RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 22 (1998)
Indiana University

Effects of Lexical Neighborhoods on Immediate Memory Span for Spoken Words: A First Report[1]

Winston D. Goh[2] and David B. Pisoni

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

---

[1] This research was supported by NIH-NIDCD Research Grant DC00011 to Indiana University. We would like to thank Luis Hernandez and Michael Vitevitch for help in programming the experiment.

[2] Also Department of Social Work & Psychology, National University of Singapore.

Effects of Lexical Neighborhoods on Immediate Memory Span for Spoken Words: A First Report

Abstract. Current theories and models of the structural organisation of short-term memory are primarily based on evidence obtained from manipulations of inherent features in the short-term traces of the presented stimuli, such as phonological similarity. In this study, we investigate whether aspects of the stimuli that are not inherent in the short-term traces, such as phonological neighborhood properties based on the mental lexicon's structure and organisation of words in long-term memory, would affect performance in an immediate memory span task. Words differing in phonological neighourhood density, phonological neighborhood frequency, and word frequency were manipulated. Memory span for lists of these words, using a repeated sampling and non-repeated sampling procedure, was measured. The results demonstrate that lexical neighborhood effects emerge only when a non-repeated sampling procedure is used. The results further indicate that these lexical effects are not dependent on short-term memory capacity as measured by the traditional digit-span task. Implications for the integration of long-term memory structure in short-term memory processes are discussed.

One of the main assumptions in short-term memory (STM) models is the proposal that verbal material is coded or represented in a phonological form. For example, in Baddeley and Hitch's (1974) well-known "working" memory model, there is a specific mechanism called the "phonological (or articulatory) loop" that handles verbal material. The impetus for positing specific mechanisms to handle phonological coding of words stems from a number of fairly robust findings in the STM literature.

First, immediate memory span for a list of phonologically similar tokens is poorer than the span for a list of phonologically dissimilar tokens across a wide variety of experimental conditions (e.g., Baddeley, 1966a; Conrad, 1964). This is called the *phonological similarity effect*. There is no corresponding effect in the immediate memory span for lists of *semantically* similar or dissimilar tokens. Semantic similarity, however, does impair performance in tasks which require the long term retention and learning of words, but acoustic similarity does not (Baddeley, 1966b). This cross-over dissociation between phonological similarity and semantic similarity in STM and long-term memory (LTM) tasks gave rise to the simple generalisation that the primary code for STM is phonological, whereas the primary code for LTM is semantic in nature (Baddeley, 1998, pp. 41-42).

Second, Baddeley, Thomson, and Buchanan (1975) demonstrated that lower spans are obtained with lists comprising words of long spoken duration (e.g., *Friday, harpoon*) than with lists comprising words of short spoken duration (e.g., *wicket*, *bishop*). This is called the *word length effect*. Word length effects are attributed to the rate at which items can be rehearsed by the articulatory loop (Baddeley, 1986; 1998). Longer words have more underlying gestures, take more time to rehearse, and therefore fewer items can be maintained in STM. Supporting evidence comes from studies that show an inverse relationship between digit-span and the articulation rate of the language used. Digit-spans tend to become smaller as one goes from Chinese, to English, to Malay, and to Welsh (Ellis & Hennelly, 1980; Naveh-Benjamin & Ayres, 1986; Hoosain & Salili, 1988; Elliott, 1992).

Third, irrelevant speech introduced during the visual presentation of digits disrupts digit-span performance (Salame & Baddeley, 1982; 1987; 1989). This is called the *unattended speech effect*. Recall is equally disrupted by meaningful words, nonsense syllables, spoken digits, and words which contain some of the phonemes of the digits (e.g., *tun*, *woo* instead of *one*, *two*). These results suggest that the unattended speech effect is caused by the phonological disruption of the STM traces, and not the semantic content; otherwise meaningful words and spoken digits should be more disruptive than nonsense syllables.

Fourth, the word length and phonological similarity effects can be abolished under certain conditions involving articulatory suppression. In these studies, participants are required to articulate "the" or some other word over and over again throughout the presentation of the stimuli. The purpose of this procedure is to prevent the use of phonological coding or articulatory rehearsal. Word length effects are generally negated by articulatory suppression (Baddeley et al., 1975). The phonological similarity effect is abolished only if the list presentation is done visually. It has been argued that participants need to phonologically recode the visually presented items, and articulatory suppression prevents this from happening (Baddeley, Logie, Nimmio-Smith, & Brereton, 1985). When the list is presented auditorily, articulatory suppression has no effect because the stimuli have direct access to the phonological ST store and do not need to be recoded into a phonological form.

The interpretations given to these findings have not gone unchallenged (see Cowan, Wood, Nugent, & Treisman, 1997; Nairne, Neath, & Serra, 1997, for alternative accounts of the word length effect; and Bavelier & Potter, 1992, for articulatory suppression and phonological similarity). However, instead of joining these debates, we would like to focus our discussion on a common feature that occurs in all these studies. This feature is the use of structural properties that are inherent in the presented stimuli (e.g., that one list of tokens share common phonological features while the other list does not) to provide insights about the underlying mechanisms and processes of the human memory system. This prevailing approach may not reveal the contribution of factors that are beyond the physical properties of the experimental stimuli. These factors include the organisational structure of the lexicon as a whole and the statistical distribution of the properties that define each individual word with respect to the entire lexicon. For example, word frequency does not exist as an inherent property of the set of experimental stimuli, but is a function of long-term acquisition and experience of a participant. Yet, word frequency has been demonstrated to influence immediate memory span tasks. The span for high frequency words is consistently longer than the span for low frequency words (e.g., Engle, Nations, & Cantor, 1990). Word frequency and the organisation of the lexicon have generally been considered to fall within the domain of the more permanent long-term memory (LTM) system, and are therefore not given as much consideration when describing STM processes.

To a large degree, the prevalent approach reflects the view that STM is functionally and structurally separable from LTM (see Richardson, 1996, pp. 120-122, for a succinct review). There is evidence from the clinical literature for this view of multiple memory systems (e.g., Scoville & Milner, 1957; Martin, 1993). However, the extent of this dissociation remains a major point of contention in the literature (see Baddeley, 1998, pp. 37-44). At some level, the permanent information in LTM must come into play in memory processes. Determining which aspects of the permanent store will influence STM or "working" memory processes, and to what degree they do so, will facilitate the development of more integrated accounts of the mechanisms underlying the human memory system.

Lexical Neighborhoods

Since the prevalent view of STM encoding is that items are translated into a phonological code, it is important to look at how the phonological properties of words are stored and organized in LTM. One way to represent this information is to consider the notion of lexical neighborhoods defined by phonological similarity (Landauer & Streeter, 1973; Luce, 1986; Luce & Pisoni, 1998). The number of words that can be obtained by substituting one phoneme position in the target word or by the single addition or deletion of a phoneme defines a similarity neighborhood. For example, the neighbors of *cat* would include *hat*, *cut*, and *cap* by phoneme substitution, and *scat* and *at* by phoneme addition/deletion.

Two factors characterise the structure of a neighborhood. *Neighborhood density* refers to the number of words in a similarity neighborhood. Dense neighborhoods contain many similar sounding words while sparse neighbourhoods contain few similar sounding words. *Neighborhood frequency* refers to the average frequency of occurrence of words in a similarity neighborhood. High frequency neighborhoods contain primarily high frequency words while low frequency neighborhoods contain primarily low frequency words. *Word frequency* is an individual word's frequency of occurrence in the language (based on text count such as Kucera and Francis (1967)) and should not be confused with neighborhood frequency, which is the average frequency of the word's neighbors.

Based on these three properties, words can be classified into those that would be "easy" and "hard" to recognize. "Easy" words are those that "stand out" because they are higher in frequency relative to their neighbors, and they also have few neighbors. They reside in low density and low frequency neighborhoods. In contrast, "hard" words tend to get "swamped" by their neighbors because of their lower relative frequency and the crowded neighborhood. They reside in high density and high frequency neighborhoods.

Lexical neighborhood properties have been shown to influence word recognition, including accuracy in perceptual identification (Landauer & Streeter, 1973; Luce, 1986), naming and lexical decision latency (Luce, Pisoni, & Goldinger, 1990), priming (Goldinger, Luce, & Pisoni, 1989), and reduction in speech (Wright, 1997). The overall pattern of results indicates that words are recognized in the context of other words in the lexicon, i.e., words compete for recognition (Luce & Pisoni, 1998; Vitevitch & Luce, 1998). Luce and Pisoni's (1998) Neighborhood Activation Model (NAM) of spoken word recognition evolved from this series of findings. The basic assumptions of the model include the proposition that words in the mental lexicon are organized in terms of the similarity neighborhoods we described earlier, and that the LTM representations of words will contain many overlapping features if they are from the same neighborhood. The word recognition system is assumed to involve a competitive discrimination process among the individual lexical items in LTM. According to NAM, when the stimulus is parsed during word recognition, all representations with similar features get activated. Through a process of competition, one token wins and is recognized. The speed and accuracy of recognition is directly dependent on the lexical neighborhood properties of the target word.

Would lexical neighborhoods have an effect on STM processes? One might ask why neighborhoods should continue to exert an influence once the word has been recognized. In the typical immediate memory span tasks used in STM studies, the words are recognized long before the participant is required to actually recall the list of items. Since the words are already recognized, why should their neighbors continue to affect anything?

If we can demonstrate that lexical neighborhood properties do influence immediate memory span, then two things become clear in terms of theoretical implications. First, neighborhood effects persist beyond word recognition. Second, the processes involved in immediate memory span tasks are also influenced by phonological information in LTM and not just the ST phonological traces of the tokens. Note the difference between how words are picked in phonological similarity studies and how they are picked in lexical neighborhood studies. Phonologically similar words are essentially picked from one lexical neighborhood, and the list making up dissimilar words are picked from multiple neighborhoods. "Easy" and "hard" lists in lexical neighborhood experiments are *both* picked from multiple neighborhoods. These lists are thus equivalent to the dissimilar word lists in phonological similarity experiments. Therefore, if there is any disparity in performance on "easy" and "hard" words, it is not because of ST phonological traces, but because of the contribution of LT phonological neighborhood properties of the specific items used in these lists. In the present study, we manipulated lexical neighborhood properties by using lists made up of only “easy” or “hard” words. We predicted that the memory span for “easy” words would be longer than the span for “hard” words.

Repeated and Non-repeated Sampling

Many of the memory span studies described above used a repeated sampling procedure – i.e., there was a fixed number of stimulus tokens which were repeated from trial to trial or list to list. Drewnowski and Murdock (1980) noted that this procedure might turn the experimental stimuli into a very restricted set. This may induce participants to encode or retrieve the tokens based on partial information of each token and their knowledge of the restricted set. If participants know that the same five tokens are going to be on the lists over and over again, it might be sufficient to simply encode or maintain features that distinguish one item from the other, since these features can subsequently be retrieved and mapped onto the appropriate token. Repeating tokens from trial to trial may therefore lead to a more limited and idiosyncratic coding in memory and fail to reflect any influence from LTM. Token repetition may also increase proactive interference (PI) from previous lists, since the same tokens are used over and over except for the order. This is similar is similar to the situation involving digit-span tests of STM as well, although the list of 10 digits from 0 to 9 form a naturally closed set.

If the tokens are unique from trial to trial, then each token is only activated once, and limited coding or PI will not obscure the influence of its LTM representation. Nairne, Neath and Serra (1997) argue that PI plays a role in the word length effect for repeated lists because it only emerges after many trials and not in the initial trials of an experiment. They argue that subjects are directly recovering list information from LTM without any interference in the initial trials. This supports an earlier study by La Pointe and Engle (1990) which showed that the word length effect is not abolished by articulatory suppression for non-repeated lists. By using new words for each list, the effects of PI and limited coding are minimized. It is plausible that articulatory suppression would not be expected to have an effect if the words are being retrieved directly from LTM for each new list.

Differences in performance between repeated and non-repeated sampling procedures are not consistent in the literature. Coltheart (1994) showed that the phonological similarity effect persisted in both repeated and non-repeated conditions. However, in Coltheart’s study, her pool of phonologically similar words in the non-repeated condition was essentially drawn from a single large neighborhood. So it is not surprising that the similarity of the unique tokens from one list to another caused interference. Recently, Nairne and Kelley (in press) showed that the phonological similarity effect could be reversed under certain conditions using non-repeated sampling. By using a delayed recall task and varying the retention interval, the advantage of phonologically dissimilar tokens over phonologically similar ones actually reverses at the longest retention interval of 24 seconds. Their explanation of this pattern is that the similarity of phonological features can be an effective retrieval cue for discriminating between lists while it may produce impairments for discriminating tokens within lists.

Such findings suggest that it may be very important to examine the differences in performance using repeated and non-repeated sampling techniques in any study of STM or working memory. Any LTM effects are probably going to be strongest when using non-repeated sampling since this procedure minimizes the effects of PI and the use of limited or item-specific coding strategies. The findings of Sommers, Kirk, and Pisoni (1997) support this assertion. In a perceptual identification task, Sommers et al. (1997) demonstrated that participants who used a “closed-set” response format (i.e., they had to identify the presented token from a set of six possible targets) showed no performance difference for “easy” and “hard” words. However, participants who used an “open-set” response format (i.e., they had to identify the presented token without any constraining sets) showed a performance advantage for "easy" over "hard" words. Sommers et al. (1997) argued that the closed-set format effectively eliminated the need to access LTM since the search is limited to the given set of possible responses. In this condition, the words were no longer competing with other words in the mental lexicon, and thus there was no access to the neighborhood structure of the lexicon. In the open-set response format, the search space is the entire lexicon. Not surprisingly, the structural organisation of the lexicon influenced performance in this condition.

Thus, in addition to the lexical neighborhood manipulation, we also manipulated the sampling procedure by using repeated and non-repeated tokens. We hypothesised an interaction between sampling procedure and lexical neighborhood such that there will be a difference between the "easy" and "hard" word spans only if non-repeated sampling is used. Lexical neighborhood effects are expected to be obscured in repeated sampling because this procedure effectively makes the limited pool of words a closed set and changes the nature of the task demands.

Method

Participants

Forty undergraduates from the Psychology Department at Indiana University participated in the study. All were native English speakers with no speech or hearing disorder at the time of testing.

Design

The experiment employed a 2 x 2 mixed factorial design. The Sampling independent variable (repeated vs. non-repeated) was run between subjects and the Word Type independent variable ("easy" vs. "hard") was run within subjects. Visual and auditory digit-spans were also obtained from all subjects as a base measurement of STM capacity. This was necessary to ensure that the subjects in the two sampling conditions did not differ in STM capacity as measured by the traditional digit-span.

Materials

Tokens of 132 spoken words (66 "easy" and 66 "hard") were drawn from a prerecorded digital database (see Torretta, 1995, for a detailed description). All were monosyllabic consonant-vowel-consonant words which were rated as highly familiar (> 6.7) on the 7-point Hoosier Mental Lexicon scale (Nusbaum, Pisoni, & Davis, 1984). The tokens recorded by the most intelligible male talker (M9, who had a mean intelligibility score of > 95%) at a "medium" speech rate were used (see Torretta, 1995, for details). Sixteen tokens (8 "easy" and 8 "hard") were selected from the 132-word pool for use in the repeated sampling conditions. All 132 words were used in the non-repeated sampling conditions.

Tokens of the 10 spoken digits (0 to 9) were obtained from the Texas Instruments 46-Word (TI46) Speaker-Dependent Isolated Word Corpus (Texas Instruments, 1991). The original tokens on the CD-ROM were in 11025 Hz 16-bit PCM-Motorola formatted files. These were edited using a digital waveform editor to remove the silent portions from either side of the token and saved as 12500 Hz, 16-bit, mono WAV files. The overall RMS (root-mean-square) amplitude levels for each digit token were digitally equated with the word tokens to ensure equal presentation levels. The tokens recorded by a male speaker whose mean intelligibility was 100%, as determined by a token identification task with ten volunteer participants, were used.

Apparatus

Gateway 2000 Pentium 133 MHz IBM compatible computers equipped with 15-inch Vivitron SVGA monitors and SoundBlaster AWE32 sound cards were used in the experiment. The auditory stimuli were presented to participants via a pair of Beyer Dynamic DT100 headphones at approximately 75 dB SPL.

Procedure

Participants were tested individually or in groups of five or fewer. The digit-span tests were always done first, followed by the word-span tests. The auditory-visual order of presentation for the digit-span tests was counterbalanced, as was the "easy-hard" order of presentation for the word-span tests. This resulted in a total of four different orders of testing. Participants were randomly assigned to one of these orders.

The procedure followed the Wechsler Adult Intelligence Scale method of increasing the list length by one after every two trials. For example, our digit-span test starts with a list length of 4 digits. This constitutes one trial. The second trial is also a list of 4 digits. The third trial is a list of 5 digits and so on. Both visual and auditory digit-span tests began with a list length of 4 digits and ended with a length of 10, resulting in a total of 14 trials for each test. The "easy" and "hard" word-span tests began with a list length of 3 words and ended at a length of 8, resulting in 12 trials for each test.

During a trial, the computer presented each token with an inter-stimulus interval (ISI) of 500 msec. For the visual digit-span test, Arabic numerals were randomly displayed one at a time in the center of the CRT monitor for 1000 msec. For all other tests, the auditory tokens were randomly presented to the participants via a pair of headphones. Participants were told to memorise the list and that they had to reproduce all the items on the list in the correct sequential order. They recorded their responses by writing on prepared answer booklets at the end of each trial, after which they initiated the next trial by pressing a key on the keyboard.

Results

Scoring

A token was considered correct if its identity and sequence position were *both* recalled accurately. A perfectly recalled trial is one where *all* tokens were recalled in the correct order. Four different scoring procedures derived or adapted from the methods described by La Pointe and Engle (1990) were used. The Highest Span score was obtained from the longest list length in which *both* trials were recalled perfectly (Broadbent, 1971). The Strict Span score was obtained by adding 0.5 points to the Highest Span score for every trial that was *perfectly* recalled thereafter (cf. Daneman & Carpenter, 1980). The Absolute Span score was obtained by counting the total number of words recalled in every perfectly recalled trial, and the Total Span score was obtained by counting the total number of words recalled correctly regardless of whether the trial was perfectly recalled (La Pointe & Engle, 1990). Of the four scoring algorithms, the Total Span score provides the most variance, while the Highest Span score contains the least variance. Table 1 gives examples of the various span scores.

Two participants did not complete the last trial in one or more of the span tests. For these participants, only the Highest Span score could be computed. All analyses involving the three other span scores did not include data from these two participants.

Table 1

Scoring Procedures

List Length	Trial	Presented list	Response
2	1	1 2	1 2
	2	3 4	3 4
3	3	1 2 3	1 2 3
	4	2 3 4	2 3 1 *
4	5	4 2 1 3	4 2 1 3
	6	1 2 3 4	1 3 2 4 *

Highest span score: 2
Strict span score: 2 + 0.5 + 0.5 = 3
Absolute span score: 4 + 3 + 4 = 11
Total span score: 4 + 5 + 6 = 15

Note. The asterisks indicate responses that contain incorrect items and/or order

Digit-span Scores

Table 2

Intercorrelations Between Visual and Auditory Digit-span Scores

	Visual				Auditory			
	Highest	Strict	Absolute	Total	Highest	Strict	Absolute	Total
Visual								
Highest	1.000	.746**	.677**	.614**	.205	.419**	.413**	.461**
Strict		1.000	.988**	.858**	$.306^a$	.502**	.484**	.540**
Absolute			1.000	.850**	$.296^a$	.464**	.451**	.485**
Total				1.000	$.313^a$	.575**	.593**	.552**
Auditory								
Highest					1.000	.702**	.661**	.621**
Strict						1.000	.977**	.850**
Absolute							1.000	.830**
Total								1.000

** $p < .01$
$^a p < .08$

Pearson correlations between the various digit-span scores are listed in Table 2. Both visual and auditory digit-span correlate significantly on all the different scoring procedures except for auditory highest digit-span with the other visual digit-span scores. This could be due to the low variability, and hence sensitivity, of the highest span procedure. However, three out of the four statistically non-significant correlations do approach significance. The mean digit-span scores are listed in Table 3. None of the differences between the visual and auditory digit-spans for any of the scoring procedures were significant. Since the modality of the digit-span did not appear to make much difference in performance, we used the auditory digit-span as the base measure of participants' STM capacity in all subsequent analyses.

Table 3

Mean Digit-span Scores as a Function of Presentation Modality

	Span Type			
Modality	Highest	Strict	Absolute	Total
Auditory				
M	5.73	6.87	44.61	70.32
SD	1.15	1.08	14.17	9.98
Visual				
M	5.88	6.87	44.63	72.39
SD	1.04	1.00	15.39	11.25

Table 4 lists the mean digit-span scores for participants across the two sampling conditions. There were no significant differences between the two conditions, all *t*s(38) < 1. Therefore, we can be confident that the average STM capacity (as measured by digit-span) of the two groups of participants was equivalent.

Table 4

Mean Auditory Digit-span Scores as a Function of Sampling Conditions

	Span Type			
Sampling	Highest	Strict	Absolute	Total
Repeated				
M	5.60	6.70	42.20	68.95
SD	1.23	1.32	16.45	11.20
Non-repeated				
M	5.85	6.93	45.35	71.75
SD	1.09	0.82	12.38	8.12

Lexical Neighborhood and Item Repetition

Pearson correlations between the various digit-span scores and the "easy" and "hard" word-span scores were not as consistent as the correlations observed between the visual and auditory digit-span scores. This was true in both the repeated and non-repeated sampling groups. Differences in the sensitivities of the various scoring procedures may account for this. However, an examination of the correlations for the Total Span scores (since these have the most variation and might be the most sensitive) listed in Table 5 revealed the following:

The "easy" and "hard" total word-span scores are significantly correlated with the total digit-span scores. This is not surprising – a larger STM capacity (as measured by digit-span) can hold more words and is therefore expected to reflect larger word-span scores. Interestingly, the "easy" and "hard" word-span scores do not significantly correlate with each other. This suggests that whatever is happening with "easy" and "hard" word-spans may be more dependent on factors other than pure memory capacity for individual items.

Table 5

Total Span Correlations

Span	Digit	"Easy" word	"Hard" word
	Repeated sampling		
Digit	1.00	.529*	.463*
"Easy" word		1.00	.150
"Hard" word			1.00
	Non-repeated sampling		
Digit	1.00	.523*	.442[a]
"Easy" word		1.00	.352
"Hard" word			1.00

* $p < .05$
[a] $p < .06$

The pattern of results using the mean span scores is identical for all four scoring procedures, so we will only report the analyses of the Total Span scores here. Figure 1 shows the interaction between the Sampling and Word Type conditions. This interaction was significant in a two-way mixed-design analysis of variance (ANOVA), $F(1, 36) = 10.43$, $MS_e = 25.80$, $p < .01$. Tests of simple effects indicated that the source of the interaction was the difference between the "easy" and "hard" word-spans in the non-repeated sampling condition, $F(1, 36) = 19.43$, $MS_e = 25.80$, $p < .001$, and the difference between repeated and non-repeated sampling performance on "hard" word-spans, $F(1,36) = 6.80$, $MS_e = 43.49$, $p < .05$. All other simple effects were not significant, all *F*s < 1.4.

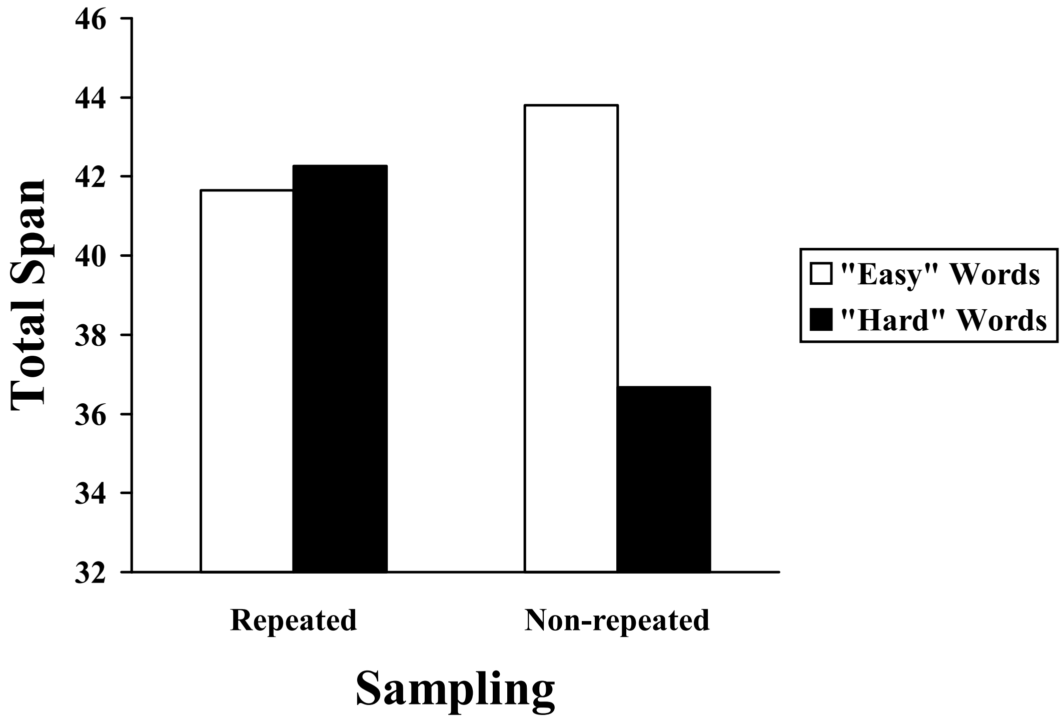


Figure 1. Total span score as a function of Sampling and Word Type.

Given that the simple correlations between the various digit-span scores and the various “easy” and “hard” word-span scores were inconsistent, we decided to check if the magnitude of the *difference* between “easy” and “hard” word-span scores for each participant in the non-repeated sampling condition was correlated with their digit-span scores. This means taking each participant’s “hard” word-span score and subtracting it from their “easy” word-span score to obtain the “easy-hard” difference, and then correlating this difference with the corresponding digit-span score. Participants with larger memory capacities may show smaller “easy-hard” differences compared to participants with smaller memory capacities. If so, we would expect a negative correlation between digit-span and the “easy-hard” difference.

The results are displayed in Table 6. None of the “easy-hard” difference scores were significantly correlated with the digit-span scores, so we are confident that memory capacity does not significantly contribute to the “easy-hard” difference observed in the non-repeated sampling condition.

Table 6

Correlations Between “Easy-Hard” Word-span Differences and Digit-spans in the Non-repeated Sampling Condition

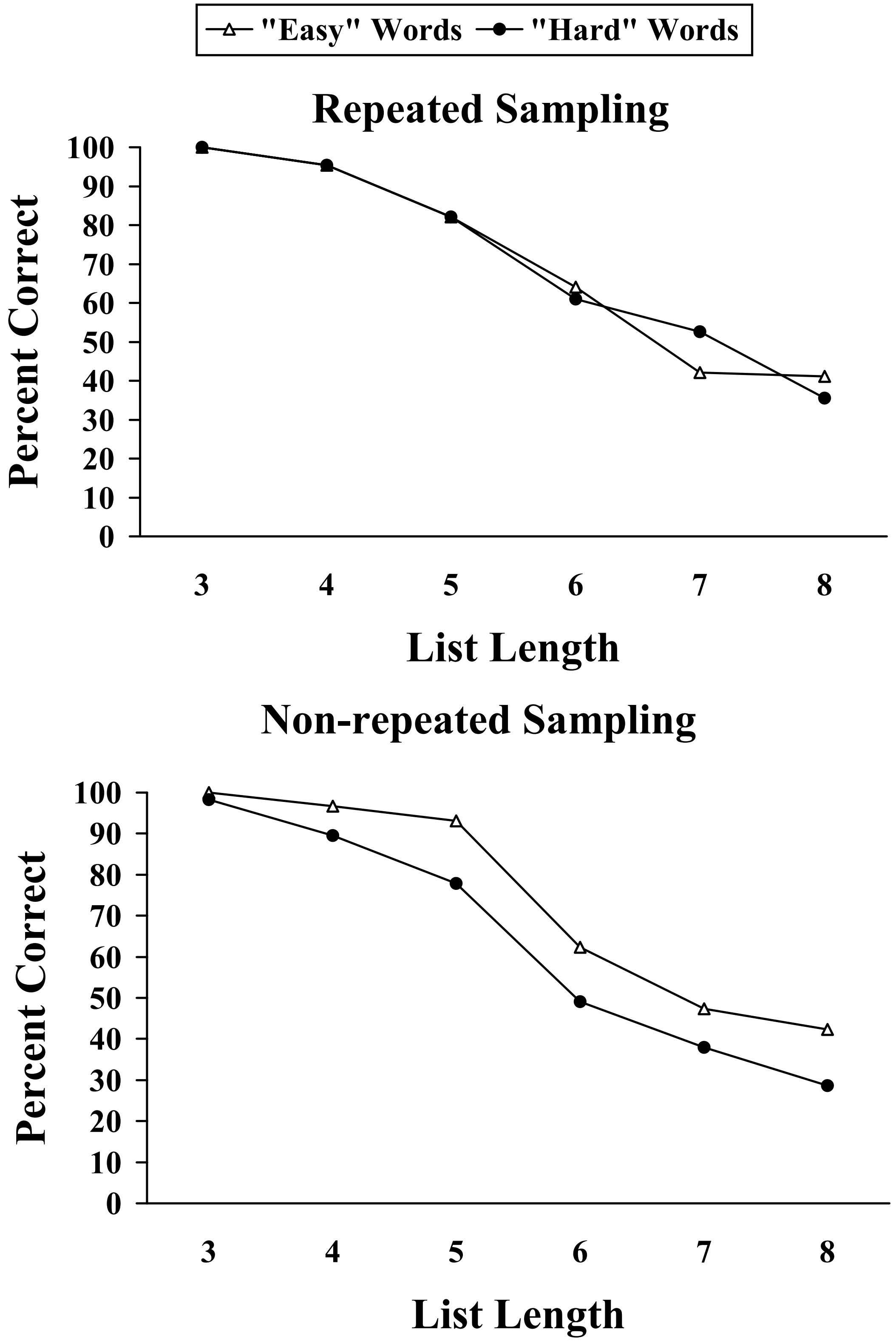
	Span Type			
“Easy-Hard” word-span difference	Highest	Strict	Absolute	Total
Highest	-.156	-.215	-.180	-.107
Strict	-.214	-.366	-.344	-.248
Absolute	-.130	-.272	-.261	-.166
Total	.067	.102	.119	-.005

List Length

We next analysed the performance of the “easy” and “hard” word-spans across each list length. The number of tokens correctly recalled across the two trials at each list length was summed and converted to a percentage correct score. A three-way mixed-design ANOVA revealed significant main effects of Word Type, $F(1, 36) = 10.63$, $MS_e = 257.03$, $p < .01$, and List Length, $F(5, 180) = 197.33$, $MS_e = 268.75$, $p < .001$. More importantly, the interaction between Word Type and Sampling was also significant, $F(1, 36) = 12.03$, $MS_e = 257.03$, $p < .01$. All other interactions and main effects were not significant. Figure 2 illustrates the main results.

Planned comparisons revealed that there were no differences between “easy” and “hard” word lists at each list length in the repeated sampling condition shown in the top panel. However, as shown in the bottom panel, there was a clear divergence between “easy” and “hard” words as list length increases in the non-repeated sampling condition. “Hard” word-spans were significantly shorter than “easy” word-spans at lengths 5 and 6, $F(1, 19) = 10.56$, $MS_e = 213.16$, $p < .01$ and $F(1, 19) = 10.50$, $MS_e = 158.90$, $p < .01$, respectively. The differences were marginally significant at lengths 4 and 8, $F(1, 19) = 3.29$, $MS_e = 143.71$, $p < .09$ and $F(1, 19) = 3.71$, $MS_e = 487.37$, $p < .08$, respectively. They were not statistically significant at lengths 3 and 7.

To determine if the “easy” word-span advantage increased with list length (which may indicate a systematic relationship to memory capacity), we compared the *magnitude* of the “easy-hard” differences at each list length with one another. Only two comparisons in the non-repeated sampling condition were statistically significant – length 5 vs. length 3, $F(1, 19) = 10.24$, $MS_e = 173.68$, $p < .01$, and length 6 vs. length 3, $F(1, 19) = 6.50$, $MS_e = 194.72$, $p < .05$. All other comparisons were not significant. This pattern


Repeated Sampling

Non-repeated Sampling

Figure 2. Percentage recall as a function of List Length, Sampling, and Word Type.

indicates that the magnitude of the "easy" word-span advantage essentially stayed the same at each list length from length 4 onwards.

Discussion

The pattern of results obtained in the present experiment demonstrates that lexical neighborhood properties exert an influence beyond word recognition, particularly when the use of limited coding and PI is minimized by a non-repeated sampling procedure. In particular, two findings strongly suggest that these lexical neighborhood effects are not influenced by memory capacity. First, the correlations between the magnitude of the "easy-hard" word-span *difference* scores and digit-span scores were not significant. This finding indicates that the observed memory span advantage for lists of "easy" words over lists of "hard" words, using non-repeated sampling, is not related to participants' STM capacity for digits in any systematic fashion. Second, the "easy" word-span advantage did not increase with increases in list length. This means the advantage is relatively constant and does not vary with the overall load placed on STM capacity by increases in list length.

It is difficult to account for these results without making a strong case for a much tighter integration between STM and LTM processes. Neighborhood properties do not exist in the phonological traces of the stimulus; they exist only in the LT structure of the lexicon and the pattern of activations within a lexical neighborhood. This means that the target stimuli (the lists of words presented) are not only activating their own phonological representations in LTM, but are also activating representations which are related to their phonological structure. Activation in LTM subsequently affects the coding and rehearsal processes in STM that are involved in the immediate memory span task.

What processes or mechanisms could account for this? One view of STM considers it to be the currently active portion of LTM (e.g., Anderson & Bower, 1973; Atkinson & Shiffrin, 1971; Engle, 1996). Consider what might be happening when the participant processes each token of a spoken word. During word recognition, the entire neighborhood of the word receives some activation (cf. Luce & Pisoni, 1998). "Hard" word lists, having denser neighborhoods, will therefore provide more interfering activations than "easy" word lists, resulting in a lower span score. Perhaps it takes time for the neighborhood activations to recede to levels where they will no longer interfere with whatever processes that are occurring to maintain the activation level of the presented words. In a priming study, Goldinger et al. (1989) showed that increasing the ISI between the prime and the target eliminated lexical neighborhood priming effects, suggesting that neighborhood activity does recede as time passes.

In the case of repeated sampling of a small set of words, the repetition of the words serves to strengthen their activations at every trial, and with enough trials, effectively makes them a closed set (cf. Sommers et al., 1997). The strength of the words' activations could possibly diminish the interference of the neighborhood activations. Therefore, lexical neighborhoods will have no effect on the maintenance of the tokens' activations. Any extraneous interference in repeated sampling would come from PI effects from previous lists.

Varying the ISI is one way to test this hypothesis. If participants are given more time to rehearse, elaborate or engage in activities that strengthen the words' activations, the disparity between "easy" and "hard" word-span scores using non-repeated sampling might diminish with longer ISIs. Another way to study this is to use tokens produced by different talkers at different speech rates. Work is currently underway in our lab on these issues.

In conclusion, our findings demonstrate that the lexical neighborhood properties of individual words influence performance in immediate memory span tasks. Our results also add to the rapidly growing literature that shows differences in performance depending on whether repeated or non-repeated sampling is used in memory span tasks. One implication for current memory models is that the specialised "phonological loop"-type modules will need to integrate LT phonological organisation and structure into their mechanisms and processes. It may turn out to be more parsimonious to consider STM and LTM processes in a more tightly integrated fashion, or even as a unitary memory system. Whatever architecture one adopts, the present findings demonstrate that the processes used in recognizing spoken words also affects their rehearsal and immediate recall and in turn, the memory span for non-repeated lists of spoken words. The present findings are consistent with the general principles of NAM – that words are recognized "relationally" in the context of other phonetically similar words in lexical memory. The process of recognising and memorising words appears to make use of a mechanism that exploits "competition" among a set of activated patterns that differ in frequency and acoustic-phonetic similarity.

References

Anderson, J.R., & Bower, G.H. (1973). Human associative memory. Washington, DC: Winston.

Atkinson, R.C., & Shiffrin, R.M. (1971). The control of short-term memory. *Scientific American*, 225, 82-90.

Baddeley, A.D. (1966a). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, 18, 363-365.

Baddeley, A.D. (1966b). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, 18, 302-309.

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

Baddeley, A.D. (1998). *Human memory: Theory and practice* (revised ed.). Boston: Allyn & Bacon.

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *Recent advances in the psychology of learning and motivation* (Vol. VII, pp. 47-89). New York: Academic Press.

Baddeley, A., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language*, 24, 119-131.

Baddeley, A.D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behaviour*, 14, 575-589.

Bavelier, D., & Potter, M.C. (1990). Visual and phonological codes in repetition blindness. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 134-147.

Broadbent, D.E. (1971). The magic number seven after fifteen years. In R.A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory*. New York: Wiley.

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

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, 55, 75-84.

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.

Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour*, 19, 450-466.

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

Elliott, J.M. (1992). Forward digit span and articulation speed for Malay, English and two Chinese dialects. *Perceptual and Motor Skills*, 74, 291-295.

Ellis, N.C., & Hennelly, R.A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology*, 71, 43-51.

Engle, R.W. (1996). Working memory and retrieval: An inhibition-resource approach. In J.T.E. Richardson (Ed.), *Working memory and human cognition* (pp. 89-119). Oxford: Oxford University Press.

Engle, R.W., Nations, J.K., & Cantor, J. (1990). Is "working memory capacity" just another name for word knowledge? *Journal of Educational Psychology*, 82, 799-804.

Goldinger, S.D., Luce, P.A., & Pisoni, D.B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, 28, 501-518.

Hoosain, R., & Salili, F. (1988). Language differences, working memory, and mathematical ability. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), *Practical aspects of memory: Current research and issues, Vol. 2: Clinical and educational implications* (pp. 512-517). Chichester: Wiley.

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

Landauer, T.K., & Streeter, L.A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behaviour*, 12, 119-131.

La Pointe, 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.

Luce, P.A. (1986). *Neighborhoods of words in the mental lexicon* (Research on Speech Perception Technical Report No. 6). Bloomington: Indiana University, Department of Psychology, Speech Research Laboratory.

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

Luce, P.A., Pisoni, D.B., & Goldinger, S.D. (1990). Similarity neighborhoods for spoken words. In G.T.M. Altmann, (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 142-147). Cambridge: MIT Press.

Martin, R.C. (1993). Short-term memory and sentence processing: Evidence from neuropsychology. *Memory and Cognition*, 21, 176-183.

Nairne, J.S., & Kelley, M.R. (in press). Reversing the phonological similarity effect. *Memory & Cognition*.

Nairne, J.S., Neath, I., & Serra, M. (1997). Proactive interference plays a role in the word-length effect. *Psychonomic Bulletin & Review*, 4, 541-545.

Naveh-Benjamin, M., & Ayres, T.J. (1986). Digit span, reading rate, and linguistic relativity. *Quarterly Journal of Experimental Psychology*, 38, 739-751.

Nusbaum, H.C., Pisoni, D.B., & Davis, C.K. (1984). *Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words* (Research on Speech Perception Progress Report No. 10). Bloomington: Indiana University, Department of Psychology, Speech Research Laboratory.

Richardson, J.T.E. (1996). Evolving issues in working memory. In J.T.E. Richardson (Ed.), *Working memory and human cognition* (pp. 120-154). Oxford: Oxford University Press.

Salame, P., & Baddeley, A.D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behaviour*, 21, 150-164.

Salame, P., & Baddeley, A.D. (1987). Noise, unattended speech and short-term memory. *Ergonomics*, 30, 1185-1193.

Salame, P., & Baddeley, A.D. (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology*, 41A, 107-122.

Scoville, W.B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20, 11-21.

Sommers, M.S., Kirk, K.I., & Pisoni, D.B. (1997). Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format. *Ear & Hearing*, 18, 89-99.

Texas Instruments. (1991). TI 46-word speaker-dependent isolated word corpus (CD-ROM). Gaithersburg: NIST.

Torretta, G.M. (1995). *The "easy-hard" word multi-talker speech database: An initial report* (Research on Spoken Language Processing Progress Report No. 20, pp. 321-334). Bloomington: Indiana University, Department of Psychology, Speech Research Laboratory.

Vitevitch, M.S., & Luce, P.A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, 9, 325-329.

Wright, R. (1997). *Lexical competition and reduction in speech: A preliminary report*. (Research on Spoken Language Processing Progress Report No. 21, pp. 471-486). Bloomington: Indiana University, Department of Psychology, Speech Research Laboratory.

Appendix

Summary Statistics of “Easy” and “Hard” Words

Word	Word frequency [a]	Neighborhood density	Mean neighborhood frequency	S.D. mean neighborhood frequency
		“Easy” words		
balm	36	13	17.769	23.392
cause	130	10	33.300	61.887
chain	50	19	39.421	60.597
check	88	15	15.933	24.546
chief	119	12	10.417	10.131
curve	45	13	15.231	28.877
death	277	10	30.700	52.686
deep	109	18	36.167	67.664
dirt	43	15	23.400	36.717
dog	75	8	11.875	10.480
down	895	20	38.700	84.053
faith	111	11	50.091	108.843
fig	72	16	44.438	91.059
firm	109	13	13.692	22.213
food	147	11	24.909	36.416
fool	37	18	73.000	72.454
full	230	15	59.533	59.517
gave	285	18	47.667	91.779
girl	220	16	6.688	14.605
give	391	7	70.429	114.394
hung	65	18	30.556	88.959
jack	92	17	74.412	232.408
job	238	19	5.316	9.995
join	65	8	27.000	30.822
judge	77	6	2.333	1.966
king	88	17	36.118	79.428
league	69	19	24.474	48.021
learn	84	16	51.000	89.634
leg	58	15	79.667	150.546
live	177	15	61.067	117.743
long	755	13	75.846	119.089
lose	58	17	65.765	148.671
love	232	11	42.455	74.908
mouth	103	7	41.857	88.207
move	171	8	16.375	22.897
neck	81	13	15.923	24.312
noise	37	4	43.500	46.573
page	66	16	52.938	70.721
path	44	14	16.571	23.533
peace	327	19	18.158	26.304

pool	111	18	25.278	24.989
put	437	14	20.000	21.711
reach	106	20	77.450	200.055
real	260	16	23.438	49.475
roof	59	13	49.692	103.900
rough	41	20	19.500	47.578
serve	107	14	24.786	30.892
shape	85	16	19.313	23.991
ship	83	19	18.211	23.930
shop	63	16	40.375	57.262
size	138	12	71.917	108.807
soil	54	13	28.385	38.372
south	240	5	22.200	45.180
teeth	103	12	19.333	28.334
theme	55	8	45.375	82.523
thick	67	13	78.462	139.452
thing	333	11	72.000	124.543
thought	515	11	36.546	34.958
vice	42	16	31.125	56.429
voice	226	7	29.000	39.942
vote	75	15	29.600	55.331
wash	37	7	65.000	64.151
watch	81	5	60.600	68.252
wife	228	15	72.933	181.523
work	760	20	47.000	88.941
young	385	7	20.000	23.252
“Hard” words				
ban	7	30	418.533	904.398
bead	1	26	299.192	1268.315
beak	1	28	298.214	1230.906
bean	5	25	396.360	1365.426
bud	9	23	216.478	911.452
bug	4	26	190.577	859.956
bum	7	24	287.375	943.625
chat	5	22	743.364	2478.973
cheer	8	27	87.222	207.708
chore	7	21	879.095	2288.535
cod	6	27	91.704	311.074
comb	6	24	92.917	198.442
con	9	22	100.091	377.804
cot	1	35	180.371	778.101
dame	7	23	117.522	226.350
den	2	33	130.636	304.239
doom	3	23	95.478	318.813
dune	1	27	120.333	330.676
fade	2	22	115.227	247.686
fin	3	30	825.333	3902.344

goat	6	26	95.923	201.477
gut	1	24	253.958	899.300
hack	3	30	438.800	1218.936
hag	1	25	479.160	1325.547
hid	6	25	711.440	1728.925
hoot	9	23	146.087	463.260
hum	5	25	232.520	613.737
hurl	3	22	199.182	641.743
kin	2	33	796.273	3725.878
kit	2	35	281.857	1475.166
knob	2	21	236.667	1004.977
lace	7	29	92.345	150.289
lad	6	34	187.382	876.051
lame	2	28	89.286	173.304
lice	2	26	138.308	289.602
mace	1	29	175.138	350.020
main	119	31	199.548	400.531
mall	3	24	192.125	609.653
mat	8	30	636.033	2126.576
mid	2	26	113.731	295.540
mitt	1	33	321.182	1519.421
moan	1	26	233.962	501.463
moat	1	31	161.387	448.038
mole	4	33	97.394	386.302
mum	1	23	144.739	402.090
pad	8	26	225.115	1001.929
pawn	2	21	366.952	1464.965
pet	8	30	96.633	185.430
pup	2	21	98.476	411.838
rat	6	37	480.270	1924.293
rhyme	4	25	121.560	346.470
rim	5	26	129.308	515.088
rout	1	21	164.476	471.632
rum	3	29	256.276	856.449
rut	1	28	221.214	830.562
sane	8	33	90.333	156.597
sill	4	35	116.714	392.673
soak	7	23	108.913	410.824
tan	9	25	379.280	875.452
wad	1	22	163.136	588.629
wade	2	24	248.375	609.158
wail	3	32	153.781	439.996
wed	2	25	295.080	663.664
weed	1	24	287.000	755.260
white	365	11	302.273	578.041
wick	4	26	432.692	1469.849

Note. The words used in the repeated sampling conditions are: balm, dirt, fool, noise, path, rough, vice and wash ("easy"); bud, con, hoot, main, pat, tan, white and wrong ("hard").

[a] Kucera and Francis (1967) counts.