected to be recalled better than a word that is similar to many other lexical entries. It is this possibility that we investigated in the following experiments.

To test this idea we needed a metric of the phonological similarity between words. We refer to the words that are similar to a target word as its *neighbors* and the set of such words as the *target word's neighborhood* (e.g., M. Coltheart, Davelaar, Jonasson, & Besner, 1977; Goldinger, Pisoni, & Logan, 1991). Words from large neighborhoods have many similar-sounding neighbors. For example, the word *void* has far fewer words that sound similar to it than does the word *wipe*. The most straightforward prediction from the redintegration hypothesis would be that words with few neighbors are easier to redintegrate, and therefore easier to recall, than words that have many neighbors. To investigate this idea we needed a definition of what constitutes a given word's neighborhood. A working definition that we have adopted is that a phonological neighborhood is the set of words that differ from the target word by a single phoneme. For example, for the word *sent*, *bent* is a neighbor, whereas *belt* is not. To return to the earlier example, by this definition *void* has only 2 neighbors (*voice* and *vied*) whereas *wipe* has at least 12 neighbors (including *pipe*, *wine*, and *weep*).

There is very little previously published data relevant to neighborhood effects in serial recall. Goldinger et al. (1991) were interested in relating perception to memory for words and used previous work on neighborhood effects in perception. Neighborhoods have two properties that have been identified as having an influence on the perceptual processing of words: the size of the neighborhood and the average frequency with which the words in the neighborhood occur in normal language usage (i.e., neighborhood frequency). When perceiving spoken words in noise, high-frequency words are better recognized than low-frequency words, words from larger neighborhoods are less well recognized than those from smaller neighborhoods, and words with high-frequency neighbors are less well recognized than those with only low-frequency neighbors (Luce & Pisoni, 1998). In other speech-perception tasks using clear presentation, such as a same–different matching task, performance is faster but not more accurate for words with fewer neighbors (Vitevitch & Luce, 1999). Goldinger et al. (1991) selected words that they argued were either easy to identify (high-frequency words from small, low-frequency neighborhoods) or hard to identify (low-frequency words from large, high-frequency neighborhoods) and used them in a serial recall task. The results showed better recall performance for the easy-to-identify words than the hard-to-identify words. However, it is not possible to draw any conclusions from this study about the potentially separate effects of the neighborhood variables on verbal short-term-memory performance because the effects obtained could easily be explained in terms of word frequency (the easy-to-perceive [high-frequency] words were recalled better than the hard-to-perceive [low-frequency] words).

From a slightly different perspective, it is also relevant that Gathercole (1995) found that nonwords that are rated as being similar to words (high word likeness) are easier to repeat than nonwords that are of low word likeness. This suggests that there may be facilitatory effects on memory performance based on the number of lexical forms activated by a given memory item. Roodenrys and Hinton (2002) have shown that serial recall of nonwords with many lexical neighbors is better than recall of nonwords with few neighbors. It is therefore not entirely clear from the available data whether one should expect facilitation or inhibition of memory performance as a function of lexical neighborhood size.

There are a number of recent models that differ greatly in how they attempt to explain the phenomena observed in immediate serial recall tasks. Some of these models can be seen as specifications of the more conceptual phonological loop and explicitly integrate verbal short-term memory with other language processes (e.g., Gupta & MacWhinney, 1997). Other models (e.g., Brown, Preece, & Hulme, 2000; Nairne, 1990) simulate serial recall with an emphasis on modality-independent processes rather than processes specific to the recall of linguistic material. One thing that many current models have in common is a mechanism that is conceptually equivalent to a redintegration process.

In some models the redintegration process is explicit. These models posit a mechanism that retains the order of the items that have been presented in some temporary storage mechanism. However, when an item is retrieved from a temporary short-term store it is then identified by comparison with information held in permanent (secondary) memory. In some models this is a traditional conceptualization of a long-term store (e.g., Brown et al., 2000; Nairne, 1990; Page & Norris, 1998), and the redintegration process relies on the activation of items in the long-term store. Activation level is determined by similarity to the information retrieved from the short-term mechanism, and the most active item is selected for output. Whether the items in the long-term store actively compete to be output or are the passive objects of the selection process, this approach predicts that, all things being equal, words with more neighbors are less likely to be accurately recalled than words with fewer neighbors.

Connectionist models of serial recall (e.g., Burgess & Hitch, 1999; Gupta & MacWhinney, 1997) offer a different conceptualization of storage from the traditional models that separate short- and long-term storage. However a redintegration process is implicit in these connectionist models. For example, in the Burgess and Hitch (1999) model, short-term weights reinstate a pattern of activation across a set of nodes. This process is noisy, and a set of long-term weights clean up the initial pattern to produce a word for recall. This approach also predicts that neighbors will actively compete with the target word unless the set of items in the model is restricted to those that are used in the list.

With these ideas in mind, we designed the following experiments to investigate the combined effects of word frequency, neighborhood size, and neighborhood frequency on immediate serial recall. Ideally, we would investigate the effects of all three of these variables in a single experiment so that all two-way interactions and the three-way interaction could be tested. However, it is not possible to find enough stimuli to fill the resulting eight cells in such a design and to adequately manipulate the levels of the independent variables. In the first three experiments we therefore examined each pairwise combination of the three variables.

To conduct the experiments, we needed to calculate neighborhood size and frequency information for a large number of words. In their experiments on auditory perception Luce and Pisoni (1998) used a definition of a neighborhood that included all words that differed from the target word by the substitution, deletion, or addition of a single phoneme in any position. For our present experiments a slightly more restrictive definition was used, in which a neighborhood is defined as all of the words that differ from the target word by the substitution of a single phoneme at any position. This definition was deemed more appropriate for the present purpose because it seems likely that words with one phoneme more or one less than the target are more likely to differ in syllable structure. In any case, calculations for three-phoneme words show that the two definitions give measures of neighborhood size that correlate extremely highly ($r = .97$). Neighborhood frequency was defined as the mean of the log adjusted frequency of all of the neighbors of the word. A detailed description of how the neighborhood frequencies were calculated is given in Appendix A.

Experiment 1

Experiment 1 examined the effects of word frequency and neighborhood size on immediate memory span. Our prediction, based on the idea that the redintegration of degraded phonological memory traces relies on speech-processing mechanisms, was that the effects of word frequency and neighborhood size found in auditory word recognition (Luce & Pisoni, 1998) will also be observed in memory span. We therefore predicted that memory performance would be poorer for low-frequency than for high-frequency words (as has been found previously; e.g., Hulme et al., 1997) and that words from large neighborhoods will be recalled less well than words from small neighborhoods.

Method

Participants. Twenty-four first-year psychology students at the University of Wollongong, Wollongong, Australia, participated in the experiment in compliance with a course requirement.

Materials. We used four sets of 16 words throughout the experiment, manipulating target word frequency and neighborhood size. High-frequency sets comprised words with a frequency greater than 50 words per million according to the norms of Kučera and Francis (1967), and low-frequency sets comprised words with a frequency of 6 words per million or less. The manipulation of frequency was highly significant (high, $M = 154$, $SD = 114$; low, $M = 2$, $SD = 2$), $t(31.01) = 7.54$, $p < .01$.

Large neighborhood size sets comprised words with 18 or more neighbors, whereas small neighborhood sets comprised words with 14 neighbors or fewer. The manipulation of neighborhood size was highly significant (large, $M = 28.8$, $SD = 7.5$; small, $M = 8.8$, $SD = 2.8$; $t[39.79] = 14.03$, $p < .01$). The words were all three phonemes in length, with a consonant-vowel-consonant (CVC) structure, and they matched across conditions on neighborhood frequency. Words were selected to minimize phonological similarity within the conditions and control for this across conditions. The word sets and their mean word frequency, neighborhood size, and neighborhood frequency are presented in Appendix B.

Procedure. The experiment was conducted in a single session of approximately 1 hr. All of the tasks in the experiment were controlled by a Power Macintosh G3 (Apple Computer; Cupertino, CA) computer using an external amplified speaker to present the previously recorded and digitized items. Each participant completed a memory span procedure with spoken recall and a speech-rate measurement for each condition. The order of testing for both factors, word frequency and neighborhood size, was counterbalanced across participants.

In each condition the participants were initially presented with each word individually and asked to repeat it to check their audibility. Following this, memory span was measured using an up-and-down method (Watkins, 1977) beginning with a list length of three words. The participants were presented with 16 lists in each condition. When a participant successfully recalled a list, the next list presented was one item longer. When an error was made, the next list presented was one item shorter. Each list of words was drawn at random without replacement from the word pool for that condition and presented at a rate of one word per second. Memory span was calculated as the average list length presented, ignoring Lists 1 and 2 but including a length for the 17th list that was not presented (the length of which was determined by performance on the 16th list).

Following the memory span task the participant's speech rate was measured. For each condition the participants were presented with eight word pairs. They were instructed to repeat each pair 10 times as quickly as possible, and the time taken to do this was recorded. The mean of these eight times was then transformed into a measure of speech rate in words per second.

Results

Table 1 shows the speech-rate and memory span scores for each condition. It is apparent that, as we expected, memory span is

Table 1
Mean Memory-Span and Speech-Rate Scores in Experiment 1

	High word frequency				Low word frequency			
	Large NS		Small NS		Large NS		Small NS	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Memory span	5.11	0.75	4.69	0.46	4.53	0.62	4.43	0.58
Speech rate	3.52	0.80	2.90	0.52	3.20	0.66	2.95	0.58

Note. Speech-rate scores are shown in words per second. NS = neighborhood size.

better for high-frequency than low-frequency words, whereas contrary to prediction, memory span appears to be better for words from large neighborhoods than those from small neighborhoods.

The memory span scores were subjected to a two-way repeated measures analysis of variance (ANOVA) in which the variables were word frequency and neighborhood size. This analysis revealed a highly significant effect of word frequency, $F(1, 23) = 36.93$, $p < .01$, $MSE = 0.12$; an effect of neighborhood size, $F(1, 23) = 15.72$, $p < .01$, $MSE = 0.10$; and a significant Word Frequency $\times$ Neighborhood Size interaction, $F(1, 23) = 4.84$, $p < .05$, $MSE = 0.12$. Tests of simple effects revealed that the effect of frequency was significant for words from both large and small neighborhoods, $t(23) = 4.83$, $p < .01$, and $t(23) = 3.65$, $p < .01$, respectively. The effect of neighborhood size was significant for high-frequency words, $t(23) = 3.79$, $p < .01$, but not for low-frequency words, $t(23) = 1.29$, *ns*.

An equivalent analysis on the speech-rate scores revealed a significant effect of word frequency, $F(1, 23) = 24.25$, $p < .01$, $MSE = 0.02$; a highly significant effect of neighborhood size, $F(1, 23) = 69.08$, $p < .01$, $MSE = 0.07$; and a significant Word Frequency $\times$ Neighborhood Size interaction, $F(1, 23) = 28.04$, $p < .01$, $MSE = 0.03$. Tests of simple effects revealed that the effect of neighborhood size was significant for both high- and low-frequency words, $t(23) = 7.90$, $p < .01$, and $t(23) = 6.09$, $p < .01$, respectively. The effect of frequency was significant for words from large neighborhoods, $t(23) = 6.52$, $p < .01$, but not for words from small neighborhoods, $t(23) = 1.14$, *ns*.

To evaluate the extent to which any effects of word frequency or neighborhood size may be explicable in terms of speech-rate differences between the stimuli, we conducted an analysis of the memory span scores, using speech rate as a covariate. This analysis partials out any variation in span scores attributable to linear effects of speech rate.

This analysis revealed a significant effect of word frequency, $F(1, 22) = 25.58$, $p < .01$, $MSE = 0.12$, which means that with the effects of speech rate partialed out a significant difference in memory span exists between high- and low-word-frequency sets. The main effect of neighborhood size was only marginally significant, $F(1, 22) = 3.29$, $p < .09$, $MSE = 0.11$, however this effect did not interact significantly with frequency, $F(1, 22) < 1$, $MSE = 0.12$.

Discussion

This experiment, in line with previous experiments (Hulme et al., 1997; Roodenrys et al., 1994; Tehan & Humphreys, 1988), reveals a highly significant effect of word frequency on immediate serial recall that is independent of differences in how quickly the words can be spoken. However, contrary to our prediction from the redintegration view and several models of serial recall, memory span is actually better for words from large neighborhoods than for words from small neighborhoods. We consider the interpretation of this surprising result after the results of two further experiments are presented.

Experiment 2

The aim of this experiment was to manipulate the frequency of the stimulus words and the frequency of the stimulus words' neighborhoods. Once again we predicted that high-frequency words would be better recalled than low-frequency words.

None of the models outlined in the introduction make predictions about the effects of the frequency of neighbors, however this variable has been shown to be important in other lexical tasks. In an auditory word-recognition task Luce and Pisoni (1998) found that words from high-frequency neighborhoods were harder to identify. In naming visually presented words, on the other hand, high-frequency neighborhoods facilitate performance (see Andrews, 1997, for a review). Given the effect of neighborhood size observed in Experiment 1, it seemed possible that we would also observe an effect of neighborhood frequency even if our models were unable to predict the effect.

Method

Participants. Twenty-four first-year psychology students at the University of Wollongong participated in the experiment in compliance with a course requirement.

Materials. We used four sets of 15 words throughout the experiment, factorially manipulating target-word frequency and neighborhood frequency. All four sets of items were matched on neighborhood size and length in terms of number of phonemes. High-word-frequency sets comprised words with a frequency greater than 54 words per million according to the norms of Kučera and Francis (1967), and low-word-frequency sets comprised words with a frequency of 10 words per million or less. The manipulation of frequency was highly significant (high, $M = 162$, $SD = 119$; low, $M = 4$, $SD = 3$), $t(29.04) = 7.27$, $p < .01$.

High-neighborhood-frequency sets comprised words with neighborhood-frequency ratings greater than 70, whereas low-neighborhood-frequency sets comprised words with a neighborhood-frequency rating of 60 or less. (See Appendix A for an explanation of the neighborhood-frequency ratings.) The manipulation of neighborhood frequency was highly significant (high, $M = 78$, $SD = 6$; low, $M = 54$, $SD = 3$), $t(42.44) = 19.40$, $p < .01$. The word sets and their mean word frequency, neighborhood frequency, and neighborhood size are presented in Appendix C.

Procedure. The procedural details were the same as for the previous experiment with the exception that speech rate was measured by randomly selecting 7 pairs from the 15 stimuli for each participant. Each participant completed a memory span procedure with spoken recall and a speech-rate measurement for each condition. The order of testing for both factors, word frequency and neighborhood frequency, was counterbalanced across participants.

Results

Table 2 shows the speech rate and memory span scores for each condition. It is apparent that, as in Experiment 1, memory span is better for high-frequency than for low-frequency words. However, memory span is worse for words from low-frequency neighborhoods than for those from high-frequency neighborhoods, and this effect appears to be larger for low-frequency words.

The memory span scores were subjected to a two-way repeated measures ANOVA in which the variables were word frequency and neighborhood frequency. This analysis revealed a highly significant effect of word frequency, $F(1, 19) = 29.86$, $p < .01$, $MSE = 0.17$; a significant effect of neighborhood frequency, $F(1, 19) = 13.72$, $p < .01$, $MSE = 0.20$; and a significant Word Frequency × Neighborhood Frequency interaction, $F(1, 19) = 6.40$, $p < .05$, $MSE = 0.12$. Tests of simple effects revealed that the effect of frequency was significant for lists of words from both high- and low-frequency neighborhoods, $t(19) = 2.87$, $p < .01$, and $t(19) = 5.37$, $p < .01$, respectively. The effect of neighborhood frequency was significant for lists of low-frequency words, $t(19) = 5.70$, $p < .01$, but not for lists of high-frequency words, $t(19) = 1.21$, *ns*.

An equivalent analysis on the speech-rate scores revealed a significant effect of word frequency, $F(1, 19) = 59.80$, $p < .01$, $MSE = 0.04$; and a significant effect of neighborhood frequency, $F(1, 19) = 514.55$, $p < .01$, $MSE = 0.06$; but no significant Word Frequency × Neighborhood Frequency interaction, $F(1, 19) < 1$, $MSE = 0.02$.

To assess the extent to which effects in the memory span scores are independent of differences in speech rate, we subjected the memory span scores to an analysis of covariance (ANCOVA) with speech rate as the covariate. This analysis revealed a significant effect of word frequency, $F(1, 18) = 16.55$, $p < .01$, $MSE = 0.16$, which confirms that the effect of frequency remains significant after the effects of speech-rate differences have been partialed out. The main effect of neighborhood frequency was not significant, $F(1, 18) = 2.44$, $MSE = 0.17$, however, there was a significant Word Frequency × Neighborhood Frequency interaction, $F(1, 18) = 9.96$, $p < .01$, $MSE = 0.10$. This interaction has the same form as the equivalent effect in the analysis of the memory span scores without the covariate. Tests of simple effects showed that the effect of frequency was significant for lists of words from both high- and low-frequency neighborhoods, $t(18) = 2.45$, $p < .05$, and $t(18) = 2.38$, $p < .05$, respectively. The effect of neighborhood frequency was significant for lists of low-frequency words, $t(18) = 3.77$, $p < .01$, but not for lists of high-frequency words, $t(18) = 1.18$, *ns*.

As in Experiment 1, this experiment has revealed substantial effects of word frequency on short-term-memory span. However, words from high-frequency neighborhoods were recalled better than words from low-frequency neighborhoods. Once again, discussion is postponed until after a third experiment is reported below.

Experiment 3

This experiment examined the joint effects of neighborhood frequency and neighborhood size. On the basis of the results of Experiment 1, we predicted that memory span would be better for words from large neighborhoods than for those from small neighborhoods. On the basis of the results from Experiment 2, we predicted that memory span should be better for words from high-frequency neighborhoods.

Method

Participants. Twenty-four first year psychology students at the University of Wollongong participated in the experiment in compliance with a course requirement.

Materials. We used four sets of 16 CVC words throughout the experiment, factorially manipulating neighborhood frequency and neighborhood size of the target words. High-neighborhood-frequency sets comprised words with neighborhood-frequency ratings of 60 or greater, whereas low-neighborhood-frequency sets comprised words with a neighborhood frequency rating of 59 or less. The manipulation of neighborhood frequency was highly significant (high, $M = 68$, $SD = 5$; low, $M = 56$, $SD = 3$), $t(46.53) = 10.94$, $p < .01$. Large neighborhood size sets comprised words with 17 or more neighbors, whereas small neighborhood size sets comprised words with 10 neighbors or fewer. The manipulation of neighborhood size was highly significant (large, $M = 25$, $SD = 5$; small, $M = 7$, $SD = 2$; $t(43.78) = 18.75$, $p < .01$. The sets and their mean word frequency, neighborhood size, and neighborhood frequency are presented in Appendix D.

Procedure. The procedural details were the same as for Experiment 1.

Table 2
Mean Memory-Span and Speech-Rate Scores in Experiment 2

	High word frequency				Low word frequency			
	High NF		Low NF		High NF		Low NF	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Memory span	5.24	0.77	5.06	0.80	4.93	0.85	4.36	0.59
Speech rate	3.72	0.48	3.53	0.46	3.39	0.44	3.14	0.34

Note. Speech-rate scores are shown in words per second. NF = neighborhood frequency.

Results

Table 3 shows the speech rate and memory span scores for each condition. As we expected on the basis of the results of Experiments 1 and 2, memory span is better for words from high-frequency neighborhoods than from low-frequency neighborhoods, and words from large neighborhoods are recalled better than words from small neighborhoods. The faster speech rate for words from large neighborhoods found here also replicates the pattern found in Experiment 1.

The memory span scores were subjected to a two-way repeated measures ANOVA in which the variables were neighborhood size and neighborhood frequency. This analysis revealed a highly significant effect of neighborhood frequency, $F(1, 23) = 11.35$, $p < .01$, $MSE = 0.11$; and an effect of neighborhood size, $F(1, 23) = 26.60$, $p < .01$, $MSE = 0.09$; but no significant Neighborhood Frequency × Neighborhood Size interaction, $F(1, 23) = 2.56$, *ns*, $MSE = 0.15$. It is possible that the failure to detect this interaction is due to power limitations in the experiment.

The speech-rate data were subjected to an ANOVA in the same way as the recall data. This revealed a highly significant effect of neighborhood size, $F(1, 23) = 18.94$, $p < .01$, $MSE = 0.12$; but no effect of neighborhood frequency and no Neighborhood Frequency × Neighborhood Size interaction, $F(1, 23) = 0.004$, *ns*, $MSE = 0.03$; and $F(1, 23) = 1.09$, *ns*, $MSE = 0.04$, respectively.

The span scores were also subjected to an ANCOVA with speech rate as the covariate. This analysis revealed a significant effect of neighborhood frequency, $F(1, 22) = 11.24$, $p < .01$, $MSE = 0.11$; a significant effect of neighborhood size, $F(1, 22) = 14.60$, $p < .01$, $MSE = 0.10$, but no Neighborhood Frequency × Neighborhood Size interaction, $F(1, 22) = 1.73$, *ns*, $MSE = 0.14$.

Discussion

The results of this experiment confirm the major findings from the first two experiments: Words from large neighborhoods and words from high-frequency neighborhoods are better recalled than words from small neighborhoods or from neighborhoods composed of low-frequency words. The effect of neighborhood size in this experiment was stronger than in Experiment 1, and this difference may be explained by the fact that the manipulation of neighborhood size was more extreme here than in Experiment 1. It may be worth emphasizing that selecting stimuli with suitable properties for this experiment was very difficult. The words selected were of low frequency, and the manipulation of neighborhood frequency was not as large as in Experiment 2. This can account for the fact that overall the memory span scores in this experiment are slightly lower than in Experiments 1 and 2.

To summarize the results so far, Experiments 1–3 yielded clear evidence that the lexical (phonological) properties of spoken words have important effects on how well they are recalled in short-term-memory tasks. They have confirmed earlier findings in showing that high-frequency words are generally better recalled than low-frequency words (e.g., Hulme et al., 1997). Of more interest, however, are the novel effects of word neighborhoods and their interaction with word frequency on memory performance. We have found that the effect of word frequency on memory span occurs for words from large and small neighborhoods (Experiment 1) and for words from low- and high-frequency neighborhoods (Experiment 2). In addition, words from large neighborhoods are recalled better than those from small neighborhoods (Experiments 1 and 3). Finally, words that have high-frequency neighbors are better recalled than words with low-frequency neighbors (Experiment 3), and this effect is stronger for low-frequency words (Experiment 2).

The word-frequency effect observed in Experiments 1 and 2 was of a considerable magnitude, showing partial eta squared values of over .6 in the analysis of the span scores. It is not possible to compare directly the magnitude of the effect of the neighborhood variables in these experiments with that of the frequency effect because there is no direct comparison of the size of the manipulation of the variables on the underlying scale. Nevertheless, neighborhood size and neighborhood frequency both had partial eta squared values over .4 in the analysis of the span scores. These effect sizes are not negligible, and such effects of word-frequency and word-neighborhood characteristics have important implications for theories of short-term memory.

Although most of the models of serial recall described in the introduction can offer accounts for the effect of word frequency on recall, the effect of neighborhood size is clearly contrary to the prediction derived from these models, and none of them predict effects of neighborhood frequency. We discuss how some of the models may be modified to deal with these effects after the results of a fourth study are reported, below.

The pattern of results in these experiments is also clearly at odds with those observed in experiments investigating the effect of neighborhood variables on speech-perception tasks. Luce and Pisoni (1998) have shown that words from large neighborhoods and words with high-frequency neighbors are harder to recognize in noise. The present findings therefore contradict previous suggestions (e.g., Hulme et al., 1997) that lexical properties of words

Table 3
Mean Memory-Span and Speech-Rate Scores in Experiment 3

	Large neighborhood size				Small neighborhood size			
	High NF		Low NF		High NF		Low NF	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Memory span	4.82	0.65	4.72	0.72	4.62	0.69	4.27	0.52
Speech rate	3.33	0.75	3.36	0.70	3.05	0.56	3.01	0.52

Note. Speech-rate scores are shown in words per second. NF = neighborhood frequency.

influence a redintegration process that may make use of speech-perception mechanisms.

One possible alternative explanation is that these effects of lexical neighborhoods on recall reflect the role of speech-production processes in immediate memory tasks. In line with this idea, we found that increases in word-neighborhood size and neighborhood frequency were associated with higher articulation rates as well as better recall scores. However, even when articulation rate was used as a covariate, the differences in span remained. This suggests that recall differences between the word sets used cannot be accounted for purely in terms of differences in maximal articulation rate. It remains possible, however, that other processes related to speech production, such as the retrieval of speech motor programs for words that have to be articulated, do account for the recall differences associated with neighborhood size and neighborhood frequency.

It is interesting to note the similarity in the pattern of facilitative effects for neighborhood size and neighborhood frequency found in the present experiments with evidence from studies of lexical access for visually presented words. Andrews (1997) reviewed this literature and concluded that for the word-naming task all of the published experiments showed facilitative effects of both neighborhood size and neighborhood frequency. Visual word naming by definition involves speech-production mechanisms but no involvement of speech-perception mechanisms. These results are therefore consistent with the idea that variations across items in how well they are recalled in immediate memory span tasks as a function of word-neighborhood size and word-neighborhood frequency may relate to differences in speech-production mechanisms rather than speech-perception mechanisms.

There are two clear limitations to our present Experiments 1–3. In the first place, because a span procedure has been used throughout, we have no evidence about recall errors, and such errors can be very informative about the mechanisms underlying recall (e.g., Drewnowski & Murdock, 1980). Second, although we are interested in the effects of three variables on recall (word frequency, word-neighborhood size, and word-neighborhood frequency), because of restrictions in selecting matched sets of items we have only been able to assess the effects of these variables two at a time. In Experiment 4, we adopted a novel approach (see Walker & Hulme, 1999, for a previous example of the same approach) to overcome these limitations. In this experiment we used a regression-based approach, in which, across a large set of items, word frequency, neighborhood size, and neighborhood frequency are used as predictors of the probability of correctly recalling an item. This approach allowed us to assess the effects of all three variables and possible interactions between them in a single experiment. In addition, by using a fixed list-length procedure in which participants were presented with lists that are just above span on each trial, we were able to obtain evidence about the way in which each of these lexical variables affects recall errors. Such errors are potentially powerful sources of information about how word frequency and neighborhood size and frequency affect the mechanisms of recalling an item.

Another advantage of this regression-based approach was that it allowed us to discount systematic differences in the rate of articulation between groups of words as a possible source of recall differences. In Experiments 1–3, as in our earlier studies of word frequency and lexicality (e.g., Hulme et al., 1997), this issue was addressed by using the average speech rate for a set of items as a covariate. Such an approach is useful but at best gives only an approximate control for speech-rate differences. The use of a regression-based approach obviates the need for using speech rate as a covariate: Given that each item in the regression study is presented equally often in different positions and in the context of all other items in the experimental set, any differences in the probability of recalling a particular item cannot be due to differences in the rate of articulation of other items.

Finally, this procedure slightly reduces the number of times each word is encountered in the experiment. It is well known that repetition of stimuli diminishes lexical effects, such as the frequency effect, in a variety of tasks (e.g., Scarborough et al., 1977). Recently, Roodenrys and Quinlan (2000) showed that this is true also for the frequency effect in the immediate serial recall task. Their procedure involved more dense repetition of stimuli than the current experiments, but the regression procedure should minimize any effect of repetition that might be masking effects of the lexical variables.

Experiment 4

This experiment investigated the effects of word frequency, word-neighborhood size, and word-neighborhood frequency as predictors of serial-recall performance across items. A number of the models of serial recall described in the introduction that include two stages to determine order of output and the identity of the item that is output would predict that lexical factors should not be related to order errors. However, if lexical properties influence item identification, one might expect these properties to predict the occurrence of intrusions of phonologically similar words from outside of the lists (cf. Drewnowski & Murdock, 1980).

Method

Participants. Fifty-six first-year psychology students at the University of Wollongong (13 males, 43 females) participated in the study in compliance with a course requirement. All of the participants were native speakers of Australian English.

Materials. A pool of 90 CVC words was selected from the CELEX lexical database (Baayen, Piepenbrock, & Van Rijn, 1993). These words ranged in frequency from 0 to 688 per million ($M = 4137$, $SD = 104.58$), had phonological-neighborhood sizes ranging from 4 to 47 ($M = 20.97$, $SD = 10.78$), and had neighborhood frequencies ranging from 50 to 79 ($M = 63.28$, $SD = 7.66$). (See Appendix E for a list of the stimuli.) Values on the three variables were relatively uncorrelated: word frequency and neighborhood size, $r = .151$; word frequency and neighborhood frequency, $r = .014$; and neighborhood size and neighborhood frequency, $r = .171$. For each participant, the stimuli were randomly assigned to create 90 lists of 6 words such that each word occurred once in each serial position, and no word was repeated within a list.

Procedure. The study was conducted over a single session of approximately 60 min. Stimulus presentation was the same as in the previous experiments. Participants were informed that they would be presented with lists of six words that they were to repeat in the same serial order. They were instructed to say "pass" for words they could not recall. Responses were recorded for later scoring and analysis.

Results

Because each word was presented in each of the six serial positions to each participant, recall scores for each word for each participant ranged from 0 to 6. Responses were scored according to whether the presented word was recalled in the correct position. A total of 51.8% of the presented items were recalled in the correct position. In addition, a number of different types of errors were classified with respect to each presented word. A response was classified as an omission (22.2% of presented items) if the participants said “pass” or if there was no response and the word that was presented in that position was also not incorrectly recalled in another position. A response was classified as an order error if the word had been presented in the list but in a different position from the one it was recalled in. A total of 9.1% of the presented items were recalled in an incorrect position.

All other responses represent intrusions of one type or another. This could be a word from outside the list but presented in earlier lists or a word not presented in the experiment. For the sake of simplicity and clarity of exposition these errors were initially pooled together for analysis.

An initial examination of the data indicated that one item (*sung*) was very poorly recalled (more than four standard deviations below the mean) and may have been consistently misheard by some participants. This item was removed from further analysis.

To examine the relationship between each of the response categories and the variables of interest, we performed a number of simultaneous multiple regression analyses, following the recommendations of Aiken and West (1991), with each of the response categories as the dependent variable. Using a data-modeling approach to the analysis, we initially entered all of the dependent variables (word frequency, neighborhood size, and neighborhood frequency) and interactions. Nonsignificant higher order terms were then removed (three-way interactions and then two-way interactions) until only main effects or significant interactions remained. The results of these analyses are reported below. The word-frequency variable was log transformed to normalize the distribution.

The appropriate analyses for these data are ones that partition variance into between- and within-subject components. Theoretical conclusions about the effects of variables on recall depend on demonstrating effects on within-subject variance; therefore, for example, conclusions that high word frequency benefits recall depend on showing that on average, high-frequency words are recalled better when considering individual participants. The relevant analytic technique has been explained by Lorch and Myers (1990) and is followed here. The reported analyses include all of the data unless otherwise stated. Alternative analyses were performed excluding outliers on the dependent or independent variables (following the suggestions of Tabachnick & Fidell, 1996), but in most cases this had little or no impact on the pattern of the results. Only effects that were robust to the exclusion and inclusion of outliers are reported.

Correct recall. In predicting correct performance, the main effects of all three variables account for significant within-subject variance: word frequency, $r^2 = .023$, $t(4983) = 10.87$, $p < .01$; neighborhood size, $r^2 = .003$, $t(4983) = 4.24$, $p < .01$; and neighborhood frequency, $r^2 = .001$, $t(4983) = -2.40$, $p < .05$. These results indicate that words are more likely to be correctly recalled if they are higher in frequency, if they have larger neighborhoods, and if they have lower frequency neighbors.

A parallel analysis examined the contribution of phonotactic properties to the prediction of recall. Gathercole, Frankish, Pickering, and Peaker (1999) showed effects of biphone frequency on the serial recall of nonwords. However Roodenrys and Hinton (2002) argued that biphone frequency was confounded with neighborhood size and showed that neighborhood size was the important variable. The measure of biphone frequency used in this research is simply the number of three-phoneme words with that particular pair of phonemes in that position, totaled for the two biphones in the CVC word. The total biphone frequency of the words in this study correlates with neighborhood size at .91. When biphone frequency is entered into the regression prior to the neighborhood variables it has a significant effect, $r^2 = .005$, $t(4983) = 5.05$, $p < .01$. The significance of word frequency and neighborhood frequency are unaffected; however, neighborhood size is only marginally significant, $r^2 = .001$, $t(4983) = 1.78$, $p = .075$. When biphone frequency is entered after the other variables, it is not a significant predictor, $t(4983) = 0.09$, *ns*. These analyses indicate that it is difficult to separate the effects of biphone frequency and neighborhood size given their very high correlation (.91) in the sample of words used in this experiment. However, in line with the findings of Roodenrys and Hinton (2002) for nonwords, the regression analyses suggest that neighborhood size is a slightly better predictor of recall than is biphone frequency.

Omission errors. Omission errors occur when an item presented is not recalled at any position in the list. In assessing the predictors of omission errors (and the other error types that follow) correct performance was entered as a predictor on the first step, prior to the simultaneous entry of the predictor variables. After controlling for significant effects of correct performance, $r^2 = .328$, $t(4983) = -49.99$, $p < .01$, we found that as neighborhood size increases, the likelihood of a word being omitted decreases, $r^2 = .001$, $t(4983) = -2.54$, $p < .05$.

Intrusion errors. After controlling for correct performance, $r^2 = .325$, $t(4893) = -49.25$, $p < .01$, only the neighborhood frequency of an item significantly predicted the rate of intrusions suffered by an item, $r^2 = .001$, $t(4893) = 2.39$, $p < .05$. Thus the higher the frequency of its neighborhood the more likely a word is to be intruded on, suggesting that words may suffer intrusions from higher frequency neighbors. To investigate this further, we performed a regression using the number of neighbor intrusions as the dependent variable. An examination of the recall protocols shows that 3.2% of all the occasions on which an item was presented resulted in a neighbor intrusion. In this analysis, after controlling for correct performance ($r^2 = .068$, $t[4983] = -19.31$, $p < .01$), word frequency, neighborhood size, and neighborhood frequency were all significant predictors of neighbor intrusions, $r^2 = .006$, .006, and .007, and $t(4983) = -5.59$, 6.07, and 5.91, respectively, all ps $< .01$. These results show that higher frequency words are less likely to suffer an intrusion from one of their neighbors, whereas words from larger neighborhoods and higher frequency neighborhoods are more likely to be intruded upon by a neighbor. Of all of the neighbor intrusions, in 76% of cases the intruded word was higher in frequency than the word it replaced.

On a smaller proportion of occasions (approximately 2% of all items) a word is recalled by the participant that appears to be a neighbor of a word presented in a different position in the list, which may or may not have been recalled itself. These occasions may represent an order error that is compounded by a neighbor intrusion. If this is the case, one would expect the neighborhood characteristics of the neighbor presented in the list to relate to the likelihood of this occurring in a similar manner as they do to the likelihood of making a neighbor intrusion in the same position. After controlling for correct performance, $r^2 = .031$, $t(4983) = -12.80$, $p < .01$, an analysis of these order plus neighbor errors showed significant effects of word frequency, $r^2 = 0.001$, $t(4983) = -2.50$, $p < .05$; neighborhood size, $r^2 = .005$, $t(4983) = 4.91$, $p < .01$; and neighborhood frequency, $r^2 = .006$, $t(4983) = 5.42$, $p < .01$. That is, neighbors of a word presented in the list are more likely to incorrectly intrude in the list if the presented word is lower in frequency, has more neighbors, and comes from a higher frequency neighborhood.

Order errors. On approximately 9% of occasions, a presented item was recalled in the wrong position. In predicting the likelihood of such an order error the likelihood of correct recall was a significant predictor, $r^2 = .050$, $t(4983) = -16.12$, $p < .01$; as were the effects of word frequency and neighborhood size, $r^2 = .004$, $t(4983) = 4.43$, $p < .01$; and, $r^2 = .001$, $t(4983) = 2.01$, $p < .05$, respectively. These results show that a word is more likely to be recalled in an incorrect position, given that it was not recalled in the correct position, if the word is higher in frequency and has more neighbors.

Discussion

The results from Experiment 4 generally confirm the pattern obtained in the first three experiments. Overall we found that high-frequency words, and words from larger neighborhoods, are better recalled (Experiments 1–4). However, whereas Experiments 2 and 3 found that words from high-frequency neighborhoods were better recalled, Experiment 4 revealed a small effect in the opposite direction. It appears, therefore, that although we have found robust effects of word frequency and neighborhood size on recall, the effects of neighborhood frequency are small and inconsistent across experiments.

One particularly interesting aspect of the findings from Experiment 4 is that the types of recall errors made were influenced by the frequency and neighborhood characteristics of the presented word. Words are less likely to be completely omitted from a list if they have more neighbors. Words are also more likely to be intruded upon by a neighbor if they are of low frequency, have many neighbors, and if the average frequency of the neighbors is high. In the majority of cases, when a neighborhood intrusion occurs the intruding neighbor is higher in frequency than the presented word.

General Discussion

We consider the implications of these findings for our understanding of immediate serial recall tasks, and more specifically, for the possible role of language processing mechanisms in these tasks. We consider each of our major findings in turn.

Word Frequency

In three experiments we found a recall advantage for high-frequency words. This pattern replicates earlier findings (Hulme et al., 1997; Tehan & Humphreys, 1988; Watkins, 1977) and demonstrates that the effects of word frequency on recall cannot be explained in terms of neighborhood properties, such as neighborhood size and neighborhood frequency, which previous studies have not considered. The traditional explanation of word-frequency effects has been in terms of a redintegration process that operates more effectively for high-frequency words (Brown et al., 2000; Hulme et al., 1997; Nairne, 1990; Schweickert, 1993). The present findings are consistent with this interpretation of frequency effects but do nothing to constrain or test theories of the mechanisms of frequency effects.

A recent reinterpretation of the word-frequency effect as a by-product of interword association values (Stuart & Hulme, 2000) is equally consistent with the present findings. Stuart and Hulme (2000) argued that the availability of item representations to support redintegration was dependent on interitem associative links between representations of list items in long-term (lexical) memory. In this view the superior recall of high-frequency words may reflect stronger preexisting associative links between those items in lexical memory. It is suggested that such links serve to support retrieval of those items in immediate serial recall tasks. We argue below that a similar view may be extended and applied to the explanation of the effects of word-neighborhood size.

Word-Neighborhood Size

The effect of neighborhood size on recall takes the opposite form to that predicted: Words from large neighborhoods are better recalled than those from small neighborhoods. This pattern has some important theoretical implications. In models of short-term recall that postulate a redintegration process, it has often been suggested that redintegration may operate by making use of speech-perception mechanisms (Hulme et et al., 1991, 1997; Schweickert, 1993). In this view the process of redintegrating a degraded memory trace has been likened to the process of trying to recognize a spoken word that is degraded. A straightforward prediction from this view is that words that are harder to recognize in speech should be harder to recall. There is evidence that spoken words from large neighborhoods are less easy to recognize than words from small neighborhoods when presented in noise (Luce & Pisoni, 1998). The present experiments therefore appear to refute the idea that recall depends on a redintegration process that makes use of speech-perception mechanisms.

Martin and her colleagues (e.g., Martin & Lesch, 1996; Martin, Lesch, & Bartha, 1999) have proposed conceptual accounts of short-term memory based on speech perception and production processes that use multilevel activation-based models. Martin et al. (1999) proposed a model in which there are separate input and output phonological buffers linked to phonological input and output units in long-term memory that comprise layers of an interactive activation model. If the redintegration process in serial recall was seen as using degraded information in the input phonological buffer to activate the network in long-term memory, it could produce facilitative effects of neighborhood size. However, because this buffer is also part of the speech-perception apparatus this would lead one to expect speech-perception tasks to show facilitative neighborhood effects, which they do not (Luce & Pisoni, 1998). This suggests that for this model to account for the facilitative effect of neighborhood size on serial recall, the locus of the effect must be on the production side of the model. This may arise from the feedback of activation from the output buffer to the phonological output units in long-term memory.

It is interesting to note at this point that similar concepts were invoked by Goldinger, Azuma, Abramson, and Jain (1997) to account for effects they observed in a delayed-naming task. In Goldinger et al.'s study, participants had to withhold the naming response on some trials, and it was found that this had a greater impact on naming low-frequency than high-frequency words. It was suggested that maintaining the word in the naming task requires a stable representation in the production system, which they termed a *resonance*. In addition, they suggested that maintaining the resonance for a low-frequency word is more attention demanding than for a high-frequency word. The serial recall task could be seen as an extreme form of the delayed-naming task, in which spoken words must be maintained for several seconds. This strengthens the suggestion that these effects arise from speech-production mechanisms used in the serial recall task.

Given that words with more neighbors are more often recalled correctly as well as being more likely to suffer from a neighbor intrusion, one interpretation is that a word's neighbors increase the likelihood that a word will be produced from the redintegration process. Generally, this is the correct word, suggesting that the neighbors in some way directly support redintegration of the target word rather than simply competing with it. This is inconsistent with the current conceptualization of the redintegration process as it is instantiated in the theories discussed in the introduction. The problem for all of these models is to specify the sort of process that would produce a facilitative effect of neighbors on recall.

However, this pattern of results is broadly in line with the reinterpretation of word-frequency effects offered by Stuart and Hulme (2000). If we assume that words in a phonological neighborhood are associatively linked in lexical memory, our hypothesis would be that such groups of associated words will all be activated to some extent by the presentation of one word from the neighborhood. A further assumption is that members of a neighborhood form a mutually supportive network of items. In this way words from large neighborhoods will receive supportive activation from more other words at recall than words from small neighborhoods. Such a mechanism might also be expected to generate significant numbers of neighbor intrusions in recall, and this was the pattern we observed in Experiment 4.

A problem for any proposed mechanism of this sort is how, in a mutually supportive network, competition between simultaneously active items is resolved. One way in which this issue has been solved computationally is in models that incorporate self-organizing maps, or Kohonen networks (see, e.g., Kohonen, 1995). In this type of model the units in the input layer are connected to all of the units in a second, output layer. Units in the output layer also receive input from other units in the output layer. Each unit in the output layer computes the similarity between its weights and the input pattern and is activated accordingly. The unit that is most active is selected for output. Most interesting for our purposes, the output-layer units have excitatory connections to nearby units and inhibitory connections to more distant units. As stimuli are repeatedly presented in the training phase, the output layer comes to be topologically organized, reflecting the distribution of the items in the training set. Usually the size of the neighborhood (sharing excitatory connections) is reduced during training to accomplish this (Kohonen, 1995). If neighborhood size is based on distance, then units in dense neighborhoods will receive more activation when their corresponding input pattern is presented than units in sparse neighborhoods will receive. It is important to note that these are winner-take-all models, therefore only one item is output, which means that having more neighbors will increase the likelihood of the correct unit being output but will also increase the likelihood of a neighboring unit being output. Such a process might fruitfully be incorporated into some of the models cited in the introduction.

Word-Neighborhood Frequency

As we noted earlier the small negative effect of neighborhood frequency on correct recall in Experiment 4 contrasts with the small facilitative effects observed in Experiments 2 and 3. This difference may well be due to differences in task demands between experiments. It appears that the finding that correct performance is negatively related to neighborhood frequency in Experiment 4 is due to the likelihood of making a neighbor intrusion. Analyses showed that words from higher frequency neighborhoods are more likely to suffer a neighbor intrusion than are those from lower frequency neighborhoods. The opportunity for neighbor intrusions to occur is much more limited in the experiments measuring memory span, because the methodology means that approximately half of all lists are recalled completely correctly. In addition, the number of errors on the other half of the lists is fairly small because the task is, almost by definition, only just beyond the capacity of the participant. This means the percentage of items on which a participant makes an error is actually very small, whereas in Experiment 4 it is almost 50%. The limited opportunity for neighborhood intrusions in the memory span experiments, compared with the regression study in Experiment 4, appears to offer the most natural explanation for why neighborhood frequency had a different effect on recall across experiments.

Patterns of Recall Errors

One interesting aspect of the results of Experiment 4 was the pattern of recall errors obtained. Such errors potentially provide powerful constraints on models of task performance. It was found that words are less likely to be completely omitted from a list if they have more neighbors. This finding appears consistent with the idea outlined above, that words with many neighbors benefit from mutually supportive links with these neighbors at the point of retrieval. Words were more likely to be intruded upon by a neighbor if they are of low frequency, have many neighbors, and if the average frequency of the neighbors is high. This pattern suggests that there is some trade-off at retrieval between the strength of activation present for the target word and the strength of activation present for neighboring words. It appears that words of low-frequency will on some occasions suffer competition from higher frequency neighbors and be substituted by them as responses. In keeping with this, in the majority of cases, when a neighborhood intrusion occurs the intruding neighbor is higher in frequency than the presented word.

Finally, it was also observed that word frequency and neighborhood size had effects on the likelihood of making an order error. Specifically, high-frequency words and words from larger neighborhoods are more likely to be recalled in the wrong position in a list. This finding seems to go against the spirit of a number of models of short-term memory mentioned in the introduction but on closer examination may be consistent with at least some of them. A number of models stress the existence of separate mechanisms for maintaining item identity and order of presentation. For example, the primacy model of serial recall (Henson, Norris, Page, & Baddeley, 1996; Page & Norris, 1998) includes a mechanism by which nodes corresponding to items in a list are activated to varying degrees according to their position in the list. The rank order of the most active node through the least active corresponds to the first through the last item. This gradient of activation, combined with a suppression mechanism, determines the order in which nodes are selected for output. When a node is selected for output the second stage determines the phonological form that is actually produced. In such a model, if the initial stage involved no phonological information one expects order errors to be unrelated to the lexical variables. However, this is not true of most of these models. Rather, phonological information is represented in both processes, and the models vary in their potential for lexical knowledge to influence the initial item selection process during recall. For example, in the Burgess and Hitch (1999) model, items become associated with a context signal that changes in a systematic manner. Essentially, a replaying of this changing context signal is used to cue retrieval of the appropriate item at the appropriate time by differentially activating the items. The nodes representing items are linked not only to nodes in the context layer but also to layers of input and output phonemes. Strong activation from long-term phonological knowledge, as might be expected from high-frequency words, when combined with partial activation from the context layer may allow these items to exceed the activation of the correct item and jump the queue in recall. The results of Experiment 4 clearly suggest that such an influence does exist.

Conclusion

To summarize, the experiments reported here show that words of high frequency and words from larger neighborhoods are better recalled in immediate serial recall tasks. Errors in recall often involve substituting a word's neighbor for the word presented. Such neighbor substitutions occur more often for low-frequency words and for words from larger and higher frequency neighborhoods. These results show that (a) the neighbors of a word play an active role in its recall and (b) the immediate serial recall of word lists is critically dependent on the nature of representations of those items in long-term (lexical) memory. However, the fact that words from large neighborhoods are better recalled, whereas previous studies have shown that such words are harder to recognize in speech, indicates that the effects of these lexical properties of items on recall cannot be mediated by speech-perception mechanisms. A facilitative effect of neighborhood size has also been observed in memory for nonwords (Roodenrys & Hinton, 2002) suggesting the effect may arise in a process that is common to the recall of both words and nonwords. We argue that an account of recall that places the locus of these neighborhood effects in speech-production processes offers a useful means of understanding these effects.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Thousand Oaks, CA: Sage.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4,* 439–461.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database.* Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Baddeley, A. D. (1986). *Working memory* (Vol. 11). Oxford, England: Oxford University Press.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158–173.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

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.

Brown, G. D. A., & Hulme, C. (1995). Modelling item length effects in memory span: No rehearsal needed? *Journal of Memory and Language, 34,* 594–621.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107,* 127–181.

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.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention & performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Coltheart, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on short-term-memory recall of repeated and novel word lists. *Memory & Cognition, 21,* 539–545.

Conrad, R. (1963). Acoustic confusion and memory span for words. *Nature, 197,* 1029–1030.

Drewnowski, A., & Murdock, B. B. (1980). The role of auditory features in memory span for words. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 319–332.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23,* 83–94.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 84–95.

Goldinger, S. D., Azuma, T., Abramson, M., & Jain, P. (1997). Open wide and say "blah!" Attentional dynamics of delayed naming. *Journal of Memory and Language, 37,* 190–216.

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 152–162.

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language, 59,* 267–333.

Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49*(A), 80–115.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30,* 685–701.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal* *of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1217–1232.

Kohonen, T. (1995). *Self-organizing maps.* Heidelberg, Germany: Springer-Verlag.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 149–157.

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

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271.

Martin, R. C., & Lesch, M. F. (1996). Associations and dissociations between language impairment and list recall: Implications for models of STM. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 149–178). Hove, England: Psychology 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.

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

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

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17,* 398–422.

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48*(A), 384–404.

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., Alban, J., Ellis, A., & Brown, G. D. A. (1994). Effects of word frequency and age of acquisition on short-term memory span. *Memory & Cognition, 22,* 695–701.

Roodenrys, S., & Quinlan, P. T. (2000). The effects of stimulus set size and word frequency on verbal serial recall. *Memory, 9,* 71–78.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 1–17.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21,* 168–175.

Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 419–425.

Sperling, G., & Speelman, R. G. (1970). Acoustic similarity and auditory short-term memory: Experiments and a model. In D. A. Norman (Ed.), *Models of human memory* (pp. 151–202). New York: Academic Press.

Standing, L., Bond, B., Smith, P., & Isely, C. (1980). Is the immediate memory span determined by sub-vocalisation rate? *British Journal of Psychology, 71,* 525–539.

Stuart, G., & Hulme, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 796–802.

Tabachnick, B. G., & Fidell, L. S. (1996) *Using multivariate statistics* (3rd ed.). New York: HarperCollins.

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.

Vitevitch, M. S., & Luce, P. A. (1999) Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40,* 374–408.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1256–1271.

Watkins, M. (1977). The intricacy of memory span. *Memory & Cognition, 5,* 529–534.

Wetherick, N. E. (1975). The role of semantic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 471–480.

Appendix A

Calculating Mean Neighborhood Frequency

The mean neighborhood log-frequency is based on a natural log-transform of the frequency of occurrence per million words according to the count of Kučera and Francis (1967). The value for this variable was calculated by transforming the Kučera and Francis frequency of each item in the neighborhood using the following formula: *Log-Freq.* = $(\log_e freq + 10) \times 20$, rounding off to the nearest integer, and then averaging across the words in the neighborhood. The purpose of the log transformation in this formula was to reduce the effect of outliers on the mean neighborhood frequency. It results in a distribution of values such that a raw frequency of 0 becomes 46, and a raw frequency of 100 becomes 94. The purpose of the multiplication in this formula was simply to produce a more user friendly scale. The calculations of neighborhood frequency used in Experiment 4 are based on the CELEX frequency count (Baayen et al., 1993) rather than the Kučera and Francis frequency count.

Appendix B

Stimuli Used in Experiment 1

High word frequency				Low word frequency			
Large NS		Small NS		Large NS		Small NS	
boat		charge		bean		chive	
book		death		bike		fudge	
face		faith		cart		garb	
game		food		cheat		geese	
lead		guess		fad		gong	
heart		house		fin		hedge	
hope		jazz		gash		keg	
job		judge		hop		mesh	
kill		leg		lard		moth	
moon		love		mole		poise	
park		north		rhyme		pouch	
rain		safe		rib		sour	
road		sharp		sage		surf	
seat		song		shale		torch	
tone		south		weed		vase	
wife		voice		whim		verb	
M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Word frequency							
145.1	87.4	163.4	137.8	2.3	1.6	2.4	1.6
Neighborhood frequency							
65.1	5.1	66.5	6.1	64.0	4.9	63.9	9.0
Neighborhood size							
29.5	8.3	9.8	3.1	28.1	6.9	7.9	2.4

Note. NS = neighborhood size.

(*Appendixes follow*)

Appendix C

Stimuli Used in Experiment 2

High word frequency				Low word frequency			
High NF		Low NF		High NF		Low NF	
dead		chief		chat		barb	
foot		church		geese		booth	
force		dark		gnome		cough	
game		death		hem		fetch	
gun		dog		hose		gouge	
house		fig		pet		jug	
jazz		gas		sieve		lash	
leg		girl		soot		merge	
mouth		job		tame		morgue	
nose		neck		thong		moth	
shot		north		torch		nap	
wife		peace		void		pouch	
wish		rock		web		puff	
wood		room		wool		tag	
youth		vote		yacht		yolk	
M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Word frequency							
147.5	134.6	176.7	103.4	4.7	3.3	3.9	3.0
Neighborhood frequency							
76.4	4.6	55.4	3.2	79.5	6.9	53.3	2.4
Neighborhood size							
16.5	7.5	16.2	5.9	14.7	7.2	14.5	5.6

Note. NF = neighborhood frequency.

Appendix D

Stimuli Used in Experiment 3

Large neighborhood				Small neighborhood			
High NF		Low NF		High NF		Low NF	
bait		boil		bush		chef	
bug		dome		chive		chirp	
cat		dice		geese		couch	
cork		jail		hedge		cure	
fan		lamb		mesh		fudge	
goat		lip		nerve		garb	
hip		mop		nudge		gnarl	
leach		peach		queer		gouge	
mouse		gag		thief		keg	
rope		nap		thong		merge	
rude		rib		tooth		morgue	
shark		rug		torch		moth	
thorn		sock		turf		poise	
tomb		tub		vogue		vague	
weed		veal		void		verb	
yacht		yawn		web		yarn	
M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Word frequency							
7.6	6.7	8.3	7.0	6.3	5.3	7.3	8.8
Neighborhood frequency							
67.2	4.2	56.7	2.7	69.2	6.5	55.8	3.3
Neighborhood size							
26.1	5.8	24.3	4.0	6.7	2.6	7.4	2.0

Note. NF = neighborhood frequency.

(*Appendix follows*)

Appendix E

Stimuli Used in Experiment 4

beef	fierce	jeep	nudge	sieve
beige	firm	jet	pain	soup
bile	fizz	kin	park	sung
born	food	knife	phone	tap
bush	garb	large	poach	tease
cape	gash	limb	poise	thumb
chafe	gauze	lurch	pool	tooth
cheat	girl	maim	psalm	tough
chirp	gout	men	pup	van
choice	halve	moose	quiche	verb
coil	hat	mope	rate	void
cork	hearse	morgue	reek	vote
darn	hid	moth	rhyme	wart
dice	hike	nag	rouge	wed
doll	home	nerve	rub	whale
dove	hutch	nook	shirt	wick
fall	jade	notch	shot	yolk
fetch	jam	noun	shove	zip

Note. Word frequency, $M = 41.4$, $SD = 104.6$; Neighborhood size, $M = 21.0$, $SD = 10.8$; Neighborhood frequency, $M = 63.3$, $SD = 7.7$.

Received April 20, 2001
Revision received April 30, 2002
Accepted April 30, 2002 ■