

SUCCESSIVE APPROXIMATIONS TO A MODEL FOR SHORT TERM MEMORY¹

G. SPERLING²

Bell Telephone Laboratories, Murray Hill, New Jersey, USA

ABSTRACT

Experimental data are considered from a simple task in which an observer looks at letters and then writes them down. Three models are proposed. Model 1 consists of only two components: a visual memory for the letters and a motor translation component to enable copying a visual memory onto paper. Model 1 is inadequate because the visual image is shown not to persist until the time of reproduction. Model 2 corrects this deficiency by incorporating the possibility of subvocal rehearsal of the stimulus letters and an auditory memory for the rehearsal. However, Model 2 cannot account for performance with extremely short duration images because of the limit on the maximum rehearsal rate. The critical improvement in Model 3 is a more detailed specification of scanning, recognition and rehearsal, including a form of memory which is inherent in the process of recognition itself. Model 3 accounts for these data and incidentally gives rise to some interesting inferences about the nature of consciousness.

1. INTRODUCTION

One major difficulty in studying human memory is that we have not yet learned how to obtain systematic physiological information. This means that the only technique available for the study of human memory is to present the subject with a variety of memory tasks, and then to record his actions and the accuracy of his performance. From these observations, we try to abstract the functions or operations that the subject performs on the to-be-remembered stimulus in order to produce the observed performance.

Because of the complexity and subtlety of human behavior, it seemed desirable to us to confine ourselves initially to a simple memory task. A subject looks at a row of random letters and then writes them down. If this situation were understood, perhaps the principles could be generalized to more complex tasks.

¹ Originally presented at the symposium on 'Memory and Action', XVIII International Congress of Psychology, Moscow, 1966.

² I use the first person plural to refer at various times to Mrs. Susan A. Speeth, Mrs. Mary W. Helms and Dr. Roseanne J. Speelman, whose able assistance I wish to acknowledge thereby.

2. MODELS

2.1. Model 1

When a row of letters is exposed briefly, i.e., for 1/20th sec, an adult subject can reproduce about 4 or 5 of the letters (CATTELL, 1885). The simplest model for the action of reproducing visually presented letters might be organized into two main components: (1) a visual memory containing the letters (called visual information storage) and (2) a translation component, which can translate a visual image of the letters into a series of motor actions; namely, copying the letters onto a piece of paper (fig. 1). The

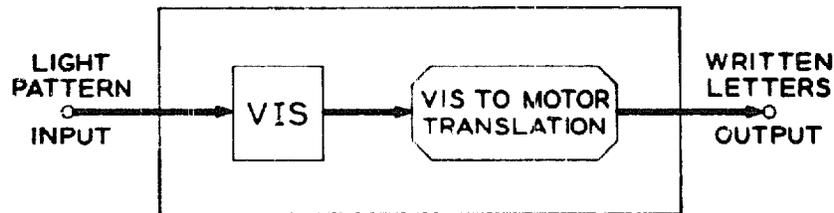


Fig. 1. Model 1. The large box represents the subject. Arrows indicate the direction of information flow. The components are visual information storage (VIS) and a translator. The translator converts an input (the memory of a letter) into an output (a series of motor actions) which result in a written representation of the letter.

limited memory span of the subject might be represented in the model by progressive deterioration—a fading into illegibility—of the contents of visual storage. While the subject is writing, the contents of his visual memory are decaying, so that when he finally comes to write the fifth or sixth letter his visual memory of the stimulus no longer is legible.

Without elaborating further on the difficulties of Model 1, we can reject it immediately for one basic reason: before the subject begins to write the letters, his visual image of the letters has already disappeared. (The measurement of the duration of visual storage is described below.) Having shown that letters are not stored visually until they are reproduced, we must now determine the form in which they are stored.

2.2. Model 2

Occasionally a subject, when he is writing down letters, can be heard to mumble the letters as he is writing them. His tendency to say the letters aloud can be emphasized by playing loud noise into his ears. Noise itself does not seem to alter performance in any other significant way. We have used this technique, together with a microphone placed near the subject's mouth, to record the actual letters the subject is saying. We also recorded automatically whenever the subject was writing. The most interesting results with this technique are obtained when the subject is required to wait (e.g.,

for 20 sec) after the stimulus exposure before writing the letters. He repeats (rehearses) the entire letter sequence several times with a pause between each repetition during the interval. Then, at the time of writing each letter, he also may speak it simultaneously.

Rehearsal suggests an obvious memory mechanism. The subject says a letter, hears himself saying it, and then remembers the auditory image. As the auditory image fades, he repeats it to refresh it. Most of our subjects do not vocalize during recall, but they all concur in stating that they rehearse subvocally. Therefore, we assume that the sound-image of a letter can enter auditory memory directly from subvocal rehearsal without the necessity of actually being converted into sound and passing into the external world. These relations are illustrated in fig. 2.

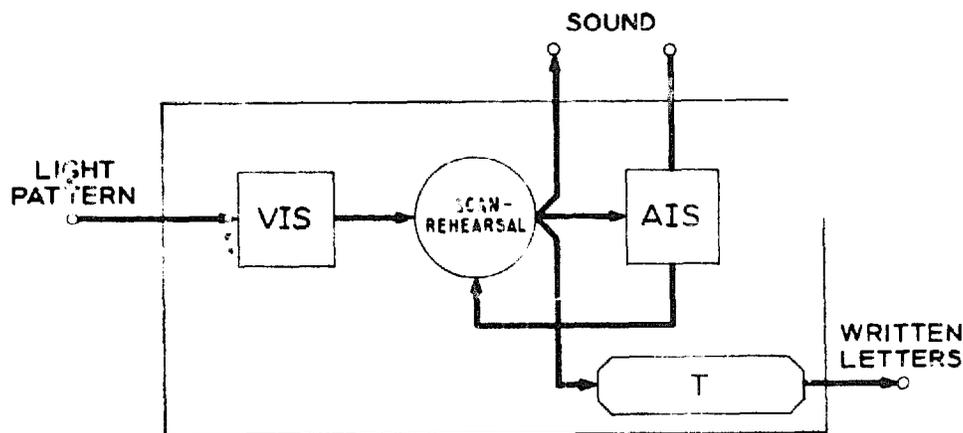


Fig. 2. Model 2. VIS = visual information storage, AIS = auditory information storage, T = translator.

The auditory nature of subvocal rehearsal can be emphasized by playing distracting speech into one's ears during rehearsal. The speech seems to emanate from one set of locations in space (the ears) while one's rehearsal is heard as an internal voice speaking from the center of the head. External sound also can be used as a clock against which to measure the rate of subvocal rehearsal. Another method of measuring the rate of subvocal rehearsal is to ask subjects to rehearse a sequence of letters subvocally 10 times and to signal when through. This may be compared to a vocal rehearsal of the same sequence. All these indirect measures of the rate of subvocal rehearsal indicate that, while it may be slightly faster than vocal rehearsal, it is basically the same process (cf. LANDAUER, 1962). The maximum possible rate is about 6 letters per second but, in memory experiments, maximum rates of about 3 letters per second are more typical.

The existence of auditory memory in visual reproduction tasks also may be inferred from the deterioration in performance which occurs when the

stimulus letters sound alike (B, C, D, etc.). We have studied a large variety of tasks in which stimuli were presented visually or auditorily and found almost the same rule to apply to both modalities of presentation. When the memory load is small (about 2.5 letters in an auditory task, 3 letters in a visual task) it makes little difference to performance whether the stimulus letters sound alike or sound different. Additional letters beyond the minimal number are remembered only about half as well when they sound alike as when they sound different. This dependence of performance on the sound of letters—even in a task which nominally involves only looking and writing—is of practical as well as of theoretical importance (CONRAD, 1963; SPERLING, 1963).

According to Model 2, stimulus letters first are retained in visual storage. They are rehearsed, one at a time (i.e., converted from a visual to an auditory form), and then remembered in auditory storage. Subsequently they may be rehearsed again and again as required until they are written down. The limits on performance may arise either from the limited duration visual storage (so that some letters decay before they can be rehearsed) or from the limited capacity of the rehearsal-auditory storage loop, depending on the stimulating conditions.

Attractive as Model 2 seems, it is inadequate for the following reason: it is possible to generate an image in visual storage which has a duration of definitely less than .1 sec and from which 3 letters can be reported. This would require a rehearsal rate of over 30 letters per second, which clearly is completely beyond the capabilities of the rehearsal processes described for Model 2. Before considering Model 3, we need to examine in more detail some properties of visual information storage.

2.3. *Short duration visual images*

When the contents of an image in visual storage exceed four or five items, they can be measured by a sampling technique which requires the subject to report only a part of the contents (SPERLING, 1960). For example, by this technique it was shown that the visual image induced by a 1/20th sec exposure may contain as many as 18 unrelated letters, and that as many as 10 items may still remain 2 sec after the exposure. The visual image of shortest duration that we have measured by this technique was produced by a stimulus exposure of 1/20th sec, preceded and followed by bright white fields. Immediately after the exposure, 14 letters were contained in visual storage; within 1/2 sec they had vanished (AVERBACH and SPERLING, 1960). To produce and to measure really short duration images, however, different methods are required.

In a letter-noise stimulus sequence, a second, interfering, stimulus (visual 'noise') is exposed immediately on termination of the letter stimulus. The duration of the letter images can be estimated by comparing them to an

auditory signal. Two different methods were used. In the first method two clicks were produced at the ears of the subject. He then adjusted the interval between the clicks until the auditory interval was judged equal to the visual duration. In the second method, the subject heard only one click at a time. He adjusted this click to occur so that it coincided subjectively with the onset of the visual image. After this judgment was complete, he made

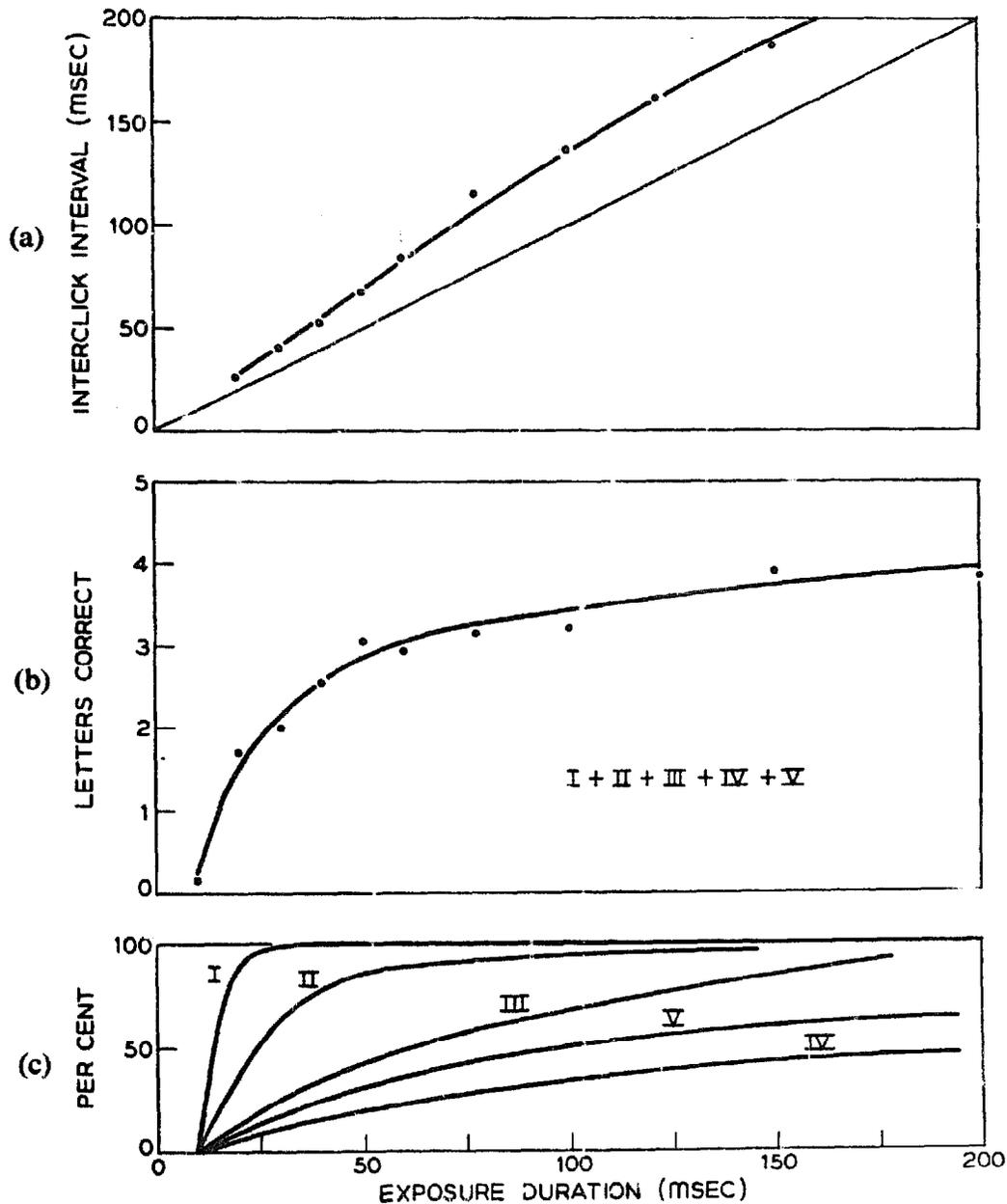


Fig. 3. (a) The apparent duration of the letters in a letter-noise presentation. Abscissa is the exposure duration of the letters. Ordinate is the duration of an interval between clicks which was judged equal to the visual letters. Data for one typical subject. (b) The total number of letters reported correctly. Five letters were presented, one in each location, I to V. (c) The percent of letters reported correctly, shown individually for each location, I to V.

another adjustment of the click to coincide with the termination of the visual image. The measured interval between clicks—taken to be the duration of the visual image—was the same by both methods. The apparent image duration of the letters in a letter-noise sequence is zero for extremely brief exposures (e.g., less than 10 msec) and then increases linearly with increasing exposure duration for durations exceeding about 20 msec (fig. 3a).

When stimuli of 5 letters, followed by noise, are exposed for various durations, the accuracy of report increases with exposure duration as shown in fig. 3b. The most interesting aspect of these data is revealed by analyzing separately the accuracy of report at each of the 5 locations (fig. 3c). The accuracy of report at each location reported increases continuously as a function of exposure duration. For this subject, the order of the successive locations which are reported correctly is generally left-to-right (I to V), except that location V is reported correctly at shorter exposures than location IV. Other subjects have different idiosyncratic orders, e.g., I, V, III, II, IV. By definition, in a purely serial process the n^{th} location is not reported better than chance until the exposure duration at which the $n-1^{\text{th}}$ location is reported with maximum accuracy is exceeded. The observation that all locations begin to be reported at better than chance levels even at the briefest exposures, may be interpreted as evidence of an essentially parallel process for letter-recognition.³ This process gives the illusion of being serial because the different locations mature at different rates (cf. GLEZER and NEVSKAIA, 1964; SPERLING, 1963). These findings are taken into account in Model 3 (fig. 4).

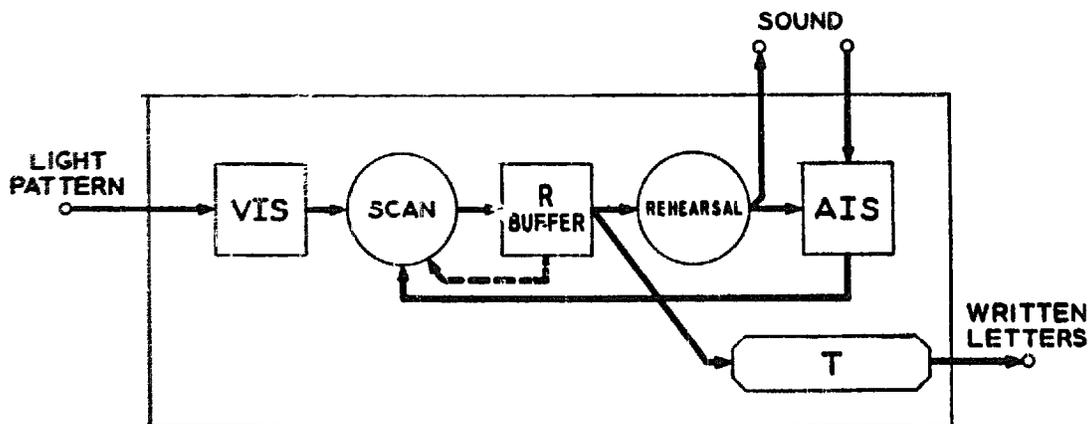


Fig. 4. Model 3. (See fig. 2) R-buffer = recognition buffer-memory.

³ An alternative interpretation is that the scan is serial but the order varies from trial to trial and/or there is great variability in the processing time per item. We tentatively consider that interpretation to be unlikely because (a) the modal scan pattern is highly repeatable from session to session and (b) parallel processing of other aspects of the stimulus is occurring, e.g., of its orientation, overall length, brightness, etc.

2.4. *Model 3*

In Model 3, the scan-rehearsal component of Model 2 is subdivided into three separate components. The first of these is a scan component which determines—within a limited range—the sequence of locations from which information is entered into subsequent components. The extent to which the subject can vary his order of scanning is a current research problem. In very brief exposures, the variation in scanning may be limited to changing the rate of acquisition at different locations—information processing beginning simultaneously at all locations. On the other hand, the overall rate of information flow through the scanner must be limited.

The second new component is the recognition buffer-memory. It converts the visual image of a letter provided by the scanner into a 'program of motor-instructions', and stores these instructions. This program of motor instructions, when it is executed by the rehearsal component, constitutes rehearsal. The important idea embodied in the recognition buffer-memory is that the program of motor-instructions for a rehearsal can be set up in a very short time (e.g., 50 msec for 3 letters) compared to the time necessary to execute it (e.g., 500 msec for 3 letters).

The recognition buffer is efficient partly because the programs for rehearsing several letters can be set up in it simultaneously. However, the major gain in speed derives from the assumption that setting up a program to rehearse a letter is inherently a faster process than executing the program, i.e., rehearsing the letter. In fact, the biological organization of motor systems is extremely hierarchical. Thus the program in the recognition-buffer could be a program to call a program, etc., and the ultimate representation at the top of such a pyramid could be called quickly.

The rehearsal component executes the rehearsal, which then is entered and remembered temporarily in auditory storage. The memory of the rehearsal in auditory storage is scanned, the auditory image is converted to motor-instructions in the recognition-buffer, and a second rehearsal is executed. This loop continues until the response is called for and the letters are written down. I know almost nothing about the translation of the memory of a letter to its written representation except that it occurs, and therefore must be represented in the model. It has been represented in parallel with rehearsal because writing a letter so often is accompanied by vocalization.

3. CONSCIOUSNESS IN THE MEMORY MODELS

One can know the contents of the consciousness of another individual only insofar as they are expressed by his behavior, particularly by his verbal behavior. In the models, this structure would induce us to look for evidence of consciousness at the level of the rehearsal unit. However, one also must

admit that a person who is unable to speak or act may still retain consciousness. The critical aspect of the contents of consciousness is that they normally are capable of being verbalized or acted upon. Within the limits of the tasks for which Model 3 was proposed, we can identify the contents of the scan component with the contents of consciousness. This is because the scan component contains the information upon which actions are performed.

There are several inferences to be drawn from this identification. When contents of visual memory are not scanned before they fade away, they never become conscious. And, we are unconscious of all contents of our auditory memory except those being scanned. Another inference is that if the contents of a memory cannot be scanned, they are not accessible to consciousness.⁴ The untransformed contents of the recognition buffer-memory are not accessible to scanning and therefore never the objects of consciousness. This makes it indeed a mysterious component; it cannot be observed directly either from within or from without! However, this inaccessibility should not surprise us. It is axiomatic that in any system which examines itself there ultimately must be some part of the mechanism which is inaccessible to examination from within. The recognition buffer-memory is such a part in the human memory mechanism.

REFERENCES

- AVERBACH, E. and G. SPERLING, 1961. In C. Cherry (Ed.), *Information theory*, London: Butterworth, 196—211.
- CATTELL, J. McK., 1885. *Philos. Stud.*, **2**, 635—650.
- CONRAD, R., 1963. *Nature*, **197**, 1029—1030.
- GLEZER, V. D. and A. A. NEVSKAIA, 1964. *Doklady Academy of Sciences USSR*, **155**, 711—714.
- LANDAUER, T. K., 1962. *Percept. Motor Skills*, **15**, 646.
- SPERLING, G., 1960 *Psychol. Monogr.*, **74**, No. 11, (Whole No. 498).
- SPERLING, G., 1963. *Human Factors*, **5**, 19—31.

⁴ In conceptualizing Sternberg's experiments, it is useful to assume that the recognition buffer-memory can be scanned for a minor aspect of its content, e.g., whether it contains an item which was entered much more recently than the others. A dotted line has been drawn from the recognition buffer to the scan component to indicate the possibility of this kind of scan.