

STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses

GUOJUN ZHANG and HERBERT A. SIMON
Carnegie-Mellon University, Pittsburgh, Pennsylvania

The capacity of short-term memory (STM) for verbal materials depends both upon the number of familiar chunks and upon the average complexity (number of syllables) of the chunks. A model that predicts STM capacity well was built, incorporating these two factors, for a number of experiments that used both Chinese and English language materials. One experiment, which used Chinese homophones, showed that STM has a nonacoustical (visual or semantic) component as well as an acoustical one. STM capacity for material encoded nonphonologically appears to be no greater than three chunks, whereas acoustical STM has a capacity of up to seven chunks. This result was confirmed by an experiment using chunks (radicals) that do not possess highly familiar one-syllable names.

Since Miller (1956) published his famous paper, "The Magical Number Seven, Plus or Minus Two," suggesting that the capacity of short-term memory (STM) was constant if measured in terms of chunks, much research has been done on chunking and its relation to STM span. Although the Simon and Chase studies (Chase & Ericsson, 1982; Chase & Simon, 1973a, 1973b; Simon, 1974) generally supported Miller's chunk-based model, other researchers (Baddeley, 1981, 1983; Baddeley, Thomson, & Buchanan, 1975; Mackworth, 1963; Salame & Baddeley, 1982; Vallar & Baddeley, 1982) have proposed a time-based model in preference to the chunk-based model. They believe that when a subject's span is measured in terms of spoken duration, STM capacity works out to approximately 2 sec, and hence should be measured in terms of the number of symbols that can be produced in that time, rather than in number of chunks.

An important goal of this investigation was to develop a model of STM capacity that reconciled these two points of view and explained the data that have been published in support of each. In our experiments, we measured STM spans for Chinese language materials (with subjects whose native language was Chinese), using a standard immediate-recall paradigm. We compared the results with data on English language materials, and will propose a model of STM capacity that accommodates both sets of data. Our experiments, using homophonous Chinese characters, also showed that there is a nonphonological component of STM with a capacity of about three chunks. This nonphonological component appears to correspond to Baddeley's "visuo-spatial scratch-pad" (Baddeley, 1983; Brooks, 1967).

In our analysis we need to pay careful attention to the differences between the English and Chinese languages.

This research was supported in part by Grant MD-07722 from the National Institute of Mental Health, and in part by a grant from the Alfred P. Sloan Foundation. The authors' mailing address is: Department of Psychology, Carnegie-Mellon University, Schenley Park, Pittsburgh, PA 15213.

English is an alphabetic writing system, whereas Chinese is logographic; written English has four principal structural levels (letter, word, phrase, and sentence), whereas Chinese has at least five (radical, character, word, phrase, and sentence); the entire English alphabet consists of only 26 letters, whereas Chinese has over 10,000 characters (about 7,000 in general use); and so on. We will presently explain in greater detail the nature and structure of Chinese radicals, characters, and words.

Among the additional characteristics of the Chinese language that are pertinent to our studies are: (1) many radicals do not have highly familiar names; (2) each character is pronounced with a single syllable and has semantic features; (3) most words consist of two characters, and most idioms consist of four characters; and (4) almost all characters have many homophones (an average of about six per character).

We tried to exploit these characteristics of the Chinese language in the design of our experiments, and in comparing the results of experiments using Chinese language stimuli with those that have been run with English language stimuli. In our experiments, Chinese materials were the main stimuli, and the main experimental paradigm was immediate recall.

EXPERIMENT 1 STM SPAN FOR NAMED AND UNNAMED CHUNKS

As a first step toward measuring STM span for Chinese language materials, we designed an experiment using radicals without names, characters, or two-syllable words as stimuli. The experiment had four goals: (1) to obtain a general idea of STM spans for the most basic language units in Chinese (radicals, characters, and words); (2) to probe the effect on STM span of the presence or absence of pronounceable names for visual stimuli; (3) to test for the existence of separate acoustical and nonacoustical

STM; and (4) to probe the effect on STM span of the number of syllables in each stimulus item.

The radical is an important construct in the Chinese language. Every Chinese character consists of a radical and possibly additional parts (see Figure 1 for more details). Chinese dictionaries are often indexed according to radicals, of which there are slightly more than 200. But, although educated Chinese people can recognize every radical, many radicals do not have commonly used oral names. In contrast to radicals, every character has a definite one-syllable pronunciation. Most Chinese words consist of two characters and are pronounced as two syllables (see Figure 1 for more details).

Since the radicals, characters, and words used in this experiment are all familiar items to literate Chinese, the chunking hypothesis of STM would predict the same STM capacity for all three kinds of items. A theory of capacity based on an articulatory loop would predict a capacity for two-syllable words only about one-half the capacity for characters. For any radicals at all to be retained in STM would require some form of nonphonological encoding in addition to, or as a substitute for, the usual phonological encoding.¹

Method

The subjects were six native Chinese graduate students and six Chinese scholars residing in Pittsburgh.

Thirty-eight sets of Chinese symbols were selected from a Chinese-English dictionary, each set consisting of a Chinese radical, a corresponding Chinese character (the radical was part of the character), and a corresponding two-character Chinese word (the character was part of the word). Two examples of the Chinese symbol sets are shown in Figure 1.

Three different types of stimulus sequences were generated: a radical sequence, consisting of a list of two to five radicals; a charac-

Set	Radical	Character	Word
1	心	爱	爱人
2	友	友	友谊

Figure 1. Two examples of the Chinese symbol sets used in Experiment 1.

Kind	Ten Basic Numbers									
Complex Chinese	壹	貳	叁	肆	伍	陸	柒	捌	玖	拾
Arabic	1	2	3	4	5	6	7	8	9	10

Figure 2. Chinese numerals used in auxiliary experiment.

Table 1
Mean STM Span of Chinese Numbers with Different Response Methods

Response	Mean	SD
Oral	9.50	1.77
Written	8.25	2.05

ter sequence, containing four to nine characters; and a word sequence, containing three to six words. Each sequence was arranged so that adjacent items did not form meaningful units. Thus, no two adjacent radicals formed a character, no two adjacent characters formed a word, and no two adjacent words formed a meaningful phrase. Each stimulus sequence was hand printed on a 35-mm slide with a blank space separating each pair of items.

Before Experiment 1 was carried out, an auxiliary experiment was run to gather baseline data on the effect of oral versus written responses on STM capacity. In this auxiliary experiment, after listening to a digit string from a tape, two subjects were asked to report the string orally; the other two were asked to write down the string using complex Chinese numerals (see Figure 2).²

The data from the auxiliary experiment are given in Table 1. We hypothesized before running the experiment that the difference between oral and written reports of the contents of STM would be significant, because written reports would take much more time than oral reports, and therefore would have a greater chance for decay of STM. [With intervening tasks, STM may decay to a level of 30% correct recall after about 9 sec (Murdock, 1961; Peterson & Peterson, 1959).] Contrary to this expectation, although the digit lists took three times as long to write as to recite orally (13.5 sec vs. 4.4 sec for the longest lists), the difference in number of items recalled in the two conditions was not significant [$t(3) = 2.40, p > .05$]. We concluded that writing time was not an important determinant of memory span, and hence, that the complexity of the characters used as stimuli would not have a significant effect on the measured spans.

In Experiment 1, the subjects were seated at a table and were run individually. The slides were shown by a projector at a rate of about 750 msec/item. That is, if there were three radicals or characters or words on the slide, it would be exposed for $750 \times 3 = 2,250$ msec. The experimenter triggered the slide projector manually, timing himself with a stopwatch. After each presentation, the subject was asked to write down the symbol sequence in the correct order. The response time was not limited. The subject was given five practice trials for each type of symbol sequence before the experimental trials were begun. Within a particular block of trials (radical, character, or word), the stimulus sequences were shown in order of increasing length, two sequences of each length. The order in which the three conditions were presented was counter-balanced across subjects.

Memory span was computed from the number of items in the correctly recalled stimulus sequences. The scores were calculated for each subject according to the following principle. Suppose that a subject could recall the pairs of a certain type of symbol sequence (say, sequences of characters) perfectly up to and including the pair that were K items long, then failed on at least one of the sequences of length (K + 1) but got N sequences greater than length K correct. The total score would then be $K + 0.5N$. (That is, each sequence longer than K that was recalled was credited as half a pair.) For example, in the case illustrated in Figure 3, $K = 3$, because all sequences up to and including the pair three items long were recalled correctly. Two longer sequences, one each of lengths four and five, were also recalled correctly, so $N = 2$. Thus, the score is $3 + 0.5 \times 2 = 4$ items.

Results

The data from Experiment 1 are shown in Table 2. The mean STM span is 2.71 for radicals, 6.38 for characters,

Length of Sequence	Sequence	Performance [+ correct - incorrect]
2 item	##	+
	##	+
3 item	###	+
	###	+
4 item	####	+
	####	-
5 item	#####	-
	#####	+
6 item	#####	-
	#####	-
7 item	#####	-
	#####	-

Figure 3. An illustration of the performance of a subject on sequences 2-7 items long; from this performance, a memory-span score is derived.

and 3.83 for two-character words. All three differences, between characters and radicals, between characters and words, and between radicals and words, are significant [$t(11) = 11.6, 8.7, 4.7$, respectively; all $ps < .001$]. During the recall of characters, subjects made 25 intrusion errors, of which 12 were homophone errors; that is, the responses were phonemically correct but graphemically wrong.

After finishing the experiment, every subject was asked to comment on his experiences in performing the task. All subjects considered it very difficult to recall radical sequences. Nine subjects thought the main reason for the difficulty was that radicals did not have common names, whereas three subjects thought the main reason was that radicals were seldom seen in isolation. Seven subjects reported that they sometimes organized the character sequences into two or three groups of characters. Five subjects reported that while writing a sequence they forgot other items.

Discussion

Although radicals appear in Chinese writing much more frequently than do the characters in which they are embedded (for each radical is used to form a number of different characters), the STM span for radicals was less than half that for characters. The subjects mentioned that radicals are seldom seen in isolation, although Chinese readers do see them in isolation when using Chinese dictionaries. Hence, the most probable reason for the very

small memory spans for these radicals is that they do not have familiar pronounceable names.

Since the span for the radicals is not zero, it appears that not all STM is encoded acoustically,³ but that there is some visual or semantic STM capacity.⁴ Thus, the experiment does provide some support for the two-STM theory (Baddeley, 1983; Brooks, 1967). However, the nonacoustical STM appears to have a capacity of only two or three chunks, much less than the acoustical memory. We will provide another estimate of the size of nonacoustical STM in the next experiment.

For the characters and words, both of which were pronounceable, STM capacity was neither a constant number of syllables nor a constant number of items (chunks). In terms of chunks, the STM span of two-character words (3.83) was distinctly smaller than that of characters (6.38), but in terms of characters or syllables, the STM span of two-character words (7.68) was larger than that of isolated characters (6.38). We will return to this finding and its explanation later.

The frequent occurrence of homophone errors (12 out of a total of 25 intrusion errors) suggests strongly that items are encoded phonemically. We will explore the homophone phenomenon from different sides in the next three experiments.

**EXPERIMENT 2
MEMORY FOR COMPLETE HOMOPHONES**

Since many studies have shown that phonemic similarity among test items reduces STM capacity (Baddeley, 1966; Conrad, 1964; Conrad & Hull, 1964; Tzeng, Hung, & Wang, 1977). It would be instructive to check the extreme case of phonemic similarity by using sequences of homophones as stimuli. If nonacoustical STM capacity is as limited as suggested by the results for the sequences of radicals in Experiment 1, then the span for sequences of homophones should not be more than two or three items.

Method

The subjects were six Chinese scholars residing in Pittsburgh.

There are many Chinese phonemes that correspond to more than six homonymous characters each, so that it was very easy to find stimuli for an experiment to test our hypothesis. Thirty-eight Chinese radicals without familiar names and 20 Chinese characters, each possessing more than five common homophones, were selected from a Chinese-English dictionary. Figure 4 shows a set of homophonic characters, all of which are pronounced "gong," with high tone.

Ten stimulus sequences were sampled randomly from the radicals, and 10 were sampled from the characters. The stimulus sequences varied in length from two to six items with two radical sequences and two character sequences of each of the five lengths. Each sequence was written on a card.

The subjects were run individually. The subject was asked to read the card at a speed of about 750 msec/item, cued by the experimenter's tapping quietly while presenting the card. The subject was then asked to write down the sequence as well as possible. There were five practice trials each for blocks of radical sequences and for blocks of character sequences before the experimental trials began.

Table 2
Mean STM Span for Three Types of Chinese Symbol Sequences (Experiment 1)

Type of Item	Mean	SD
Radicals	2.71	.52
Characters	6.38	1.08
Words	3.83	.75

Chinese Character	English Translation
工	work, labour
弓	bow
公	public, common
功	meritorious service
攻	attack, accuse
供	supply
宫	palace, temple
恭	respectful
龚	a surname

Figure 4. Set of homophonic characters, all pronounced "gong," with high tone.

Results

The data are shown in Table 3. The mean STM span of radicals is 3.00 and the span of homophones is 2.83. The difference between the two conditions is not significant [$t(5) = .67, p > .05$].

When asked after the experiment to report their experiences, every subject said that it was hard to recall the stimuli, and one subject complained, "It seems that my memory doesn't work today."

Discussion

The results support our hypotheses. Recall was the same, about three items, for both unnamed radicals and homophonic characters. For nonhomophonic characters, where an acoustic code could be used, the STM span approached seven items. Thus, the evidence for both acoustic and nonacoustic (visual or semantic) encodings is unequivocal, as is the smaller capacity, in chunks, of the latter.

Table 3
Mean STM for Two Kinds of Chinese
Symbol Sequences (Experiment 2)

Type of Item	Mean	SD
Radicals	3.00	.41
Homophones	2.83	.47

These results with Chinese language materials are especially interesting because it has often been claimed that Chinese readers, unlike readers of alphabetic languages, encode the ideographic characters directly from visual to semantic form, without going through an intermediary acoustic encoding. The high rate of homophonic intrusions in Experiment 1 and the low measured span of nonacoustic STM make this claim dubious. It would appear that the oral language is an essential intermediary in the extraction of meanings from both kinds of texts.

EXPERIMENT 3 STM SPAN FOR NON-HOMOPHONIC CHARACTERS

Because of the relatively high frequency of homophone errors in Experiment 1, it seemed desirable to measure the STM span for Chinese characters that do not possess homophones. However, it is hard to find such characters. Therefore, we decided to use Chinese family names as stimuli; most Chinese family names do not have other family names as homophones. Hence, if subjects knew that the stimuli were intended to represent family names, they would not be troubled by homophonic interference.

Method

The subjects were six Chinese scholars residing in Pittsburgh. None had participated in the earlier experiments.

Twenty-four Chinese family names without homonymous family names and 12 pairs of homonymous Chinese family names were selected. Sixteen stimulus sequences were sampled randomly from the nonhomophone group. The sequences varied in length from 3 to 10 characters, and there were two examples of each length. Another 16 stimulus sequences were sampled randomly from the homophone group, representing the same array of lengths. Each of the sequences was written on a card. No two items in a single homophonic pair were put in a single stimulus sequence.

The subjects were run individually. The stimulus sequences were presented from the shortest one to the longest. The subject was asked to read the card aloud at a speed of about 750 msec/item and then to write down the sequence as accurately as possible.

Before the experiment, the subject was told: "Note that all stimuli are Chinese family names."

There were five practice trials each for the homophone and nonhomophone conditions before the experimental trials began.

The measure was the number of items in the correctly recalled stimulus sequences, the score being calculated exactly as in Experiments 1 and 2. However, two measures of STM span were calculated for each condition. The first counted a sequence as correct only if all the graphemes in it were written down correctly; homophones were counted as errors. The second measure allowed homophonic substitutions for the correct graphemes. Clearly, any sequence correct by the first criterion would also be correct by the second.

Results

The results are shown in Table 4. Using the criterion of graphemic as well as phonemic correctness, the STM span in the nonhomophone condition is 7.08 items, whereas the STM span in the homophone condition is 5.33, a difference that is significant at the 5% level with the *t* test. Using the weaker criterion of phonemic cor-

Table 4
Mean STM Under Four Conditions

Type of Name	Graphemically Correct		Phonemically Correct	
	Mean	SD	Mean	SD
Nonhomophone	7.08	1.34	7.67	1.31
Homophone	5.33	.80	7.92	.67

rectness, the STM span in the nonhomophone condition is 7.67, whereas the span in the homophone condition is 7.97, a nonsignificant difference at the 5% level with the *t* test.

During the experiment, a few subjects complained that, for some family names in the homophone group, they were able to recall the sound but were not sure which family name in the corresponding homophonic pair had been displayed.

Discussion

Experiment 3 provides strong evidence that the subjects usually first translate the visual information into acoustical information, then decode the acoustical information back into the corresponding visual information. During the decoding process, one or another of the homophonic characters will be evoked and the corresponding grapheme written. Of course, with nonhomophonic stimuli, only one grapheme response (the correct one) could be evoked.

The span for homophonic names in this experiment (5.33) was substantially larger than the span for homophonic words (2.83) in Experiment 2. This difference can be accounted for by the fact that family names seldom have more than two homophones, whereas words have, on average, six. Hence, the dominant homophone is more likely to be correct in the former than in the latter case. Moreover, to recall the *nondominant* family name, the subjects would have to remember only one bit of information in addition to the pronunciation. This interpretation is supported by the results of Experiment 4 below, which measured the effects of relative homophone frequency.

If we measure the acoustical STM span by the number of syllables that were phonemically correct, the span in this experiment was between 7 and 8. In an earlier experiment, we saw that the span for orally presented digits was about 9.5, comparable to the spans for (English language) digits of English-speaking subjects. The larger span for digits, as compared even with nonhomophonic names, may be due to the greater familiarity of the former, which allows subjects to chunk them into pairs or triplets. The true span, in chunks, would then be smaller than the measured span. We have no independent evidence for this hypothesis.

EXPERIMENT 4 EFFECT ON STM SPAN OF HOMOPHONE FREQUENCY

When a subject has the task of writing a character that has homophones, but retains only acoustical information about the stimulus, we would expect homophones that occur frequently in the language to be substituted for less common homophones, but not vice versa. If this were so, then the measured STM span for frequently occurring characters would be larger than the span for less frequent characters. We will say that homophone A is dominant over homophone B if A is of more frequent occurrence in written Chinese than B. The purpose of this experiment was to test this hypothesis, that the effect of frequency on span could be explained in terms of homophone dominance.

Method

The subjects were six Chinese scholars residing in Pittsburgh.

First, 252 Chinese homophone groups were selected from a Chinese-English dictionary, each containing more than five homophonic characters. Then, 84 first-class characters, 84 second-class characters, and 84 third-class characters were selected from the 252 groups (allocating the groups randomly to the classes) according to the following criteria, applied by five Chinese scholars:⁵ (1) A character of the first class has no dominant homophones; (2) a character of the second class has only one dominant homophone; and (3) a character of the third class has more than four dominant homophones.

Fourteen stimulus sequences were sampled randomly from the 84 first-class characters. The sequences varied in length from three to nine items, with two examples of each of the seven lengths. Similar sequences were constructed from the second- and third-class characters. Each of the sequences was written on a card.

The subjects were run individually, counterbalancing the order of presentation of the conditions across subjects. In each condition, sequences were presented from the shortest to the longest.

Before the experiment, the subject was told: "When you are writing down a sequence, if you are not sure of one or more characters in the sequence but remember the pronunciation, please use a homophone substitution." The subject was asked to read every stimulus sequence aloud at a speed of about 750 msec/item according to the rhythm given by the experimenter, and then to write down the sequence as correctly as possible. There were 10 practice trials before the experimental trials began.

Results

The data are shown in Table 5, in terms of both the criterion of grapheme correctness and the criterion of phonemic correctness. By the criterion of graphemic correctness, the mean STM spans for the three classes, respectively, are 5.50 (first), 4.08 (second), and 2.67 (third). All differences between pairs of conditions are significant at the .01 level, using the *t* test. By the criterion of

Table 5
Mean STM Span Under Six Conditions

Type of Sequence	Graphemically Correct		Phonemically Correct	
	Mean	SD	Mean	SD
First-Class Characters	5.50	.91	7.25	.75
Second-Class Characters	4.08	1.20	6.75	.80
Third-Class Characters	2.67	.24	5.92	.79

phonemic correctness, the spans for the three classes are 7.25 (first), 6.75 (second), and 5.92 (third). The difference between the first and third is significant at the .05 level, using the *t* test; the other two differences are insignificant. The differences between spans measured by the two criteria were significant at the .01 level for second- and third-class stimuli, but only at the .05 level for first-class stimuli (all *t* tests).

After the experiment, when asked to give a retrospective protocol, every subject said that it was more difficult to recall the second-class characters than the first-class characters, and more difficult to recall the third-class characters than the second-class characters. When asked the reason for the difficulty, three subjects answered, "No idea"; three other subjects answered, "Because of the different familiarity with the stimuli." No subject indicated directly an awareness of the homophone issue.

The data of Experiments 1 and 2 indicated that visual STM has a span of only about three chunks. In the present experiment, the span, with graphemic correctness, of third-class characters was only 2.67, whereas the phonemic span of these same characters was 5.92. These data fit very well our hypothesis that, when only acoustical information can be retained, acoustically retained syllables will be coded in terms of the most frequent grapheme in the corresponding homophone class. Hence, acoustical information alone will seldom permit correct retention of a third-class character. On the other hand, in lists consisting of dominant homophones, the difference in spans measured in terms of graphemic and phonemic correctness, respectively, was barely significant. We interpret these findings as very strong evidence for the hypothesis that the effect of character frequency upon STM span is mediated through the mechanism of homophone dominance.

EXPERIMENT 5 EFFECT ON SPAN OF STIMULUS FAMILIARITY

As the results of Experiments 3 and 4 show, dominant homophones are retained better than nondominant homophones in STM. However, in these experiments, dominance is confounded with homophone familiarity. It would be helpful to know if familiarity has an effect on span when only dominant items are employed. As we know, the difference in familiarity between native- and second-language materials may be large, even if a person has a good knowledge of the second language. So, we thought that a comparison of Chinese digits and English digits, with Chinese bilinguals as subjects, would be suitable for such an experiment.

Method

The subjects were five Chinese scholars and one Chinese graduate student, none of whom had majored in English or English literature; all currently resided in Pittsburgh.

Twenty spoken lists of Chinese digits and 20 English digits were recorded on a tape. The length of the Chinese lists ranged from

Table 6
Mean STM Span of Chinese and English Digit Lists by Spoken Presentation and Oral Repetition (Experiment 5)

Type of Digit	Mean	SD
Chinese Digits	9.50	1.98
English Digits	5.67	.69

4 to 13 digits; the English lists ranged from 2 to 11 digits. All of the digits were chosen randomly. There were two digit lists of each length. Both Chinese and English lists were recorded from the shortest to the longest. All lists were spoken at a rate of about 750 msec/digit. Responses were oral. The measure of span was the number of items in the correctly (orally) recalled sequences. Scores were calculated by the same method as in Experiment 2.

Results

The data are shown in Table 6. The STM span for Chinese digits is 9.50 items and for English digits, 5.67 items. The difference between the Chinese condition and the English condition was significant [$t(5) = 5.452$, $p < .01$]. Subjects were able to recall many more Chinese than English digits. (Note that their span for Chinese digits was comparable to the typical spans of English-speaking adults for *English* digits.)

After the experiment, when asked to give a retrospective report, every subject reported that it still seemed much easier to remember Chinese digits than English digits, even after 1 year or more in the U.S. One subject said that Chinese digits sounded clearer than the English ones; another, that Chinese digits seemed to last longer than English digits after a digit list was heard.

Discussion

The difference between the STM spans for Chinese and English digits, respectively, is substantial. Hence, familiarity of stimuli would appear to be an important factor for STM. The experiment itself provides no explanation for the difference, although it rules out homophony, which was not present for any of the stimuli. A possible explanation lies in Baddeley's (1981) hypothesis that STM capacity is determined by the amount of material that can be rehearsed in a fixed time (about 2 sec). It is likely that speed of rehearsal will be strongly correlated with familiarity of the material being rehearsed. If so, a native speaker of Chinese could rehearse, in a given interval, more Chinese digits than English digits. In our next experiment, we sought some converging evidence for this hypothesis.

EXPERIMENT 6 THE EFFECT OF CHUNK SIZE ON STM SPAN

In our discussion of Experiment 1, we noted that, with increase in number of syllables per item, STM span measured in syllables grows, whereas STM span measured in chunks (items) declines. In the present experiment, we tested this finding further, using Chinese characters of one syllable, Chinese words of two characters (hence two syl-

lables), and Chinese idioms consisting of four characters (and syllables). We will fit an equation to our data that takes account both of numbers of chunks and numbers of syllables, and will provide an interpretation of the equation that reconciles the chunking theory of STM capacity with Baddeley's hypothesis of an articulatory loop.

Method

The subjects were six Chinese scholars residing in Pittsburgh. None had served as subjects in previous experiments.

A total of 104 Chinese characters, 50 Chinese words, and 40 Chinese idioms were selected from a Chinese-English dictionary. Sixteen sequences were sampled randomly from the character set, the sequences varying in length from three to seven items. Ten sequences, ranging in length from three to seven items, were sampled randomly from the word set. Another 10 sequences, ranging in length from two to six items, were sampled randomly from the set of idioms. Two sequences of each length were constructed for each set of stimuli, and each sequence was written on a card.

The subjects were run individually, and the order in which the three conditions was presented was counterbalanced across subjects. In the same stimulus set, sequences were presented from shortest to longest. The subject was asked to read each stimulus sequence aloud at speeds of about 750 msec/character, 1,125 msec/word, and 1,875 msec/idiom, according to the rhythm given by the experimenter, and then to write down the sequences as correctly as possible. Five practice trials were given for each condition before the experimental trials were begun. Different sequences were used in practice and test trials.

Results

The data are shown in Table 7. In terms of chunk size, the average STM spans for character, word, and idiom, respectively, are: 6.58 (C1), 4.58 (W1), and 3.00 (I1). Measured in numbers of characters or syllables, the corresponding average spans are: 6.58 (C2), 9.16 (W2), and 12.00 (I2).

The pairwise differences in spans between C1 and W1, W1 and I1, and C1 and I1 are, respectively, significant at the .001, .01, and .001 levels, by t test. The pairwise differences in span between C2 and W2, W2 and I2, and

C2 and I2 are, respectively, significant at the .001, .05, and .001 levels, by t test.

Discussion

The present experiment confirms previous findings, with both Chinese and English materials, that the more syllables in each chunk, the smaller will be the STM span measured in chunks and the larger will be the STM span measured in syllables. To explain these data, we formed the following hypothesis, which provides a role for both chunks and syllables in determining the STM capacity:

Short-term memory is limited by the amount of material that can be rehearsed (explicitly or implicitly) in a fixed time interval, T msec. Rehearsal requires an interval of time (a msec) to bring each new chunk into the articulatory mechanism and an interval of time (b msec) to articulate each syllable in the chunk beyond the first. Therefore, if S is the average size of a chunk, in syllables, then the STM capacity measured in chunks, C, will be related to the time interval, T, by

$$T = C[a + b(S - 1)] \tag{1}$$

or

$$C = T/[a + b(S - 1)] \tag{2}$$

In fitting these equations to data on memory spans, we have one extra degree of freedom which we can use by picking a plausible value of T, say something in the range of 2 to 3 sec. Fitting Equation 2 to the data in Table 7, and setting T = 2,000 msec, we find that a = 309.09 msec and b = 120.06 msec. The value of T was selected to give an articulation rate of about 6 syllables/sec, consistent with known rates for such tasks as reciting the alphabet. With these values for the parameters, Equation 2 becomes:

$$C = 2000/[309.09 + 120.06(S - 1)] \tag{3}$$

In Table 8, we compare the STM span measured in chunks from the experimental data with the spans computed from this equation and find an excellent fit. In fact, a chi-square test shows no significant difference between the experimental and predicted values (p > .05).

Nor is the relation expressed by Equations 1 and 2 limited to Chinese language materials. Table 9 shows the result of fitting Equation 2 to the data reported by Simon (1974), using English language stimuli. Again, we set T = 2,000, and found least squares values of a = 281.23 sec and b = 53.17 msec. Again, a chi-square test shows that there is no significant difference between the experimental and predicted values (p > .02). Furthermore, the value of the parameter, a, is very close to the value estimated for the experiment with Chinese materials—both are approximately 300 msec. In terms of the model underlying Equations 1 and 2, we interpret this parameter as the time required to bring a new chunk into the articulatory mechanism and produce its first syllable.

Table 7

Mean STM for Three Kinds of Chinese Symbols (Experiment 6)

Type of Item	Chunks Recalled		Syllables Recalled	
	Mean	SD	Mean	SD
Characters	6.58	.84	6.58	.84
Words	4.58	.67	9.16	1.34
Idioms	3.00	.58	12.00	2.31

Note—Each character, word, or idiom is counted as 1 chunk in the lists. Characters contain 1 syllable each; words, 2; and idioms, 4.

Table 8

Predicted (by Equation 3) and Experimental Mean STM for Three Kinds of Chinese Symbol Sequences

Type of Item	Chunks	
	Predicted	Experimental*
Characters	6.47	6.58
Words	4.66	4.58
Idioms	2.99	3.00

*From Table 7.

Table 9
 Predicted (by Equation 2, with $T = 2.000$, $a = 281.23$,
 $b = 53.17$) and Experimental Mean STM for
 Five Kinds of English Materials

Words and Phrases	Chunks	
	Predicted	Experimental*
1-syllable	7.11	7
2-syllable	5.98	7
3-syllable	5.16	6
2-word	3.84	4
8-word	2.90	3

*From Simon, 1974.

The values of b in the Chinese and English language experiments differ by a factor of two—120 msec as compared with 53 msec. We interpret this as meaning that the English syllables were rehearsed about twice as rapidly as the Chinese syllables. More experimentation will be required to determine whether this is a reasonable interpretation.

Baddeley et al. (1975) measured reading rates for words with various numbers of syllables. We fitted the following least squares equation to his data:

$$T = 399.8 + 83.7(S - 1), \quad (4)$$

where T is the reading time per word and S is the word length in terms of syllables. We see that the parameters in this equation are close to the parameters estimated from the immediate-recall data of Simon's (1974) short-term memory experiment (399.88 vs. 281.23 msec, and 83.7 vs. 53.17 msec, respectively, for parameters a and b —well within a factor of 2). Hence, Baddeley's data on reading speeds provides good support for our interpretation of the STM span data.

CONCLUSION

In this paper, we have examined short-term memory capacity in terms of numbers of chunks and numbers of syllables in the stimulus sequences. We found that STM capacity is constant neither in terms of chunks alone nor in terms of syllables alone, but can be expressed in terms of a weighted sum of chunks and syllables. This linear relation can be interpreted in terms of Baddeley's hypothesis of an articulatory loop with a fixed duration of 2 or 3 sec. It appears to require $\frac{1}{3}$ to $\frac{2}{5}$ sec to bring each chunk into the articulatory mechanism and express it, and $\frac{1}{20}$ to $\frac{1}{6}$ sec for each additional syllable in the chunk. These estimates correspond, in order of magnitude, to estimates derived from direct measures of reading rates.

The articulatory loop hypothesis, as modified by this study, assumes that subjects in the immediate recall experiment encode stimuli acoustically. The presence of numerous homophones among Chinese characters permitted us to verify this hypothesis, and to show, also, that there is a small visual or semantic short-term memory capable of retaining two or three chunks. We were able

to show also that, as would be expected, STM capacity for stimuli not possessing common pronounceable names (some Chinese radicals) is also only two or three items. This is true in spite of the fact that these same radicals are components in complete (and nameable) characters that have a span of about seven items.

Experiment 5 told us that not all the effect of familiarity upon memory span is mediated by homophony, for the span (with Chinese subjects) for Chinese digits was substantially greater than the span for English digits. We intend to examine the hypothesis that this difference can be explained, in accord with the articulatory loop hypothesis, by differences in reading rates for the two encoding of the digits.

REFERENCES

- BADDELEY, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, **18**, 362-365.
- BADDELEY, A. D. (1981). The concept of working memory: A view of its current state and probable future development. *Cognition*, **10**, 17-23.
- BADDELEY, A. D. (1983). Working memory. *Philosophical Transactions of the Royal Society, London*, **B302**, 311-324.
- BADDELEY, A. D., THOMSON, N., & BUCHANAN, M. (1975). Word length and structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, **14**, 575-589.
- BROOKS, L. R. (1967). The suppression of visualization by reading. *Quarterly Journal of Experimental Psychology*, **19**, 289-299.
- CHASE, W. G., & ERICSSON, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The Psychology of learning and motivation* (Vol. 16). New York: Academic Press.
- CHASE, W. G., & SIMON, H. A. (1973a). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.
- CHASE, W. G., & SIMON, H. A. (1973b). Perception in chess. *Cognitive Psychology*, **4**, 55-81.
- CONRAD, R. (1964). Acoustic confusion in immediate memory. *British Journal of Psychology*, **55**, 75-84.
- CONRAD, R., & HULL, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, **55**, 429-432.
- HAYES, J. R. M. (1952, January/March). Memory span for several vocabularies as function of vocabulary size. *Quarterly Progress Report*.
- MACKWORTH, J. F. (1963). The relation between the visual image and post-perceptual immediate memory. *Journal of Verbal Learning and Verbal Behavior*, **2**, 75-85.
- MILLER, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, **63**, 81-87.
- MURDOCK, B. B., JR. (1961). The retention of individual items. *Journal of Experimental Psychology*, **62**, 618-625.
- PETERSON, L. R., & PETERSON, M. (1959). Short-term retention of individual items. *Journal of Experimental Psychology*, **58**, 193-198.
- SALAMÉ, 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 Behavior*, **21**, 150-164.
- SIMON, H. A. (1974). How big is a chunk? *Science*, **183**, 482-488.
- TZENG, O. J. L., HUNG, D. L., & WANG, W. S. (1977). Speech recoding in reading Chinese characters. *Journal of Experimental Psychology: Human Learning and Memory*, **3**, 621-630.
- VALLAR, G., & BADDELEY, A. D. (1982). Short-term forgetting and the articulatory loop. *Quarterly Journal of Experimental Psychology*, **34A**, 53-60.

NOTES

1. Note that the radicals used in this experiment do not have common pronounceable names. Although subjects might try to invent names for the radicals on the spot, the conditions of the immediate recall experiment do not give time to associate such names with the visual stimulus.

2. These are not the numerals in everyday use, but special and more elaborate ones used on paper money, to avoid check-kiting, and so on. They are familiar to literate Chinese, who can read them but not write them without some effort and thought.

3. We will follow common practice in using "acoustic" as a generic term to refer to the whole gamut of phenomena associated with the au-

ditary modality, from the sound waves and the auditory encoding to the phonemic analysis, if any, and the articulation.

4. Alternatively, radicals may have been recalled by embedding them in characters and remembering the pronunciation of the characters. This is an unlikely possibility, especially in the light of the results we will report below for experiments with homophones.

5. At the time this experiment was performed, we did not have available a list of character frequencies. A subsequent test against such a list indicates that the scholars who made these classifications were generally correct in their judgments of relative homophone frequencies.

(Manuscript received February 3, 1984;
revision accepted for publication March 15, 1985.)