

# Retention of visually presented information in children's spelling

S. FARNHAM-DIGGORY and HERBERT A. SIMON  
*Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213*

One hundred and thirty-six 8-year-old children participated in two studies of their ability to spell familiar or nonsense words which they had seen or heard for 2-5 sec. A visual or auditory task was interpolated for 12-15 sec before spelling began. Spelling of meaningful words was most accurate when words were presented visually, except when the interpolated task was watching random letters of the alphabet. Nonsense spelling did not show a similar pattern. The theoretical implication of this difference is that, under certain conditions, visual word presentation provides more efficient pointers to long-term memory information than auditory presentation does. A model of the facilitation process is offered.

Traditional short-term memory research derives from relatively simple models. A set of stimuli are introduced into a buffer of some kind and after an interval are emitted. Research has been concerned with the nature of the material, the nature of the buffer, interval variations, additional demands upon the buffer or upon the processor generally, and response conditions (Norman, 1970). It has been assumed that such activity could be distinguished from long-term memory, considered to require more time for both input and output of stimulus materials, and to be organized in ways that are relatively immune to the interval variations or attention stressors affecting short-term memory function.

Most real-world tasks, however, require an interplay between the two systems. *Spelling* is an example. Spelling requires subjects to maintain some portion of a stimulus word in short-term memory, while retrieving and applying information—such as word sounds and orthographic rules—from long-term memory. Clearly, a model to account for such behavior must be concerned with the active integration and control of information stored in several ways, and resulting from different histories of presentation and rehearsal.

Presented here is a set of experimental explorations into the nature of such a model. We studied children because we wanted to investigate the early stages of memory consolidation and activation. It is presumed that adults spell in accord with their childhood training—i.e., that early stages of development are especially critical to an understanding of this particular type of adult behavior.

The experiments reported here arose from D. F. Simon's (unpublished) discovery that seeing a word

led to more accurate spelling than hearing it did, even after a 30-sec conversational interval.

Ordinarily, we give a child a spelling task by pronouncing the word to be spelled—often in the context of a sentence, in order to resolve the ambiguities of homonyms and near-homonyms. Since *spelling* means producing the string of letters that corresponds in the written language to the spoken word, it would appear that there could be no such thing as a visually presented spelling task: if a word was presented visually, the child could simply copy it without error. We know, however, if the child must look up a word in the dictionary, he may forget the spelling while walking back across the room to his desk. Similarly, if a word is presented visually to a child, and the child then performs an interpolated task for 15 to 30 sec, we might expect very little of the orthography of the word to be remembered across this interruption (Peterson & Peterson, 1959; Waugh & Norman, 1965). Hence, if spelling tests, one oral (*sound-spelling*) and the other visual (*sight-spelling*), using the same set of words and the same interpolated tasks, were administered to comparable groups of children, little difference might be expected in the performance of the two groups. In both conditions, the effective cue would be the spoken word—all that could be retained in short-term memory through the interruption—and it would be spelled on the basis of such orthographic information as the child already had associated with it or its phonemes in long-term memory.

However, in her pilot experiment, D. F. Simon established that this was not the case: *sight-spelling* is much more accurate than *sound-spelling*, even after 20 to 30 sec of interpolated activity. Apparently, additional information is conveyed by the visual stimulus and preserved in some memory through the interruption.

Simon and Simon (1973) have theorized that, except for words whose spellings are already

We are grateful to Darcy Miller, Ted Schiffman, and Steve Silverstein for technical assistance, and to Lee Gregg and Dorothea Simon for manuscript critiques. The research was supported by Public Health Service Grant MH-07722 from the National Institute of Mental Health.

overlearned, spelling begins with a self-generated list of sounds (phonemes) which are then used to retrieve associated orthographic representations (graphemes). Both phonemes and associated graphemes are in LTM along with rules for generating and testing various phoneme-grapheme correspondences (D. F. Simon, Note 1). Thus, their spelling theory includes recall and recognition components. Phonemes and graphemes are initially recalled, but once they are written, recognition processes are invoked. The correctness of a spelling word is tested by deciding if it "looks right."

More specifically, we can model spelling of orally presented or self-dictated words as a hypothetical series of operations:

- 1 determine *meaning* of word
- 2 is its spelling stored?  
if so, write it  
if not, go to 3
- 3 notice its *sound frame*
- 4 get next *phonemic chunk*
- 5 hold it in short-term memory
- 6 get an associated *graphemic chunk* and write it
- 7 does the written chunk match an item in the word-recognition list (does it "look right")?  
if so, return to 4 (or 3)  
if not, return to 6

As the model illustrates, the first step in deciding how to spell a word is to determine its meaning. Before writing *principle*, we need to decide if we mean a school official or an abstract rule. For that reason, a teacher who is dictating spelling words often provides explanatory sentences. Second, if the spelling of the word is already associated with its sound in long-term memory, we simply recover this information and produce the spelling. In Step 3, the *sound-frame* of the word is activated—we sound out the word as a whole, preparatory to segmenting it. Characteristics of the sound frame include the length of the phonemic string, vowel sounds, and intonation patterns. These characteristics are parsing cues; they point to *phonemic* (or syllabic) *chunks*. We decide which part of a word to spell first after sounding out the word as a whole. We may also recheck the sound frame, part way through a word, to decide which chunk to spell next. Once the phoneme has been factored out of the sound frame, a *graphemic chunk* can be selected to represent it. When that has been displayed (written), it can be corrected by comparison with spelling patterns and words stored in recognition memory.

In the case of sight-spelling, this entire process could be bypassed, and the word spelled directly from the visual information—provided the visual information could be retained in some memory over the interpolated task. The main purpose of these

experiments is to discover how much of this information is retained, and what the characteristics are of the memory in which it is held.

To this end, the materials used were both meaningful English words and nonsense words, and both visual and auditory interpolated tasks were employed. Differences in retention of meaningful and nonsense words, if they appeared in the data, would indicate something about the size and nature of the memory chunks, and could show what information was being accessed in long-term memory in the case of the meaningful words. Short-term retention is modality-specific, as the research reviewed by Reed (1973, pp. 111-122) details. Under many circumstances, retention is impaired more by an interpolated task in the same modality as the stimulus than by an interpolated task in a different modality.

These were the theoretical considerations that guided our first, exploratory study. No specific predictions were made, but generally it was anticipated that differences between nonsense and meaningful spelling would appear. The questions were: Would sight-spelling of meaningful words be superior to sound-spelling? Would this also be true of nonsense-word spelling? Are the effects specific to visual and auditory components of the spelling process, as revealed by error patterns? Would an interpolated task in the same modality as the stimulus word impair spelling? Would the modality of the interpolated task differentially influence sound-frame, phoneme, and grapheme errors?

Answers to these questions led to the design of a second study which varied not only modality, but also the content of the interpolated tasks, in a further effort to pinpoint the nature of the memory effects.

## EXPERIMENT I

In the first study, meaningful and nonsense words were presented in visual or auditory forms, and they were followed by 15-sec interpolated tasks which were also either visual or auditory in nature. Then the subject wrote the word that had been presented to him. The sound-frame, phoneme, and grapheme errors, were examined. An index of "word frame" memory, the number of letters, right or wrong, that were written down, was also considered. When we decide that a word "looks right," we are probably matching seen to remembered aspects of a word template or envelope. This will be called a *visual frame*. The number of letters in a word, the number and location of risers and descenders, and certain other cues comprise this frame. Only the difference between actual and remembered letter count was used as an index of frame memory.

The response time for words spelled correctly and for those spelled incorrectly, the latter classified by type of error, was also measured.

The subjects were third graders (8- to 9-year-olds).

In the first experiment, the subjects were drawn from an inner city school, from two racial populations. There were no effects of race or sex on error rates in either nonsense or meaningful conditions, hence data will not be reported here by sex or race.

## Method

**Materials and Procedure.** Ten words of medium difficulty for third graders selected from the Iowa Lists (Greene, 1954) were judged by local teachers to be within the spelling range of our sample of children. A single randomized list was used for all subjects. In the visual condition, the words were presented on a memory drum; in the auditory condition, they were tape-recorded by a female voice. A matching set of nonsense words were prepared by scrambling the real words. The real words were: *apple, October, green, dream, stamp, store, rather, along, sixty, and farmer*. The nonsense words were: *plape* (pronounced "pla-py"), *berocto* ("be-rock-toe"), *eengr* ("een-gr"), *amdre* ("am-dr"), *ampst* ("ampst"), *resto* ("res-toe"), *therra* ("ther-ra"),<sup>1</sup> *ongal* ("on-gal"), *ixtys* ("ix-tees"), and *merfar* ("mer-far").

The presentation of each stimulus, whether visual or auditory, required about 2 sec, and was followed within 2 sec by a 15-sec interpolated task. The visual interpolated task consisted of pictures from the Peabody Language Development kits, shown for 3 sec each. These were 5 x 7 colored pictures similar to those found in children's books. No two pictures were alike. The subjects were instructed not to talk about them, but merely to watch the pictures until the signal was given to pick up the pencil and begin spelling.

The auditory interpolated task was 15 sec of nursery rhyming, begun by the experimenter saying "Let's say 'Jack and Jill' together." The "togetherness" routine was intended to protect the subject from feeling that his rhyme memory was being tested. All the rhymes were familiar, though not always known fluently by third graders. At the end of 15 sec, the intervening task was terminated and the experimenter said, "All right, now you can write the word."

Subjects were provided with lined response sheets similar to paper used for school spelling practice and with soft pencils. They were requested to print. The children were told that we wanted to find out how well they could spell, but that this had nothing to do with school work or grades. Subjects in the nonsense conditions were advised that they might hear or see some words they did not know, but to go ahead and try to spell them anyway. All subjects were further advised that although the experimenter would be using a stopwatch, this did not mean that they were being tested for speed, and that they could take as long as they wished. They were asked to replace the pencil on the table as a signal that they were through writing a word.

If a subject struggled for more than 60 sec without being able to complete his writing (average word-writing time was 10 to 15 sec), the experimenter said, "All right, never mind, let's go on," and the word was deleted from the child's protocol. (This happened 26 times, distributed over 1,504 spelling instances of the two experiments.)

The words were presented one at a time, and total testing took about 15 min. Each child was tested individually in a private room.

**Measures.** In most of the tables, spelling accuracy (or rather inaccuracy) will be expressed in terms of percent of words spelled incorrectly. Incorrect spellings will be attributed to sound-frame errors, phoneme errors, or grapheme errors. Each incorrectly spelled word will be placed in one of these three classes according to the following procedure:

*Sound-frame errors* include incorrect ordering of phonemes, missing syllables, or other gross misrepresentations (e.g., *damer* for *dream*). *Phoneme errors* involve substitution of phonemes different from those in the stimulus (*dreen* for *dream*). Phoneme errors were counted only for words whose sound frames were correct. *Grapheme errors* are incorrect orthographic renditions of the phonemes (as in *driem* for *dream*). Grapheme errors were counted

only for words whose sound frames and phonemes were correct.

While these types of errors are of major theoretical interest, their conditional probability distributions are undetermined. For analyses of variance, therefore, only the total error proportion was used—the proportion of words spelled incorrectly (for whatever reason) per subject.

*Percent of visual frames correct* is the number of words having the correct number of letters divided by number of words spelled. The visual frame count does not depend on whether or not the letters are correct, or are in the correct order, but simply on whether the length of the spelled word agrees with the length of the stimulus word.

*Response time per correct word* and *response time per wrong word* were measured in seconds. Each subject was instructed to keep his pencil on the table and to place it back on the table as soon as he had finished a word. The watch was started when the subject picked up his pencil and was stopped when the pencil was put back down. A mean time was computed for each subject. Mean time per wrong word was also computed for each type of error (sound-frame, phoneme, and grapheme).

**Subjects.** Sixty-four third graders from an inner city school were tested. Equal numbers of boys and girls, whites and blacks, were assigned randomly to each condition. The experimenters were white females, as were the children's teachers. The children were well motivated and cooperative.

**Design.** Once sex and race were eliminated as influential factors, the stimulus and interpolation modalities were crossed in a 2 x 2 analysis of meaningful and nonsense-word spelling separately. A child either saw or heard words to be spelled, and after each word spent 15 sec either saying nursery rhymes or looking at pictures before beginning to write. No subject received more than one treatment, and there were eight subjects per condition.

## Results and Discussion

**Errors.** The main effect of stimulus modality was significant in the meaningful condition [ $F(1,28) = 9.22, p < .01$ , for the per subject error proportions]. As Table 1 shows, the error proportion for words presented visually was .214; for words presented aurally, the proportion was .406. This effect supports D. F. Simon's pilot findings that sight-spelling was superior to sound-spelling, even after an interval exceeding the life-span of the Sperling (1960) memory.

The main effect of the modality of the interpolated task was not significant in the meaningful condition. The percent correct following visual interpolation was .303, and the percent correct following auditory interpolation was .315. There was a significant interaction, however: when the stimulus word had been presented visually and was then followed by the pictorial interpolated task, the overall error rate was low: .114. When the stimulus word had been presented visually and was then followed by rhyming interpolation, the overall error rate was relatively high: .320. This probably resulted from the fact that rhyming interpolation was more difficult than picture watching, a discrepancy that was corrected in the second experiment. The error proportions following auditory presentation were uniformly high, regardless of the nature of the interpolation. The  $F(1,28)$  for this interaction was 9.30,  $p < .005$ .

When the error rates for sound-frame, phoneme, and grapheme errors are examined, there is no

Table 1  
 Mean Proportions of Spelling Errors and Mean Total Proportions of Errors for Nonsense and Meaningful  
 Word Conditions: Experiment I (N = Eight Per Group)

Modality		Error Categories			Total Proportions of Errors	SE
Stimulus	Interpolation	Sound-Frame	Phoneme	Grapheme		
<b>Meaningful Words</b>						
Visual	Visual	.051	.051	.012	.114	.0388
Visual	Auditory	.107	.160	.053	.320	.0530
Auditory	Visual	.114	.342	.038	.494	.0777
Auditory	Auditory	.085	.155	.070	.310	.0636
<b>Nonsense Words</b>						
Visual	Visual	.295	.423	.090	.808	.0777
Visual	Auditory	.215	.380	.139	.734	.0919
Auditory	Visual	.042	.278	.555	.875	.0318
Auditory	Auditory	.074	.353	.441	.868	.0282

evidence that the relative distribution was specific to the modality of the interpolated task. Following auditory interpolation, the sound-frame and phoneme error rates together totaled .253, while the grapheme error rate was .062. Following visual interpolation, the sound-frame and phoneme error rates together totaled .278, and the grapheme error rate was .025. Relative levels of the specific error rates were no more strongly affected by the stimulus modality. Both sound and sight errors were proportionately lower when the word had been seen than when it had been heard. The combined sound-frame and phoneme error proportion was .182 when a word had been seen, and .353 when it had been heard. The grapheme errors were .033 and .053, respectively.

The single main effect that appeared with meaningful words may be expressed by saying that the children were able to spell 59% of the words correctly when these were dictated to them, while, when the words were shown to them visually, they picked up and retained enough additional information about the spellings to spell 79% correctly. This additional information was retained even better when the interpolated task was the relatively easy one of picture watching, compared to the more difficult one of reciting nursery rhymes. The outcome was, however, contrary to predictions that could be made about same-modality interference with short-term retention. We will return to this point.

In the nonsense condition, the error rate was high—only about 18% of the words were spelled correctly, as compared with 69% (overall) in the meaningful condition. Moreover, in the nonsense condition, no significant effect was produced by the modality of either the stimulus or the interpolated task. The visual stimuli produced slightly better spelling than the auditory stimuli (error rates of .770 and .871, respectively), but the difference was both insignificant and smaller than the difference in the meaningful condition. This might be thought especially surprising for the following reason: In the nonsense condition, there was no objectively "correct"

spelling for aurally presented stimuli—the subjects in this condition were given no information that would allow them to discriminate between phonetically identical spellings. Therefore, it might be expected that performance with auditory stimuli would be particularly poor in the nonsense condition, and the deficiency as compared with the visual presentation to be larger than in the meaningful condition. Since this expectation was not fulfilled, it will be necessary to explain later why additional visual information, which proved important in the meaningful case where it was redundant, was unimportant in the nonsense case where it provided genuinely additional information. An explanation will be proposed after the reviewing of data from both experiments has been completed.

One other effect that is easier to understand appeared in the nonsense condition. When the stimulus was visual, 66% of the errors were sound-frame and phoneme errors, and only 11% were grapheme errors. When the stimulus was auditory, 37% were sound-frame and phoneme errors, while 50% were grapheme errors. With the auditory stimulus, sound-frame and phoneme errors were only slightly higher in the nonsense than in the meaningful condition (nonsense sound-frame plus phoneme errors, 37%; meaningful, 35%), while with the visual stimulus, they were almost four times as high (66% vs. 18%). The reverse was true for grapheme errors: with the auditory stimulus, grapheme errors rose from 5% in the meaningful condition to 50% in the nonsense condition, while with the visual stimulus, they rose only from 3% to 11%. Clearly, different kinds of information about the nonsense words were transmitted by the two modalities of presentation. But the children could not use the visual information to facilitate overall spelling, as they could in the case of meaningful words.

**Correct visual frame rate.** Table 2 shows the percentages of words spelled for which the number of letters in the spelled word agreed with the number of letters in the stimulus word. In meaningful spelling, visual stimuli produced a significantly higher

Table 2  
Mean Proportions of Correct Visual Frames During Meaningful and Nonsense Spelling: Experiment I (N = Eight Per Group)

Interpolation Modality	Stimulus Modality			
	Visual	SE	Auditory	SE
Meaningful Words				
Visual	.937	.0318	.725	.0707
Auditory	.912	.0353	.763	.0565
Nonsense Words				
Visual	.381	.0848	.462	.0742
Auditory	.662	.0813	.417	.0388

percentage correct:  $F(1,16) = 12.6$ ,  $p < .005$ . This effect was independent of the modality of the interpolated task. In nonsense spelling, the percentage correct did not vary much with stimulus modality, although there was an interaction with modality of the interpolated task that would be predicted by a same-modality interference hypothesis [interaction  $F(1,16) = 8.61$ ,  $p < .01$ ]. But more important, the percentage correct was higher for all forms of stimulus and interpolated task in the meaningful than in the nonsense condition. Seeing (or hearing) the stimulus "word" was an inadequate substitute for having information about a real word already available in long-term memory.

**Time per correct response.** There were no effects of stimulus modality upon speed of response. Interpolation modality, however, had an effect: Spelling time was about 2 sec per word longer following auditory interpolation, compared to visual interpolation (9.04 and 11.27 sec per correct word, respectively). This effect was replicated in the second experiment, which also used meaningful spelling words, but did not appear in the nonsense condition.

**Time per wrong response.** Table 3 shows the mean spelling times for each of the error groups, for both experiments. The general picture is clear: although overall time levels may differ from experiment to experiment, words containing sound-frame errors took longer to spell than words containing phoneme errors, which in turn usually took longer to spell than words containing grapheme errors. This is true even

for the nonsense words. The one exception in Table 3 is the rather high grapheme error time in the meaningful condition of Experiment I, but that may be a chance result of the small number of observations (13) in the set.

With reference to the model sketched in the introduction, it is tempting to speculate that grapheme errors may be produced relatively quickly because the subject is looping back through only the last steps (6 and 7) of the spelling program. Phonemic and sound-frame errors should take progressively longer to generate, since they represent loops back at least through Step 2.

## EXPERIMENT II

The outcome of the first experiment revealed that visual presentation of a meaningful spelling word provides information, retainable over an interpolated task, that is not available after auditory presentation of the word. Since only one form of interpolated task was used in each modality, it was important, for interpreting these findings, to learn how sensitive they might be to the exact nature of the interpolated task. To investigate this question, a second experiment was conducted to contrast interpolation tasks of three types in two modalities. *Meaning* interpolation was provided by pictures, as before, in the visual modality; the matching auditory condition used tape-recorded sentences about the same pictures (which were not seen in the auditory condition). *Letter* interpolation was provided by random sequences of printed English letters (visual condition) or spoken ones (auditory condition). *Frame* interpolation was provided by sequences of Hebrew words (unrecognized as such by the subjects) either written (visual condition) or spoken (auditory condition). Only meaningful words were spelled.

It was conjectured that the specific interpolated tasks might impair specific stages of the spelling process. Thus, for example, the auditory frame task might increase the sound-frame error rate, the visual frame task might increase the visual-frame error rate, and the letter interpolations might increase the

Table 3  
Mean Number of Seconds Writing Each Word

	Time Per Wrong Word						Time Per Correct Word	
	Sound-Frame Errors	N*	Phoneme Errors	N*	Grapheme Errors	N*		N*
Experiment I								
Meaning	18.59	27	15.23	54	17.92	13	10.09	211
Nonsense	18.16	48	17.86	107	15.97	81	14.07	51
Experiment II								
Combined Conditions	13.82	28	10.19	91	8.49	105	7.55	633

\*Number of words on which each mean is based.

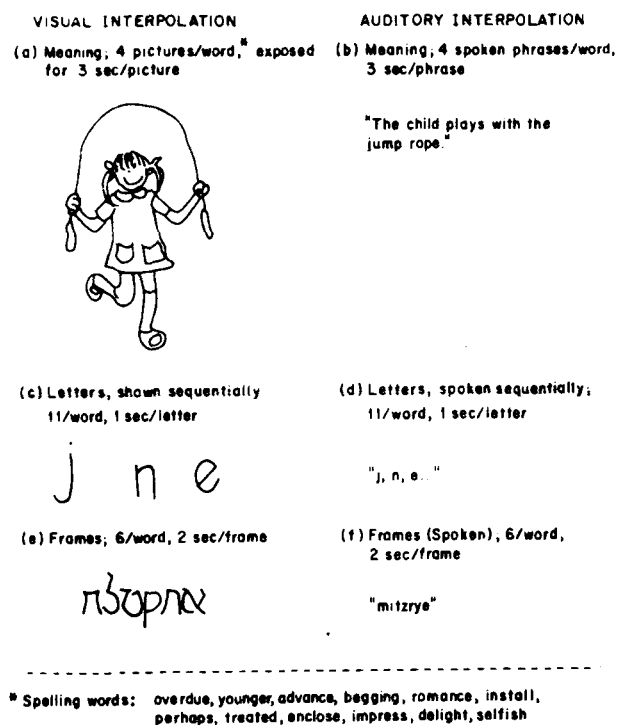


Figure 1. Samples of experimental materials used in Experiment II.

grapheme and phoneme error rates when the interpolations occurred in the visual and auditory modalities, respectively. The meaning interpolation (pictures or sentences) was included primarily to see if the relatively high error rate caused by auditory interpolation in Experiment I would replicate if the recitation aspect of the interpolated task were eliminated. In Experiment II, the subjects were merely to listen (to interpolated auditory materials) rather than to recite.

An important general change was made in the auditory presentation. In Experiment I, the visual presentation contained information about how the words were spelled, whereas the auditory presentation did not. To equate the information actually presented in visual and auditory modes, Experiment II gave the spelling, as well as the pronunciation, of each word in the auditory stimulus conditions. Thus, for example, the taped voice would say: "Advance, a-d-v-a-n-c-e, advance." Each word was pronounced, spelled aloud, and pronounced again, before the interpolation began. In the visual stimulus condition, the total exposure time equaled the time required for the auditory presentation.

## Method

**Materials and Procedure.** Figure 1 shows samples of the interpolation stimuli, matched as closely as possible across the auditory and visual modalities. For the meaning condition, pictures were drawn by an art student specializing in children's illustrations. They were in black and white, and depicted simple activities of interest to children. The matching auditory interpolated task was

generated from these pictures— simple declarative sentences were constructed, describing the pictures. For any given set of four pictures shown in the visual condition, the matching set of four sentences, in the same order, was used for the auditory interpolation. No pictures or sentences were ever repeated.

The interpolation letters were randomly selected by a computer program. The one constraint on the selection was that for every set of 11 letters, the number of vowels must equal the number in the spelling words, approximately 2 per set of 11 letters. The letter names ("ay," "bee," "cee") were said, rather than phoneme sounds, on the assumption that the letter names are what one says to oneself during spelling.

The interpolation frames were Hebrew words (not always correctly spelled), either written or spoken. (There were no children in this particular sample who understood them). The size of the written frames was selected to cover about the same amount of space as would be covered by a child writing a seven-letter word. The spoken frames contained approximately the same number of syllables as the spelling words, but of course the vowel sounds and intonation patterns differed from English.

All of the sound stimuli were recorded by the same male voice, that of a professional actor. The visual stimuli were presented in hinged plastic frames (Flip Files), paced by a time click on a tape. Twelve different tapes were prepared, incorporating the various auditory stimuli and pacers. For example, in the auditory-visual meaning condition, the experimental tape presented a spelling word, four pacing signals that controlled picture changes and exposure time, and a signal (a double click) that told the child to begin to write. The experimenter stopped the tape recorder until the child had finished, and then started it up again for the next trial. Thus, the 12 experimental conditions were virtually run by the tape recorder, with a little help from the hands of the experimenter. Each stimulus word was exposed for 4.5 sec, and the interpolated task began within .5 sec of stimulus offset. The interpolation periods each lasted for 12 sec. Different numbers of items (e.g., letters or pictures) were necessary to sustain processing for this constant interval.

Because the decision had been made to sample from a suburban, middle-class population, a new set of more difficult spelling words was selected. They were from the Iowa fourth-grade 10% lists, i.e., they were words spelled correctly only 10% of the time by a normative fourth grade group. Twelve words were selected, beginning equally often with vowel sounds and consonant sounds. Concrete nouns were eliminated, to decrease the imagery component. Each word was seven letters long, making possible a more precise time per word estimate. They are shown in Figure 1.

**Subjects.** Three boys and three girls were assigned randomly to each of the 12 experimental conditions. They were from two suburban third grades. Each child was tested individually and privately in a 15-min session. No child received more than one treatment. All were highly motivated and cooperative.

**Design.** As before, when it was ascertained that sex was not an influential factor, the data were analyzed in a 2 x 2 x 3 analysis of variance crossing stimulus modality, interpolation modality, and interpolation content (meaning vs. letters vs. frames). Following that overall analysis, separate analyses were conducted of the different interpolation content conditions, as will be explained.

## Results and Discussion

**Errors.** When the interpolation tasks were combined, there was a significant effect of stimulus modality on the total error rate [ $F(1,60) = 24.6$ ,  $p < .001$ ]. Despite attempts to increase the auditory information by dictating the actual spelling of the word, sound-spelling remained significantly worse than sight-spelling. The overall error proportion was .398 for auditory word presentation and .126 for visual word presentation.

There was also a significant main effect of

Table 4  
Mean Proportions of Spelling Errors and Mean Total Proportions of Errors for Three  
Conditions of Interpolation Content: Experiment II (N = Six Per Group)

Modality		Error Categories			Total Proportions of Errors	SE
Stimulus	Interpolation	Sound-Frame	Phoneme	Grapheme		
<b>Meaning Interpolation</b>						
Visual	Visual	None	.055	.027	.083	.0204
Visual	Auditory	.014	.056	.014	.084	.0734
Auditory	Visual	.083	.292	.111	.486	.1674
Auditory	Auditory	.014	.070	.211	.296	.1143
<b>Letter Interpolation</b>						
Visual	Visual	.098	.127	.141	.366	.0571
Visual	Auditory	None	.029	.014	.043	.0408
Auditory	Visual	.028	.181	.292	.500	.1429
Auditory	Auditory	.014	.153	.222	.389	.0734
<b>Frame Interpolation</b>						
Visual	Visual	.042	.042	.027	.111	.0612
Visual	Auditory	.014	.014	.041	.069	.0408
Auditory	Visual	.028	.111	.222	.361	.1143
Auditory	Auditory	.056	.141	.155	.352	.0979

interpolation [ $F(1,60) = 4.32, p < .05$ ]. The greater impairment resulted from interpolation in the visual modality. The error proportion for visual interpolation was .318; that for auditory interpolation was .206. As can be seen from Table 4, most of the information loss from visual interpolation occurred when the stimulus was presented in the auditory modality: .486, .500, and .361 were the highest error rates in their respective conditions. Except for the letter condition, errors following visual interpolation were low when the stimulus was presented in the visual modality.

Although the overall analysis of variance was not sensitive to this interaction, separate analyses show it clearly. Each interpolated task can be treated as if it comprised a separate experiment. Thus, instead of one large factorial study, there are three smaller ones: a 2 (stimulus modality) by 2 (interpolation modality) analysis of meaningful interpolation, a similar 2 by 2 analysis of letter interpolation, and a third 2 by 2 analysis of frame interpolation. In the meaning and frame analyses, only the main effect of stimulus modality was significant: the meaning  $F(1,20)$  was 7.88,  $p < .025$ ; the frame  $F(1,20)$  was 9.98,  $p < .005$ . In the letter condition, however, both the main effect of the stimulus modality [ $F(1,20) = 7.36$ ], and the main effect of the interpolation modality [ $F(1,20) = 6.09$ ] were significant,  $p < .025$  in both cases.

There were no other effects of interpolation task on errors. In particular, sound-frame interpolation (Hebrew words) did not increase the sound-frame error rate, nor did spoken letter interpolation increase the phoneme rate. There was only some slight evidence, in Table 4, that the visual letter interpolation increased the grapheme error rate, which was .217 in that condition, compared to .125 and .069 in the frame and meaning conditions,

respectively. The effect of auditory interpolation on sight-spelling disappeared when recitation was not required.

**Correct visual-frame rate.** As before, visual presentation of the spelling words led to a higher rate of visual frames correct than aural presentation, regardless of type of interpolated activity [ $F(1,49) = 23.5, p < .001$ ]. The rate was not decreased by the interpolated task of watching Hebrew words.

**Time per correct response.** An analysis of variance revealed only an effect of interpolated task modality [ $F(1,48) = 4.55, p < .05$ ]. Following visual interpolation, the mean time per correct response was 7.13 sec; following auditory interpolation, it was 7.97 sec. This replicated the effect found in the first experiment, although the absolute response times were lower, due, probably, to the increased general ability of the middle-class sample.

**Time per wrong response.** These and other times are shown in Table 3. While the absolute times are faster, they show the same general ordering for different error categories as in the first experiment.

## GENERAL DISCUSSION

Two experiments have explored the question of the extent to which the sight of a word, as compared to its sound, improves an 8-year-old's ability to spell it. We should like now to propose an explanation of our main findings—an explanation that is consistent with widely accepted views about the organization of memory.

We postulate a conventional short-term memory, whose capacity is limited to a small, fixed number of chunks, or familiar units. In accordance with the analysis of Waugh and Norman (1965), we postulate that during the performance of an interpolated task,

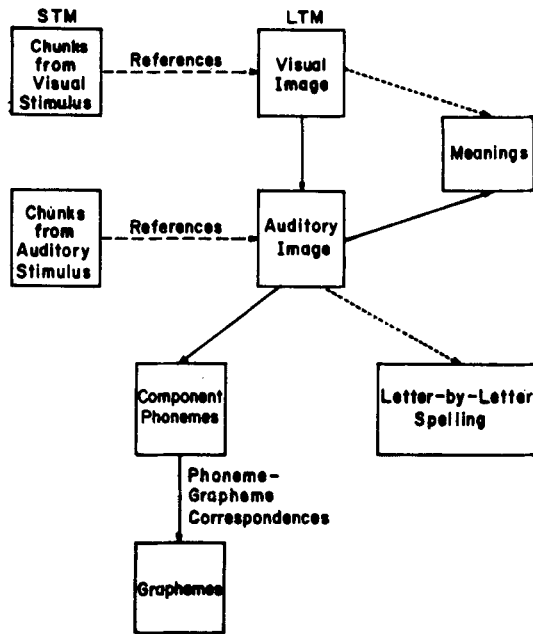


Figure 2. Hypothesized organization of associations in long-term memory (—→ long-term memory associations; ···→ possible long-term memory associations; ····→ short-term-to-long-term-memory access path).

only one or two chunks of previously stored material can be retained. A chunk in short-term memory is to be regarded as a "pointer" to information about a familiar item stored in long-term memory, by means of which access can be gained to that information.

In accordance with the Simon and Simon (1973) spelling theory, when a familiar word is presented aurally to a subject, the chunk referencing that word gives access to its auditory image, its meaning, and information (which may be incomplete) about its spelling, all stored in long-term memory. If the word is very familiar, its visual image, more or less complete, may be accessible from the auditory chunk, but it need not be. Since the written language is normally learned after the aural language, the aural image of a word will always be accessible by association with its written image, but the written image will not always be accessible from the auditory image. This fundamental asymmetry in the organization of long-term memory is depicted in Figure 2. Of course, a (more or less complete) visual image can be accessed upon presentation of the familiar word in its visual form.

When a spelling word is presented orally, a single chunk retained in short-term memory during performance of the interpolated task will serve to access the auditory image and associated spelling information when the time comes to spell the word. The individual phonemes of the word do not need to be held in short-term memory during the interpolation interval; this is the reason the word to be spelled can be remembered over that interval.

When the spelling word is presented visually, a

chunk referencing the *visual* image in long-term memory can be retained over the interpolation (Hintzman & Summers, 1973; Parkinson, 1972; Scarborough, 1972). This chunk, however, gives access to the visual image (i.e., at a minimum, the information used to recognize the word) as well as to the auditory image, and all the information associated with the latter, including the spelling program outlined earlier in this report. Hence, the chunk in short-term memory referencing the visual image will always provide as much or more information about the word's spelling as the chunk in short-term memory referencing the auditory image. From this asymmetry, it follows that after the interpolation, the visually presented word will be spelled more accurately than the aurally presented word. It is important to notice that in neither form of presentation is the spelling information retained in short-term memory during the interpolation. This is why we do not find that an interpolated task in the same modality as the stimulus word causes more spelling difficulty than an interpolated task in another modality. Experiments which do show same-modality interference require the subject to be actively maintaining stimulus material in the same buffer that is utilized for the interpolated task. But a spelling word is not being maintained in such a buffer. The word is in long-term memory, and only a clue, or a pointer, is being held during the delay interval in short-term memory.

When a nonsense word is presented, whether visually or aurally, there is (by definition) no single familiar chunk that can be stored in short-term memory to reference it. Any information that is to be employed after performance of the interpolated task in order to spell it must be held in short-term memory during that task. According to Waugh and Norman (1965), there is room in short-term memory for only one or two chunks—that is, references to only one or two letters or phonemes. Of course, a subject who notices a resemblance of the "word" to an English word or spelling pattern that is familiar to him may be able to store a little more and/or apply some of his stored spelling rules inventively. In any event, the subject in the nonsense condition will be able to spell very few five-letter or seven-letter "words" correctly.

If the foregoing arguments hold, sight-spelling of meaningful words should not only be more accurate, but should also be somewhat slower than sound-spelling. The reason is that the visual-to-auditory pathway is slightly longer than the direct auditory route activated in sound-spelling. In Table 5, we have divided response times following visual presentation from those following auditory presentation. These are error times only. When the children knew the correct spelling of a word, they wrote it so quickly that our hand-operated stopwatches could not detect a modality difference.



We believe one exists, however, and hope to detect it in future experiments using more precise laboratory apparatus.

Modality differences in the error data are apparent even by our gross timing procedures, as Table 5 shows. A type of replication can be produced by dividing errors into three categories, thereby showing that the response times following visual word presentation are repeatedly longer than the response times following auditory presentation. The fact that this is true even for nonsense words indicates that some aspects of nonsense-word spelling and meaningful spelling are held in common, although the fact that nonsense-word spelling is always longer and more erroneous indicates that other aspects are not held in common.

The results of Experiment II have been replicated with adults, using different, and more difficult, spelling words. Pilot data (Christie, Note 2) have now been collected on college students spelling the following words: sergeant, hushion, idyllic, calyculus, klystron, cicisbeo, wassail, enfeoff, estafette, abomasum, aculeate, phlegm, asyndetron, and amanuensis. These were presented either visually or aurally, using the procedures developed for Experiment II. The interpolated task was writing down sets of random digits, also presented either visually or aurally, for 30 sec. The percentage correct for both adults and children (Experiment II) is shown in Table 6. It is clear that the main effect of visual presentation remains strong [ $F(1,224) = 91.3$ ]. There was no main effect of interpolation modality in Christie's data, and no significant interaction between the stimulus and interpolation conditions.

An interaction between short-term and long-term memory like that of the spelling experiments has also been noted independently by Charness (1974), who studied retention of information about a briefly viewed chess position after performance of a 30-sec interpolated task. He, too, found levels of retention requiring the assumption that much of the information was already stored in familiar units in

Table 6  
Mean Proportions of Words Spelled Correctly by Children\* and by Adults†

Modality		Children	Adults
Stimulus	Interpolation		
Visual	Visual	.813	.659
Visual	Auditory	.935	.728
Auditory	Visual	.551	.317
Auditory	Auditory	.655	.317
Visual Stimuli		.873	.693
Auditory Stimuli		.603	.317
Visual Interpolation		.682	.488
Auditory Interpolation		.794	.521

\*Experiment II: frame, letter, and meaning conditions averaged.  
†Christie, 1975.

long-term memory, referenced by one or two chunks in short-term memory. He found the additional confirmatory fact that after an interpolated task, subjects needed several seconds' additional time before they could begin reconstructing a position, as compared with the latency when there was no interpolated task. This effect was especially pronounced when the interpolated task was visual, i.e., when it was in the same modality as the chess task.

In our spelling task, total spelling time (we did not measure latency of onset) was increased following an auditory interpolated task. If this time increase is also modality-specific, then the implication is that the spelling program is associated with auditory images, as Figure 2 suggests. Such an association is consistent with data recently reported by Nelson and Warrington (1974). They showed that children who were both poor spellers and poor readers made more phoneme than grapheme errors (our terminology, but their measures were comparable). Children who were poor spellers but good readers made relatively few phoneme errors. Other research in our laboratory has suggested that reading is monitored by the auditory system (Farnham-Diggory & Gregg, 1975). If that system is impaired, then "sounding-out" errors should appear in spelling as well.

Table 5  
Mean Number of Seconds Writing Each Incorrectly Spelled Word in Experiments I and II

Stimuli	Sound-Frame	Error Categories				
		N*	Phoneme N*	Grapheme N*	N*	
Meaningful Words: Data from Experiments I and II Combined						
Visual	17.56	25	14.47	38	9.96	24
Auditory	15.33	30	10.75	100	9.43	95
Nonsense Words: Experiment I Only						
Visual	18.47	40	18.21	63	17.50	18
Auditory	16.63	8	17.36	44	15.54	63

\*Number of words on which each mean is based.

REFERENCE NOTES

1. Simon, D. F. Spelling—a task analysis. Technical Report, University of Pittsburgh. Learning Research and Development Center, February 1974.
2. Christie, J. M. Modality differences in the retention of spelling information. Unpublished manuscript, Carnegie-Mellon University, 1975.

REFERENCES

- CHARNES, N. Memory for chess positions: The effects of interference and input modality. Unpublished doctoral dissertation, Carnegie-Mellon University, 1974.

- FARNHAM-DIGGORY, S., & GREGG, L. W. Short term memory function in young readers. *Journal of Experimental Child Psychology*, 1975, in press.
- GREENE, H. A. *The new Iowa spelling scale*. Iowa City: State University of Iowa, 1954.
- HINTZMAN, D. L., & SUMMERS, J. J. Long term visual traces of visually presented words. *Bulletin of the Psychonomic Society*, 1973, 1, 325-327.
- NELSON, H. E., & WARRINGTON, E. K. Developmental spelling retardation and its relation to other cognitive abilities. *British Journal of Psychology*, 1974, 65, 265-274.
- NORMAN, D. A. *Models of human memory*. New York: Academic Press, 1970.
- PARKINSON, H. R. Short-term memory while shadowing: Multiple-item recall of visually and of aurally presented letters. *Journal of Experimental Psychology*, 1972, 92, 256-265.
- PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, 58, 193-198.
- REED, S. *Pattern recognition and information processing*. New York: Academic Press, 1973.
- SCARBOROUGH, D. L. Stimulus modality effects on forgetting in short-term memory. *Journal of Experimental Psychology*, 1972, 95, 285-289.
- SIMON, D. F., & SIMON, H. A. Alternative uses of phonemic information in spelling. *Review of Educational Research*, 1973, 43, 115-137.
- SERLING, G. The information available in brief visual perceptions. *Psychological Monographs*, 1960, 74, No. 11.
- WAUGH, N. C., & NORMAN, D. A. Primary memory. *Psychological Review*, 1965, 72, 89-104.

## NOTE

1. This word was misheard as "Sara" by almost all the children, and was therefore deleted from data analyses.

(Received for publication October 10, 1974;  
revision received January 9, 1975.)