GEORGE SPERLING

Short-Term Memory, Long-Term Memory, and Scanning in the Processing of Visual Information

Early Experience and Visual Information Processing in Perceptual and Reading Disorders

EDITED BY FRANCIS A. YOUNG / DONALD B. LINDSLEY

Proceedings of a Conference held
October 27–30, 1968, at Lake Mohonk, New York,
in association with the
Committee on Brain Sciences,
Division of Medical Sciences,
National Research Council

Available from
Printing and Publishing Office
National Academy of Sciences
2101 Constitution Avenue
Washington, D.C. 20418

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1970

GEORGE SPERLING

Short-Term Memory, Long-Term Memory, and Scanning in the Processing of Visual Information

A MODEL OF VISUAL-INFORMATION PROCESSING

In reading, as in most visual tasks, the eye gathers information only during the pauses between its quick saccadic movements. The normal input to the visual system is thus a sequence of brief exposures. I would like to propose here a model of the way people process the information they receive in one such exposure. I shall be concerned with the simple situation in which a person is shown briefly an array of letters and then asked to write them and the closely related situation in which he hears spoken letters and is required to write them.

The model shown in Figure 1 summarizes the results of numerous experiments. The squares indicate short-term memories. The first box represents a very-short-term visual memory, which, in the past, I have called visual-information storage.[15] It contains a great deal more information than the subject ultimately will be able to report, but its contents normally fade rapidly, usually within about one fourth of a second. These conclusions are derived from a partial-report procedure: the subject is required to report only a small fraction of the stimulus contents on any trial and does not know in advance which aspects he will be required to report. The methods and results have been described in detail elsewhere.[2,15] It is easily proved that a great deal of information from a visual stimulus gets into the subject's very-short-term visual memory; the information is lost to recall because later processes are unable to use it.

Ultimately, stimulus letters are "recognized"; that is, the subject says or writes them. He makes an appropriate motor response. In terms of the model, it is useful to distinguish between actually executing the motor response (saying, subvocally rehearsing, or writing a letter) and having decided which response is to be executed. This kind of distinction is most often made in discussing computers, and perhaps the terminology that has been developed to deal with it in that domain will help to clarify it here.

Saying a letter may be conceived of as executing a long program that consists of hundreds of instructions to various muscle groups. Recognizing a letter may be considered as having decided which program to execute. In practice, a program is designated by its location, or address: the address is the location of the first program instruction to be executed. The second short-term memory box in the model designates the recognition buffer-memory. It is a short-term memory for letters that are about to be spoken or rehearsed subvocally, i.e., a memory of the addresses of the programs for saying them.

The kinds of data that require the concept of a recognition buffer-

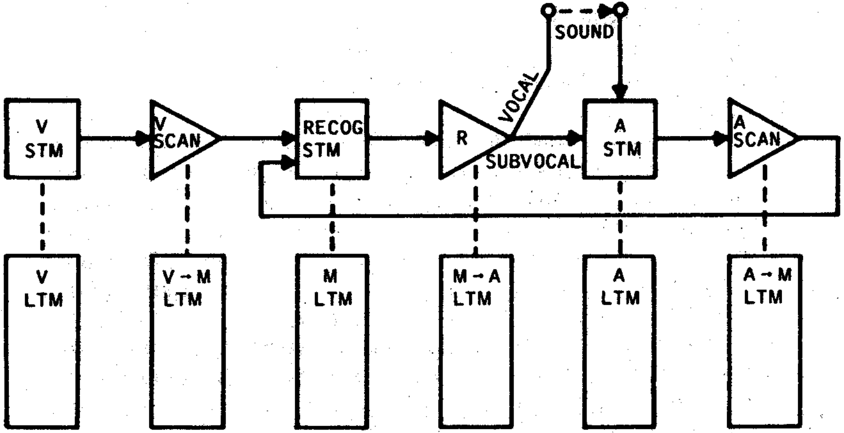

FIGURE 1 Model of visual information processing. Squares indicate short-term memories, rectangles indicate long-term memories, and triangles indicate scan components that transform signals from one modality into another. V, visual; A, auditory; M, motor; R, rehearsal; RECOG, recognition buffer-memory; →, direction of association.

memory have been described.[14] The basic idea is that three or four letters can be recalled from visual presentations even if the effective duration of the presentation—e.g., VIS—is so short that there is not time for the rehearsal of even one letter. The recognition buffer-memory can hold at least three letters (i.e., the addresses of the motor programs for rehearsing the letters) for a period of about 1 sec, until they have been rehearsed.

A scan component is needed to transform the visual information in very-short-term visual memory into the motor-address information of the recognition buffer-memory. The visual scan component is designated by a triangle in Figure 1 to indicate that it is not a memory and that it transforms information from one modality into another. Actually, the visual scan component has at least three distinguishable functions: deciding which areas of the visual field contain information on which further processing should be performed ("prescan"[8]); directing processing capacity to the locations selected by the prescan ("attention"); and converting the visual input from the selected locations into the addresses of motor programs ("scanning").

The maximal rate at which letters are scanned can be measured from visual presentations in which the persistence of the information from an initial letter stimulus is obliterated by a subsequent visual "noise" stimulus. The measured rates are quite high—say, one letter every 10–15 msec, which is equivalent to rates of up to 100 unrelated letters per second [10]

The middle triangle in Figure 1 designates rehearsal. In vocal rehearsal, the motor instructions designated by the recognition buffer-memory are executed, and a spoken letter results. Because it indicates a change of modality or dimension, a triangle is used to designate the rehearsal component; in this case, the transformation is from movements to sound. The sound produced by a vocal rehearsal is heard and remembered in auditory short-term memory.

In principle, although not in detail, the auditory scan is exactly analogous to the visual scan. The auditory scan selects some contents of auditory memory (e.g., the sound representation of one letter) and converts them into the address of a motor program. The address is remembered in the recognition buffer-memory, the program is executed by the rehearsal component, and the sounds are re-entered into auditory memory. By means of this rehearsal loop, information can be retained for a very long time in auditory short-term memory—many times longer than the decay time of the memory itself.

Perhaps in young children and some adults, the output of the rehearsal component must first enter into the outside world as sound before it can enter auditory memory, but most adults seem to have evolved a shortcut, which I have designated "subvocal rehearsal." In subvocal rehearsal, the subvocal output of the rehearsal component is entered into the auditory short-term memory just as though it had been a vocal output; i.e., auditory memory contains a memory of the sound of the letter. The rate of subvocal rehearsal can be measured,[6,10] and it is very interesting to note that it is identical with the rate of vocal rehearsal.

DISTINCTIONS BETWEEN SHORT- AND LONG-TERM MEMORY

Neural Distinctions

A short-term memory is a patch of neural tissue that is used over and over again for every appropriate input to the modality. For example, the retina undoubtedly serves as a short-term memory; a particular neuron in the retina might, by appropriate stimulus positioning, be activated by every letter that could be presented. But I suggest that the neurons involved in long-term memory are extremely specialized and are active only when their key is found. This does not mean that only one stimulus can activate a neuron in long-term memory, but rather that its range is infinitesimal, compared with the range of possible stimuli.

There is now fairly widespread agreement[1,9,12,18] that short-term memory is short-term not because its neurons remember poorly (although that is probably a factor) but because every new stimulus overwrites its predecessor or at least pushes it away from the fore of memory. Even silence or darkness, the absence of stimulation, is an input to short-term memory that must be recorded and that therefore inevitably drives out the record of previous stimulation.

Structural Distinctions

A short-term memory can be likened to a register in a computer; a long-term memory, to a section of core memory.[13] That is, a short-term memory is complicated and expensive (involving many neurons per unit of information stored), because the information in it is capable of being manipulated in many ways. For example, one bit of information can be compared with another bit of information, can be shifted, and so on. Every operation of this sort requires many connections. In computers, core memory is made as starkly economical as possible. So much is sacrificed to economy that no operation whatever (except perhaps erasure) is possible on the contents of long-term memory before they have been removed to a register. I propose that the same overriding principles that guided the evolution of computers to have a very few (but very intricate) registers and to have a great many (but very simple) core memory cells guided the evolution of nervous systems to have a few intricate short-term memories controlling great masses of long-term memory.

Functional and Behavioral Distinctions

The contents of short-term memory are retrieved by asking for the contents of the particular sensory memory, i.e., by giving the name of the memory. What did I just hear? What did I just see? The contents of long-term memory are retrieved by giving an association, i.e., a complex, highly specific input. For example, I say: "My telephone number is 582–2644. What is my telephone number?" You answer by asking yourself what was the last thing you heard. That it is Sperling's telephone number is irrelevant to the retrieval of the digits. However, if I meet you on the street tomorrow and ask you to repeat my telephone number, no short-term memory could possibly be equal to the job. You would need a memory that could be entered with the name "Sperling" (and perhaps some other concomitant bits of information) and that, when so prodded, would return the correct digits.

SIX LONG-TERM MEMORIES

Each of the active components in the model (Figure 1) is associated with a long-term memory. The long-term memory was constructed by the subject out of his past experience, long before his participation in any of my experiments. The three triangle components each use an intermodality long-term memory. The visual scan is served by an intermodality long-term memory that associates the address of the motor program for saying a letter with the visual features of that letter. The rehearsal component is served by a long-term memory that associates the auditory features of a sound with the motor program for producing that sound. The auditory scan is served by a long-term memory that associates the address of a motor program for producing a sound with the auditory features of that sound.

These intermodality long-term memories represent skills. As children, we learned to imitate sounds that we heard. We learned how to recognize letters, that is, to say the name of a letter when we saw it. Later, we learned how to read without speaking.

Beneath each short-term memory square in Figure 1 is a long-term memory of events within that modality. For example, long-term visual memory might contain the information necessary to recognize a particular face as familiar, even if no name or occasion can be associated with it. A preschool child would recognize some letters as familiar, even if he could not name them. Similarly, we have auditory memories of auditory events. Finally, we have the memory of the motor sequence necessary to say a letter.

The proper development of all six of these long-term memories is a prerequisite for the effective operation of the information-processing system outlined before.

Quantitative theories of short-term recall performance find it necessary to take into account a small amount of information that is getting into long-term memory from each trial and that, when there are repeated trials, significantly affects performance (see especially Atkinson and Shiffrin[1]). Although the experiments I have dealt with probably involve very little long-term memory (because each stimulus is viewed only once), it is obvious that something is entering the various long-term memories, at least occasionally.

I will concentrate now on the two aspects of the model that are of greatest relevance to reading: visual scanning and auditory memory.

VISUAL SCANNING

The Use of Visual Noise to Estimate Processing Rate

Brief visual exposures, by themselves, are useless for determining the rate at which visual information is processed. This is so because stimulus information persists in very-short-term visual memory for some undetermined time after the exposure, for at least 0.1 sec and usually for 0.2 sec or longer. If the duration of visual availability is undetermined, processing rate cannot be determined; duration of visual persistence and processing rate are complexly intermingled.

The way around this difficulty is to follow exposure of the stimulus letters by a "noise" postexposure field (Figure 2). The visual noise that I use looks like scattered bits and pieces of letters, and it effectively obliterates the visual persistence of the stimulus letters. By delaying the onset of the noise postexposure field, we allow the subject more time to scan the letters. Each 10–15 msec of delay enables him ultimately to report one additional letter, up to about three or four letters. This processing rate can be shown to be independent of the number of letters presented and of many other variations in procedure.

Serial or Parallel Processing?

In a brief exposure, are letters scanned one at a time, a new letter in each interval of 10–15 msec, or is information being gathered about several letters simultaneously at an overall rate equivalent to one new letter per 10–15 msec? A positive answer to the first question defines a serial scanning process, and to the second, a parallel process. I will go into greater depth in considering the problem of serial versus parallel processing, because it offers a good illustration of current research in information processing. The nonspecialist reader may have difficulty here, but I hope that he will persevere and obtain at least an appreciation of some contemporary methods and theories and of their potential power for studying the way in which words are read.

METHOD 1

When I first confronted the serial-versus-parallel problem, I sought the answer by examining the rate at which information was acquired about each individual letter in a stimulus instead of looking only at the overall rate.[14] Subjects were presented with five random letters followed, after various intervals, by a noise postexposure field (Figure 2). Their task was to report correctly as many letters as they could, from all the locations. If they processed letters in a purely serial order, I would expect only the letter in the first location to be reported correctly at the briefest exposure; the first and second letter to be reported at longer exposures; then the first, second, and third; and so on. Let p_i be the

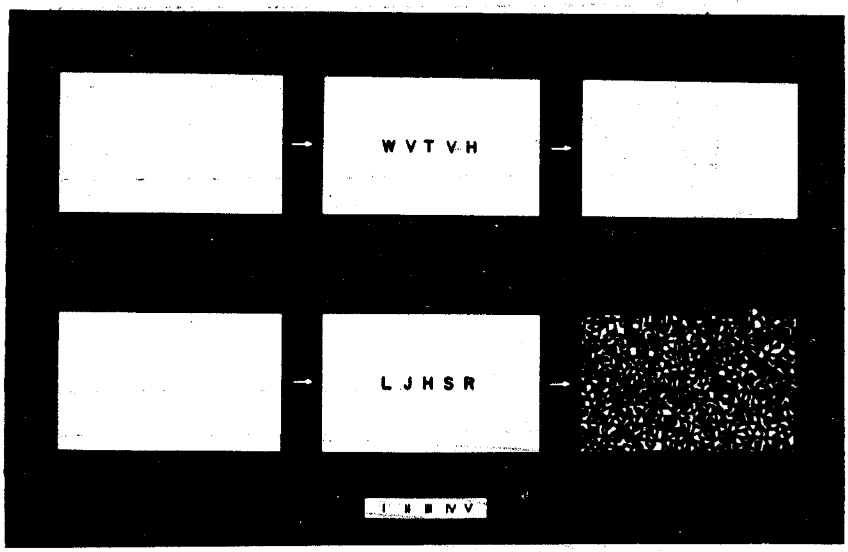

FIGURE 2 A normal tachistoscopic exposure sequence (top) and a postexposure visual noise sequence (bottom).

probability of correctly reporting the letter in the ith location. Considering each of the five letter-locations separately and plotting these p_i's as a function of exposure duration should yield a set of functions like those illustrated in Figure 3a. That is, the p_i functions in Figure 3a would be produced by a serial left-to-right scanning process whose overall theoretical performance best matches the observed performance. The first two letters are scanned quickly, the next two are scanned more slowly, and scanning of the last letter has hardly begun even at the longest exposure.

A purely parallel scanning process, in which information is retrieved at an equal rate from all five locations, would predict identical p_i at all locations (Figure 3b). Because all p_i's are the same, this p_i function also represents the observed overall percentage of correct responses.

The results of an actual test are shown in Figure 3c. The data illustrated are for one subject; tests of other subjects, including myself, yielded basically similar data. The downward concavity of all the observed p_i functions means that information is acquired, at each letter

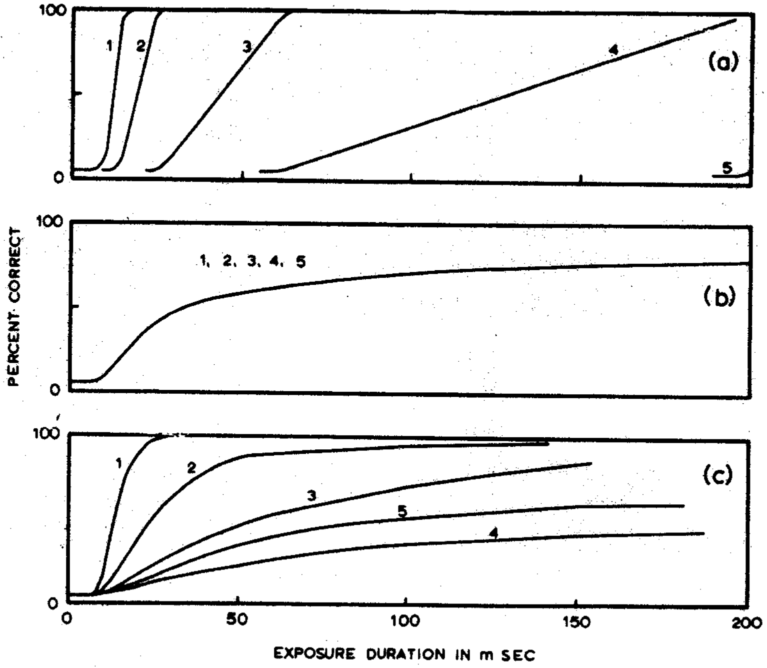

FIGURE 3 Accuracy of report of the letter at each location (1, . . . , 5) of a five-letter stimulus as a function of the exposure duration when exposure of the letters is followed by visual noise. (a) Theoretical data generated by a serial scan process with fixed order of scan. (b) Theoretical data generated by a parallel scan process having the same rate of information acquisition at all five locations. (c) Data of a typical subject (after Sperling[14]). These data are not corrected for chance guessing.

position, most rapidly immediately after the letter stimulus is turned on and that the rate diminishes as the exposure continues.* Information is acquired more rapidly at the first position than at the second, and so on, except that this subject acquired information more rapidly at the fifth position than at the fourth. Other subjects had different idiosyncratic orders.

* Percentage correct is a nonlinear (but monotonic) function of information retrieved. Plotting the results in terms of bits of information retrieved would exaggerate the concavity and strengthen the conclusion.

METHOD 2

Although the interpretation I have just given is stated in terms of parallel processing, one cannot rule out the possibility of some complex form of serial processing. To make a more sensitive test, more intricate stimulus sequences were required. Therefore, I gave up research for a year and worked at programming a computer to display visual stimuli on a cathode-ray oscilloscope.[3] The computer-produced demonstration that provides the strongest evidence of parallel processing is very similar to the procedure just described. Five letters are presented and followed by visual noise. The basic difference is that one of the letters is changed midway during its exposure (Figure 4). When this is done, for example in the fifth location, then almost invariably the first letter that appeared in that location is the one that is reported (that is, if the subject correctly reports anything at all from that location). This result with very brief exposures is just the opposite of the usual result when exposures are long (greater than 50 msec) or no postexposure noise field is used. In the latter circumstances, the second letter that occupies a location is the one that is reported.[2,5,11]

PREDICTIONS OF THE THEORIES OF SERIAL AND OF PARALLEL SCANNING

In a serial process, increasing exposure duration improves performance (increases p_i), because the ith location is more likely to have been scanned during a longer interval. Consider, for example, an exposure duration ΔT_1, which is just long enough so that $p_i = \Delta p$. Now con-

FIGURE 4 A computer-generated stimulus sequence for testing serial versus parallel processing. The initial stimulus is M T K L X. M, T, K, and L persist continuously until the onset of the postexposure visual noise; X is changed to Z in the middle of the exposure interval. Two consecutive noise fields are used to increase the effectiveness of the noise.

sider the additional exposure ΔT_2 that is needed to increase p_i to $2\Delta p$. In serial-scanning theory, an increase of Δp in p_i during ΔT_2 means that as many letter scans are made in ΔT_2 as in ΔT_1. If occasionally the *i*th position is scanned twice during the exposure, then more letter scans must be occurring in ΔT_2, inasmuch as occasionally a letter that was scanned in ΔT_1 will be rescanned in ΔT_2, and that would be a wasted scan. Serial-scanning theories can be characterized as basically "top-heavy." That is, when p_i is large–i.e., near the top of a graph like Figure 3b–then as many or more scanning attempts are needed to raise it by a given amount, Δp, compared with the number when p_i is small.

Parallel-processing theory assumes that information is accumulated continuously. To increase p_i from 0.50 to 0.95, for example, requires less than one bit of information, whereas to increase p_i from 0.05 to 0.50 requires 3.3 bits (when there are 20 equiprobable stimulus letters). This example illustrates a general property of information-gathering systems: the first few bits of information change the probability of being correct only very slightly, and the last few bits cause big changes. Thus, parallel-processing theory is "bottom-heavy." The weighty processing occurs while p_i is small, i.e., near the bottom of Figure 3b.

To relate these theories to data, let us restrict ourselves, for the moment, to locations 3, 4, and 5, and to exposure durations of less than 100 msec. For example, consider an exposure of 50 msec and divide it, conceptually, into two consecutive intervals of 25 msec. Figure 3c shows that there is an equal or greater increase of p_i between 25 and 50 msec than between 0 and 25 msec in these three cases. Suppose now that at location 5 a different letter is presented in each of the two intervals–the experiment described above. According to the serial-scanning theory, an equal or greater amount of scanning occurs in the second interval, and so we would expect the second letter to be reported at least as often as the first letter.

In the parallel-scanning theory, in this instance, about 60% more information accumulates in the first 25 msec than in the second 25 msec, so we would expect the letter from the first interval to be reported more often. For parallel theory to predict quantitatively how much more often the first letter is reported than the second would require additional assumptions.

The experimental result was that the first letter is nearly always reported. We therefore reject the serial-scanning theory and tentatively accept the parallel-scanning theory. In 50 msec, the visual system achieves sufficient information, in parallel, from a letter array to recognize about three letters.

This conclusion is potentially important for understanding the reading of words. It means that the visual system has the capacity to process a word not merely letter by letter or by its overall shape, but as a complex pattern. Whether a word is recognized directly as a visual pattern, or the letters are recognized first and then the letter pattern is recognized as a word, or both processes occur together, we do not yet know. But we do know now that the visual system has the capacity to gather enough information simultaneously—i.e., in parallel—from an array of letters (a word) to identify uniquely most ordinary words.

Extremely Rapid Visual Search in a Continuous Task

The experiments described above measured visual scanning speeds from single exposures only—that is, the speeds achieved in single bursts of scanning. Could subjects maintain the same high scanning speed in a continuous search task? The following experiment was devised to test this possibility. A computer[3] generates arrays of random letters and displays them on the cathode-ray oscilloscope. Figure 5 shows a sequence of 3 X 3 arrays. All the arrays except one consist entirely of random letters; the critical array contains the numeral "2" in a randomly selected location. The subject does not know in advance of the trial which array in the sequence will be the critical one, nor in which location the critical character will occur. His task is to look at the whole sequence of arrays and to say at which location the critical character has occurred. From the proportion of times the subject is

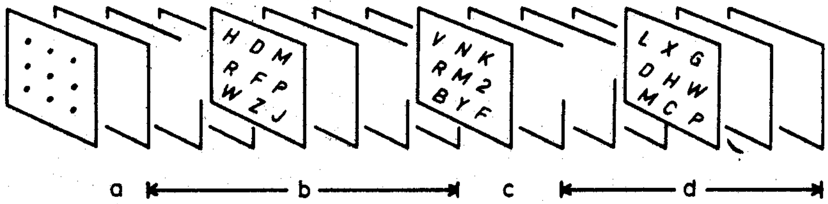

FIGURE 5 Diagram of the stimulus sequence in the sequential search procedure. a, fixation field. b, 6 to 12 letter arrays (randomly determined). c, the critical array, in this instance containing a "2" in the middle-right location. d, 12 more letter arrays.

able to make the correct response, we can deduce the speed with which he is able to scan characters to determine whether each is a "2." We have also trained a subject to detect the occurrence of any numeral among letters. The discrimination of an unknown one of ten numerals takes only slightly longer than the discrimination of a known single numeral.

We[16] have studied arrays containing from 1 to 25 letters, and presented new arrays at rates of 3 to 200 per second. We have not yet completed all these experiments, but the main results are already clear.

Subjects achieve the same high scanning speeds in the continuous-search procedure as were previously demonstrated for single bursts, 10–15 msec/letter. The highest scanning speeds are achieved at presentation rates of about 25 arrays per second with stimuli containing nine or more letters. Under these conditions, the fastest subject has broken solidly through the 10-msec barrier; he can scan characters for the absence of the numeral "5" faster than one letter per 8 msec. When nine-letter arrays are presented at a rate of 25 arrays per second (40 msec/array) he can identify the location of the critical character correctly about 70% of the time. That means that he is effectively monitoring five of the nine locations.* In terms of the parallel-scanning theory, this subject can process a fresh batch of five letters every 40 msec.

When the presentation rate is lowered, response accuracy improves, indicating that additional locations are being scanned. For example, my fast subject scans the equivalent of about 16 locations from a 25-letter array when new arrays are presented every 160 msec. His scanning speed goes down to about one letter per 10 msec at this rate, indicating that locations outside the most favored six are scanned more slowly. Sixteen positions are the maximum that he can scan in a brief exposure; lowering the rate does not improve his response accuracy. A more typical observer can scan three locations in 40 msec and a maximum of 10 locations in a single exposure.

In conception, these search experiments follow the pioneering work of Ulric Neisser,[7] who was the first to study rapid scanning of this kind. His subjects searched long lists for the presence of a critical item and signaled when they had found it. The important difference between our procedures is not that I use a detection method and he a reaction

***The estimate of the number of locations monitored depends somewhat on the guessing strategy that the subject is assumed to be using when he has not seen the critical character. If he could use absolutely the most efficient strategy, he could achieve a probability of being correct of 0.7 even when he monitored only 5.3 locations.**

method, but that in Neisser's experiments the sequence of visual inputs is controlled by the subject's own eye movement, and in my experiments, by a computer. The optimal scanning rate in the searching for a "2" or a "5" occurs at presentation rates that are five times higher than the rate of eye movements. When the presentation rate of stimuli is lowered so that it is comparable with that of eye movements (e.g. 200–250 msec), then the processing capacity is virtually idle for the second half of the interval; it has done all or nearly all that it can do in the first half. With more complicated processing tasks, of course, processing times would be longer and the rate of eye movements might not be the limiting factor.

Although it is technically very difficult to implement, the method of searching sequentially presented displays is most promising for estimating processing times and will yield much of importance for reading. It already has provided one nontrivial conclusion: In simple search tasks, the limiting factor in performance is the rate at which eye movements can be made, and not the rate at which information can be processed.

AUDITORY SHORT-TERM MEMORY

Auditory Memory in Visual-Recall Tasks

I claim that the same factors limit recall of letters from brief visual exposures (assuming that the letters are clearly visible) and from auditory presentations, to such an extent that visual recall can be predicted from auditory recall.[17]

The original evidence of auditory components in visual-recall tasks was introspective (all subjects said they rehearsed subvocally) and indirect (subjects did not begin writing until a second or more after the exposure and their visual memory had decayed by then, so auditory memory was the only logical alternative).[10] The observation[15] that subjects suffered auditory confusion in visual recall (for example, D and 2 for T) was promising but not powerful. The powerful evidence comes from the measurement of "AS deficits," a technique that was introduced independently and almost simultaneously in three laboratories by Conrad, Wickelgren, and me (see Sperling and Speelman[17]), although it could and should have been invented 100 years earlier.

An AS deficit is defined as the decrement in performance caused by replacing a stimulus composed of acoustically different letters (for example, F, H, Q, and Y) with acoustically similar (AS) letters (for example, B, C, D, and G). The deficit technique can be applied to other dimensions, such as visual similarity, semantic similarity, and pronounceability. The main finding that concerns us here is that, in the usual test of visual recall, visual-similarity deficits are small, whereas AS deficits are large.[4] That auditory similarity should be a significant factor even in a task that involves only looking at letters and writing them—and never any overt auditory representation—is *prima facie* evidence of a role for auditory memory in visual-recall tasks.

To determine quantitatively how much of the memory load in visual-recall tasks is carried by auditory memory is more difficult. However, we[17] have been able to predict AS deficits in visual-recall tasks, in which subjects viewed a dozen letters exposed simultaneously, from the AS deficits of the same subjects in auditory tasks, in which they heard spoken lists of letters and were required to recall them. We could make these predictions from lists spoken at either one or two letters per second but not from lists spoken at rates of four letters per second. The rate of silent rehearsal was previously estimated to be three letters per second.[6] This rate seems to be critical for auditory presentations of random letters; at higher rates, recall performance deteriorates rapidly. I would conclude, pending evidence to the contrary, that the same factors limit recall from simultaneous visual presentations and limit recall of auditory sequences spoken at rates lower than four letters per second.

A Phonemic Model of Short-Term Auditory Memory

The results of 38 experimental conditions in which Mrs. Speelman and I measured recall of auditory stimuli could be predicted quite accurately from rules based on a phonemic model of short-term auditory memory.[17] (The predictions accounted for 0.96 of the variance of the data.) The phonemic model assumes (1) that individual phonemes are retained and forgotten independently in auditory memory; (2) that, when some of the constituent phonemes of a letter are forgotten, the letter is reconstructed as well as possible on the basis of the remaining phonemes; and (3) that, when the remaining phonemes do not suffice to identify the letter uniquely, a choice is made from among the most probable alternatives. According to this theory, the reason that stimuli composed of letters chosen from AS alphabets are poorly recalled is that they contain phonemes that do not help to discriminate among alternative members of the alphabets. For example, in the alphabet consisting of B, C, D, and G, the phoneme e is useless for discriminating among members, and retention of this phoneme in memory is a waste of space—a precisely predictable waste.

It is reasonable to call the memory into which an unrehearsed auditory stimulus enters an "auditory memory." Because the predictions of the model apply equally well to conditions in which there is little subvocal rehearsal and conditions in which there is a great deal of subvocal rehearsal, there is no need to postulate different memories for rehearsed and unrehearsed material. Finally, because the same generalizations govern recall of visual stimuli, there is no need to postulate a different memory for visual recall.

I should add that a really satisfactory paradigm for differentiating between the recognition buffer-memory and the auditory short-term memory has not yet been discovered. Therefore, when I say "auditory memory," I have to include in it the contribution of the rehearsal buffer-memory. That is not much of a complication, because, if the contribution of the recognition buffer is small, then it does not matter much, and if its contribution is large, then we can say that it must be very much like an auditory memory, in that the phonemic model (of auditory memory) accounts for so much of the evidence.

RECAPITULATION

A model of the processing of information from an array of letters has been proposed. It consists of the following components: a very-short-term, very-high-capacity visual memory; a visual scan component that converts the representation of a letter in visual memory into the address of the motor program for rehearsing the letter; a short-term memory for this address (recognition buffer-memory); a rehearsal component that converts the subvocal rehearsal into an auditory representation; an auditory short-term memory for the sound of the letter; and an auditory scan component that converts the auditory representation into the address of the motor program for rehearsing the letter.

Neural, functional, and behavioral criteria for distinguishing between short-term and long-term memory have also been proposed. A short-term memory is made up of neurons that are used over and over again by all inputs to the modality; complicated functions can be carried out on the contents of the memory; to retrieve the contents of memory requires knowledge only of the memory's name (i.e., the modality being served). The neurons that form a long-term memory are activated only by very specific inputs; no functions are carried out directly on the contents of memory; and the contents of memory can be retrieved only by means of very specific "associations." The components of the processing model are served by six kinds of long-term memory: visual, auditory, and motor long-term memories; and visual–motor, auditory–motor, and motor–auditory association long-term memories.

Experiments with visual postexposure noise fields are interpreted to mean that information is gathered simultaneously–i.e., in parallel–from three or more letter locations at an initial rate of one letter per 10–15 msec. The visual system thus has, in principle, the capacity to analyze a word not letter by letter nor by overall shape, but from information gathered, in parallel, from its component letters.

In the sequential-search procedure, a subject searches a computer-produced sequence of letter arrays for a character at an unknown location in one of them. The highest processing rate occurs when a new array occurs every 40 msec. This maximal rate of 25 arrays per second is 5 times the rate of eye movements. Lowering the sequence rate to the rate of eye movements grossly impairs search efficiency. The best subject was able to scan five locations every 40 msec and a maximum of about 16 locations (achieved in 160 msec) in a single brief exposure. It is concluded that, in simple visual-search tasks, the rate of eye movement will be a limiting factor in search rate.

The recall of visually presented arrays of letters is shown to suffer in a predictable way when acoustically similar letters (for instance, B, C, D, and G) are used. By comparing the recall of visually presented arrays with the recall of auditory letter sequences, it is concluded that visual letters are rehearsed at fewer than four letters per second (probably three per second) and that the rehearsal is stored in auditory short-term memory. Even when letter arrays are presented visually and are reported by writing (never overtly represented in an auditory mode), they are remembered in auditory short-term memory, as if they had been presented acoustically.

In this brief account, I have not considered how eye movements are controlled, how information from successive eye movements is integrated, how long-term memories are formed, or how subjects deal with words and bigger units of meaningful materials. These problems are relevant and important for the study of visual-information processing; some are considered elsewhere in these proceedings, but most, unfortunately, are far from solution.

REFERENCES

1. Atkinson, R. C., and R. M. Shiffrin. Human memory: a proposed control system and its control processes, pp. 89–195. In K. W. Spence and J. T. Spence, Eds. The Psychology of Learning and Motivation. Vol. I. Advances in Research and Theory. New York: Academic Press, 1967. 381 pp.
2. Averbach, E., and G. Sperling. Short term storage of information in vision, pp. 196–211. In C. Cherry, Ed. Information Theory. Washington, D.C.: Butterworth Inc., 1961. 476 pp.
3. Budiansky, J., and G. Sperling. GSLetters. A general purpose system for producing visual displays in real time and for running psychological experiments on the DDP24 computer. Bell Telephone Laboratories Technical Memorandum, 1969. Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.
4. Cimbalo, R. S., and K. R. Laughery. Short-term memory: effects of auditory and visual similarity. Psychon. Sci. 8:57–58, 1967.
5. Greenberg, M., M. S. Helfer, and M. S. Mayzner. Information processing of letter and word pairs as a function of on and off times. Perception and Psychophysics 4:357–360, 1968.
6. Landauer, T. K. Rate of implicit speech. Percept. Motor Skills 15:646, 1962.
7. Neisser, U. Cognitive Psychology. New York: Appleton-Century-Crofts, Inc. 1967. 351 pp.
8. Neisser, U. Preattentive and focal processes in perception. Invited address, 76th Annual Convention of the American Psychological Association. 1968.
9. Postman, L. Short-term memory and incidental learning, pp. 145–201. In A. W. Melton, Ed. Categories of Human Learning. New York: Academic Press, 1964. 356 pp.
10. Sperling, G. A model for visual memory tasks. Hum. Factors 5:19–31, 1963.
11. Sperling, G. Information retrieval from two rapidly consecutive stimuli: a new analysis. Perception in Psychophysics. ~~(in press)~~ 9: 89-91, 1971.
12. Sperling, G. Phonemic model for short-term auditory memory. Proc. Amer. Psychol. Assoc. 4:63–64, 1968.
13. Sperling, G. Structural factors in models of memory. Acta Psychol. (in press)
14. Sperling, G. Successive approximations to a model for short term memory. Acta Psychol. 27:285–292, 1967.
15. Sperling, G. The information available in brief visual presentation. Psychol. Monogr. 74:1–29, 1960.
16. Sperling, G., J. Budiansky, G. J. Spivak, and M. C. Johnson. Extremely rapid visual search. Bell Telephone Laboratories Technical Memorandum, 1970. Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.
17. Sperling, G., and R. G. Speelman. Acoustic similarity and auditory short-term memory: experiments and a model, pp. 151–202. In D. A. Norman, Ed. Models of Human Memory. New York: Academic Press, 1970.
18. Waugh, N. C., and D. A. Norman. Primary memory. Psychol. Rev. 72:89–104, 1965.

DISCUSSION

DR. KAGAN: Can I recall something you said a few minutes ago only because I have been rehearsing it? Would it not be stored in long-term memory?

DR. SPERLING: Certainly. Even very brief events often leave lasting memories; I wish I knew more about how and why. The stimulus materials in the experiments I have been discussing are random letters and numerals; they almost never get into long-term memory in just one trial. They can be recalled accurately for only a few seconds. To recall them after intervals of, say, 10 sec, a subject must rehearse them vocally or subvocally and must not be forced to accept any new information into his short-term memory. If either of these conditions is violated, the stimulus is forgotten.

Repeated rehearsal not only maintains the stimulus in short-term memory, but helps it to get into long-term memory. We do not know whether it is the act of rehearsing itself that is responsible, or whether it is merely that the longer a stimulus resides in short-term memory the likelier it is to enter long-term memory. To reiterate, the essence of short-term memory is that the same patch of neural tissue is used over and over again by new inputs. Obviously, this same tissue cannot also serve as a long-term memory.

DR. ULLMAN: Is the use of short-term memory a prerequisite for the formation of long-term memories?

DR. SPERLING: I would say that visual inputs pass through visual short-term memory, and auditory inputs pass through auditory short-term memory. Given the complexity of long-term memory, I would be rash to venture beyond that simple statement.

DR. SCHUBERT: The leaders in the field of reading would have us believe that some children are visually minded and some children are kinesthetically minded. When you say that your subjects rehearse subvocally and you relate their performance to this kind of rehearsal, are you referring in particular to kinesthetically minded subjects?

DR. SPERLING: No. What I am saying is that, in the particular recall tasks that we have devised with random-letter stimulus materials, auditory memory is so much more effective than visual that we barely detect an effect of visual memory. If we were dealing with words and language, or with pictures, it might be quite different. Incidentally, Dr. Michael Siegal is using our acoustically similar stimuli for memory tests on children with eidetic imagery, and finds that even these extremely visually competent subjects do not behave differently from subjects *on these tasks.*

DR. HOCHBERG: Can you predict one kind of memory from the other?

DR. SPERLING: No. I did not say that I could predict the capacity of a subject's visual memory from the capacity of his auditory memory, but rather that I could predict a subject's performance on the recall of visual stimuli from his performance on the recall of auditory stimuli. The reason is that the stimuli that Mrs. Speelman and I used are remembered in auditory memory even ~~if~~ [illegible] they are presented visually. That is, when we make this assumption, we can predict performance.

I do not wish to be put into the position of saying that there is no visual memory; there certainly is. But except for the very-short-term visual memory, visual memory seems to be basically unadapted to recall, and so we do not find much evidence of it in recall tests. To find out about short-term visual memory, or perhaps intermediate-term visual memory, we have to use recognition procedures. Even that is not sufficient in itself. ~~If~~ Where efficient verbal codes exist, they will be remembered in auditory memory and in other memories and thereby override the visual phenomena that we are trying to measure. The stimuli to be recognized visually have to be made nonverbalizable. Or they have to be so constructed that a verbal description of them would be so inefficient that subjects would not be tempted to try it or, if they did, it would not aid them. I use a computer to generate visual stimuli and, with small modifications in the program (occasionally unintentional), it produces good characters for a recognition experiment. These are made of basically the same segments as letters, but joined in different ways. They look like elements from an unfamiliar Eastern scrawl (see Figure 4). The computer produces an almost limitless variety of different characters, so that none of them becomes familiar.

In our tests, we show the subject a stimulus twice, with an interval of a few milliseconds to 16 sec between the two presentations. The stimulus is composed of six or ten of these characters. In the second presentation, one of the characters is altered, and the subject's task is to say which character. From the accuracy of his response, we deduce how many characters he is remembering correctly. In preliminary experiments with this method, we again found the very-short-term, very-high-capacity visual memory. Beyond the first quarter of a second, performance was disappointingly poor. Subjects are able to retain enough information about only two or three characters to recognize that they have been changed. However, the time constant of forgetting was, surprisingly, so long that I could not estimate it properly. These experiments, like most others that have been used to investigate visual memory, have their problems (Hochberg, J., in R. N. Haber, Ed. Contemporary Research and Theory in Visual Perception. New York: Holt, Rinehart and Winston, Inc., 1968. pp. 309-331), but I cite them to show that measurements of short-term visual memory are being made (see also Shepard, R. N., J. Verbal Learning Verbal Behav. 6:156-163, 1967).

DR. SHANKWEILER: It seems to me that you should not attribute your findings to auditory memory. I suggest that subjects are coding into language.

DR. SPERLING: The kind of auditory memory I have been discussing is basically very simple, although some of its properties are very complex and may surprise us. If you had available a pile of neurons, I could tell you how to connect them to make an auditory memory. In conception, it is very much like a sound spectrograph; the same basic construction would serve either a mouse or a man. It is a memory for sounds; let us reach semantic agreement on that point.

To construct a memory that remembers not merely sounds but linguistic units would be incredibly more difficult. I should think that one would not even undertake it unless one already had a very good auditory memory for sounds. But that is a philosophic answer. That I like to keep things simple does not mean that nature does. In fact, your hypothesis about linguistic memory probably could be formulated specifically in terms of an alternative model and subjected to experimental investigation. I invite you to do so.

Sperling, G. Short-term memory, long term memory, and scanning in the processing of visual information. In F. A. Young and D. B. Lindsley (Eds), Early Experience and Visual Information Processing in Perceptual and Reading Disorders. Washington, D.C.: National Academy of Sciences, 1970. Pp. 198-218.