Sperling, G. (1963). A model for visual memory tasks. *Human Factors, 5,* 19-31.

A Model for Visual Memory Tasks[1]

GEORGE SPERLING, Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey.

A model for visual recall tasks was presented in terms of visual information storage (VIS), scanning, rehearsal, and auditory information storage (AIS). It was shown first that brief visual stimuli are stored in VIS in a form similar to the sensory input. These visual "images" contain considerably more information than is transmitted later. They can be sampled by scanning for items at high rates of about 10 msec per letter. Recall is based on a verbal recoding of the stimulus (rehearsal), which is remembered in AIS. The items retained in AIS are usually rehearsed again to prevent them from decaying. The human limits in immediate-memory (reproduction) tasks are inherent in the AIS–Rehearsal loop. The main implication of the model for human factors is the importance of the auditory coding in visual tasks.

INTRODUCTION

I shall be concerned with the apparently simple situation in which an observer looks briefly at a complex visual display and then attempts to reproduce part or all of it. Understanding this visual, immediate-memory task is important both in practical problems and in basic psychological problems. From the practical point of view, it is relevant to everyday situations such as a person looking at a number in a telephone book and then attempting to dial it, as well as to the esoteric problems that arise in matching complex visual displays to human capabilities. I shall not be concerned directly with specific applications; rather, some general principles will be evolved.

The brief visual exposure has a special theoretical significance. Normally, the eye moves in brief, quick motions between its steady fixations upon objects. These movements, called saccads, were first noticed by Javal (1878) and described by Erdmann and Dodge (1898). In such varied tasks as reading, looking at stationary objects and even in visually tracking most moving objects, the eye moves in saccads and takes in information only during the fixation pause between the saccads. As there are several fixations per second, the eye codes information from the environment into a rapid sequence of still pictures. It is natural, therefore, that the problem of what can be seen in a single brief exposure has fascinated researchers for over 100 years.[2] My purpose is to describe some of the components and properties of a preliminary model for the information processing that begins with the observation of a brief visual stimulus and that ends with the observer's response.

Historically, most research in immediate-memory (or span of attention) has been confined to the problem of capacity. That is, experimenters usually have presented subjects with a great variety of stimuli and measured simply the number of items reported correctly. While such experiments reveal something about the capacity of a memory, they do not usually reveal much about its structure. For example, by structure I mean such properties as—to use computer terminology—whether a memory has random access or whether restraints limit the sequence in which items can be remembered or recalled. To ferret out structural details, the experimental technique requires the presentation of one kind of stimulus over and over to determine the various capabilities and limitations of the observer in dealing with this stimulus. In other words, one must determine the limits of the observer's ability to cope with this stimulus depending upon instructions and

[1] Invited address, *Symposium on Information Processing in Man: Research Frontiers,* sponsored jointly by the Los Angeles Chapter of the Human Factors Society and the University of Southern California. Held at USC, June 23, 1962.

[2] Volkmann (1859) coined the word "tachistoscope" for a device he invented. It utilized falling shutters to replace the electric sparks which had been—up to that time—the primary means of producing brief exposures.

"irrelevant" stimuli such as various kinds of interference. In principle and in intent, this method has a precedent in the work of Kulpe (1904) and his followers.

The Model

In order to facilitate the exposition, it is desirable at this point to give a brief sketch of the model as follows. (1) The observer sees the stimulus material for a short time. (2) He scans it, selecting certain information to rehearse (3) He later reports what he remembers of his rehearsal. The experiments seek to clarify and quantify these three main aspects of the model. In the first place it will be established that even in a brief exposure the observer may have available much more information than he can later report. Because the duration for which he "sees" the stimulus normally exceeds the stimulus exposure duration, this component of the model is called visual information storage. The second concern is with the rate at which the observer can select and utilize the visual information. This is called scanning or read-out from visual storage. When it is not limited by the requirement of making muscular responses, the scan rate can be quite high.

The third component of the model deals with the subvocal and/or vocal rehearsal of the selected items and the memory for this rehearsal. The memory is called auditory information storage. It will be suggested that the auditory component may be the limiting factor in a wide range of both visual and auditory reproduction tasks. This single limiting process, common to many tasks, would help account for the item constancy of the so-called span of immediate-memory (cf. Pollack, 1953; Miller, 1956).

I. VISUAL INFORMATION STORAGE

Much of the material in this section has been published elsewhere (Sperling, 1960b; Averbach and Sperling, 1961) but it is presented briefly for completeness and because the results are needed for the model. Subjects are characteristically able to report about six or fewer items from brief stimulus exposures. This finding dates to the last century (Cattell, 1885; Erdmann and Dodge, 1898). In my own experiments involving immediate memory for visual stimuli, this finding occurs with great consistency. For example, in one of my studies using five experienced subjects, it was found that the average number of symbols correctly reported was equal to the number of symbols in the stimulus when stimuli contained four or fewer symbols, and equal to about 4.5 symbols when stimuli contained five or more symbols. This held for various spatial arrangements of the symbols, and for mixtures of letters and numbers as well as for letters alone. There was, moreover, no change in the number of letters correctly reported within the entire range of exposure durations tested, from 15 to 500 msec. In these presentations the stimulus was preceded and followed by a dark field. With another apparatus capable of shorter exposures, I found no change in the number of letters reported correctly even for exposures as short as 5 msec and with light pre- and post-exposure fields.[3] At such short exposures the apparent contrast of the stimulus letters was greatly reduced (black letters appeared light grey) but this did not affect the number of letters reported correctly. These results taken together, define the invariant span of attention, or span of immediate-memory.

The span of immediate-memory, however, is not due to a limit on what the subject can see. This was first proved by using a partial report procedure in the following experiment (Sperling, 1960b). Subjects were presented stimuli consisting of 12 letters and numbers in three rows of 4 symbols each. The exposure duration was 50 msec. The stimulus exposure was immediately followed by a tonal signal. The subjects had been told to report only one row of letters and the signal indicated to the subject the particular row to be reported. Subjects were able to report correctly 76 per cent of the called-for letters even though they did not know in

[3] Unpublished experiments conducted at the Bell Telephone Laboratories, 1958.

advance which particular row would be called for. This result indicates that after termination of the exposure, subjects still had available (somewhere inside them) 76 per cent of the 12 symbols, that is, 9.1 symbols. However, when the tonal signal was delayed for only one second, the accuracy of report dropped precipitously from 76 to 36 per cent. Note that 36 per cent of 12 symbols is 4.3 symbols; the previously measured memory span for this material was also 4.3 symbols.

The explanation for these results is that the visual image of the stimulus persists for a short time after the stimulus has been turned off, and that the subjects can utilize this rapidly fading image. In fact, naïve subjects typically believe that the physical stimulus fades out slowly.

The visual basis of this very short-term information storage can be demonstrated more convincingly by comparing two different kinds of stimulus presentation. In the first type, the pre- and post-exposure fields are dark; in the second, the pre- and post-exposure fields are light. It is well known that the first kind of presentation can produce persisting after-images of the stimulus and the second kind of presentation does not. Stimuli of 18 letters were exposed for 50 msec. After a variable delay following the exposure, any one of six different tonal combinations was presented to indicate the particular letters to be reported. The results of the partial report procedure applied to these stimuli are shown in Fig. 1.

The reader will recall the procedure for estimating the number of letters available to the subject. The fraction of letters reported correctly in random samples (partial reports) of the stimulus is multiplied by the number of letters in the stimulus. These values are indicated by the ordinates of Fig. 1. The abscissa indicates the time after the exposure at which the instruction calling for the partial report was given. Figure 1 shows that letters in excess of the memory span were available to the subject for one-half second when the pre- and post-exposure fields were light, and for nearly five seconds when the pre- and post-exposure fields were dark. The exposure itself was exactly the same in each case. The ability of noninformational visual fields occurring before and after the stimulus to control the accuracy of partial reports strongly suggests their dependence on a persisting visual image. Conversely, the number of letters given in the usual immediate-memory report is almost independent of the exposure conditions and does not change appreciably if the subject is required to delay his report for five seconds (or longer) instead of being allowed to report immediately.

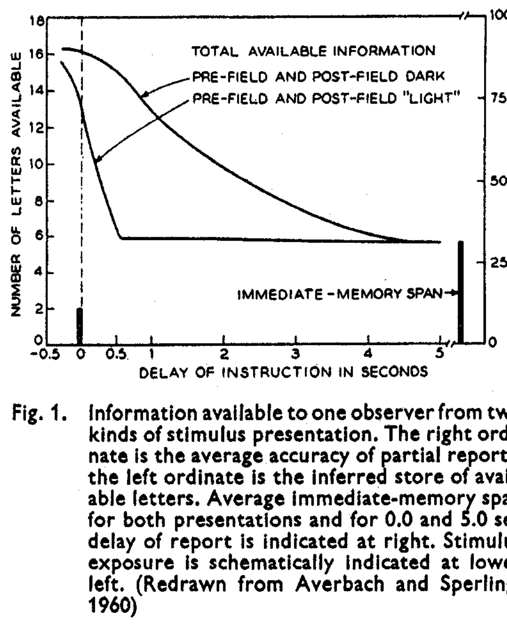

Fig. 1. Information available to one observer from two kinds of stimulus presentation. The right ordinate is the average accuracy of partial reports; the left ordinate is the inferred store of available letters. Average immediate-memory span for both presentations and for 0.0 and 5.0 sec delay of report is indicated at right. Stimulus exposure is schematically indicated at lower left. (Redrawn from Averbach and Sperling, 1960)

Therefore, the number of letters given in memory reports can be presented by a single bar in Fig. 1.

The short-term memory for letters in excess of the immediate-memory span will be called visual information storage (VIS). Some established properties and some hypotheses regarding VIS are listed below.

(1) The effective input for VIS is a local change in retinal light intensity. The contents of VIS depend on visual stimulation. Such factors as stimulus intensity, contrast, duration, pre- and post-exposure fields, etc., are particularly important.

(2) The contents of VIS can become available to subsequent components of the model as a sequence of items through "scanning" or "read-out."

(3) VIS is two-dimensional (for example, contents can be scanned either in a vertical or horizontal sequence).

(4) Maximum information in VIS is at least 17 letters when measured with stimuli containing 18 letters. Stimuli containing more than 18 letters would probably have yielded higher estimates. However, the resolution is disturbed when items that can be resolved individually are spaced too closely together.

(5) Subsequent stimuli can replace or interfere with the previous VIS contents (see below). This process is not the passive addition of a new stimulus to the fading trace of its predecessor but the active replacement of an earlier stimulus by a later one (Sperling, 1960a).

(6) Contents of VIS normally decay rapidly, decay times varying from a fraction of a second to several seconds.

(7) Long durations of visual storage can occur in the form of after-images which appear to move when the eye moves and therefore are probably localized in the retina. It has not been determined how the central nervous system is involved in the kind of VIS discussed here.

(8) It is tempting to speculate on the purpose of VIS because its properties seem so well suited to the requirements of a system like the eye, which processes information in temporally discrete chunks. The function of the persistence (the storage aspect of VIS) seems to be to maintain a visual image from one fixation of the eye to the next. The function of erasure is to permit the new image following a saccad to overwrite the trace of the previous one without interference to itself and also to "erase" the blur resulting during movement of the eye. The minimum duration of storage that has been recorded (¼ sec) is still long enough to preserve the image between eye movements. The minimum time between saccads is typically too long to allow the image produced by the second saccad to interfere with more than the tail end of the image produced by the first. Thus VIS acts as a buffer which quickly attains and holds much information to permit its relatively slow utilization later. VIS also segregates and isolates from each other successive bursts of visual information (e.g., images).

II. SCANNING

In order to determine the rate at which information can be utilized, it is necessary to gain precise control of actual stimulus availability, that is, the contents of VIS. This cannot be accomplished by controlling exposure duration alone. No matter how brief a *single* stimulus flash may be, if it is of sufficient contrast for easy legibility of letters, then it will be available for about a quarter of a second or longer.

The idea that stimulus duration does not determine stimulus availability is not new. For example, in 1868, Exner published a psychophysical study of the apparent duration of short flashes in which he found that short flashes exert their effect over a considerable time span. Baxt (1871), also working in Helmhlotz's laboratory, performed a logical sequel to this work. He followed the stimulus exposure after a delay with a bright second flash which was intended to obliterate the persisting after-image of the stimulus.[4] In one of his experiments he used a stimulus flash of 5.0 msec and a bright second flash of 120.0 msec duration. Observers viewed stimuli of 6 or 7 printed letters and reported the number of letters they could see with various time delays between the two flashes. Fig. 2 contains a graph that I prepared from Baxt's tabular,

[4] The method of Baxt was described by Ladd (1887) and James (1890) in their textbooks but it is no longer well known. Consequently, it has been rediscovered, recently by Lindsley and Emmons (1958), Gilbert (1959) and by the author.

introspective data. The abscissa represents the time from the onset of the stimulus to the onset of the "interfering" second flash; essentially, this is the time that the observer had to look at the stimulus. The ordinate represents the number of letters the observer said he could see. The parameter is the intensity of the interfering flash in terms of the arrangement that produced the intensity (three, two, or one lamps, one lamp far). The slope of the data points show that

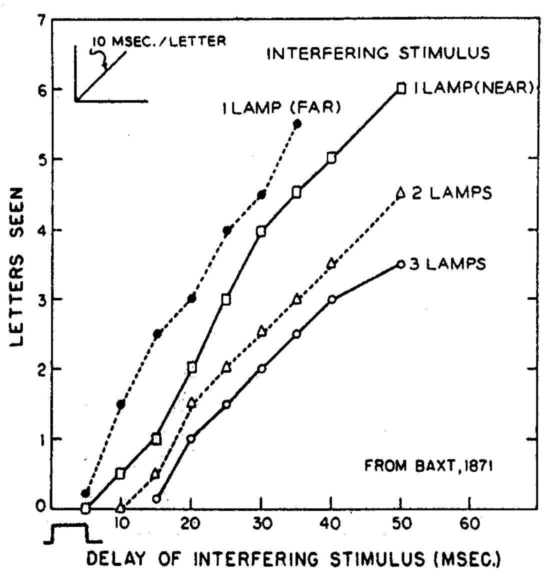

Fig. 2. The number of letters seen as a function of the delay between the onset of a 5.0 msec lettered stimulus and a 120 msec blank interfering stimulus. The parameter is the intensity of the interfering stimulus. The time course of the lettered stimulus is indicated at the lower left. (Based on the protocols of Baxt, 1871)

Baxt could see an additional letter for approximately each 10 msec of delay of the interfering flash.

The data of Fig. 2 were obtained with dark letters on a white background. Baxt obtained similar results when viewing light letters on a dark surround. He himself did not notice the simple relation of constant slope, perhaps because he did not graph his results.

Baxt's results show that the minimum delay at which detection of a letter is first possible depends on the intensity of the interfering second flash. This flash exerts its interfering effect more rapidly when it is more intense. A similar situation occurs in masking experiments. An intense masking flash can cause marked threshold changes for detection of a test flash even when the test precedes it by as much as forty msec. Had Baxt used hundreds of lamps instead of just three for his brightest interfering stimulus, the mimimum delay necessary for detection of a letter might have been increased from 20 msec (with three lamps) to 40 msec or more.

The results of Baxt's 2-flash experiment are more reasonable than those of the single flash experiments. It will be recalled that reducing the duration of a single flash reduced the apparent contrast of the exposed letters but the number reported correctly did not change with exposure duration *per se* except at ridiculously short exposures. It does not make sense to conclude, for example, that if four letters can be read from a single exposure of one msec that the time for reading a letter is ¼ msec. Nevertheless, even the scanning rate of one letter per 10 msec (or 100 letters/sec) in Baxt's experiment seemed so remarkably fast that it seemed worthwhile to attempt to reproduce some of Baxt's conditions. With certain reservations, comparable results were obtained. Some minor changes intended to improve the procedure were also tried. Subjects were required to report letters and were scored for the number correct. (Baxt had asked his observer only to say whether or not he saw a letter.) The blank second field was replaced with a visual "noise" field, consisting of densely scattered bits and pieces of letters. The noise stimulus is much more effective interference than a homogeneous field. For example, letter detection is virtually stopped when the noise and letters are viewed simultaneously at the same intensity. Baxt's procedure requires a second flash many times more intense than the stimulus to achieve comparable interference.

The main objection to Baxt's procedure, however, is a surprising one: I found that the second flash may produce a clearly visible negative after-image of the first stimulus and thereby indirectly fail to stop its persistence. In fact, in certain conditions of presentation, the blank interference produced a negative after-image of the stimulus without a prior positive image being seen (Sperling, 1960c). Ironically, here is a case of simultaneous perfect interference (the positive stimulus is completely invisible) and of perfect persistence (all the stimulus information is available in the after-image). Use of the noise field as the interfering stimulus mitigated this problem somewhat since even an after-image of the stimulus is better hidden by the noise. However, such paradoxical effects as simultaneous interference with and yet persistence of information, certainly argue for caution in interpreting the results of Baxt-type experiments.

Using the noise technique, we sought further clarification of the Baxt procedure. Baxt's results were that the scan rate is the same no matter when the scan is ended. That is, bright flashes presumably ended visual availability of the stimulus sooner than dim flashes. One obvious question that needed to be answered was "is the scanning rate the same no matter when the scan begins." If it were, then this would provide more confidence in the method.

To delay scanning, we used two different pre-exposure fields. One was a dark field which presented no visual recovery problem and the other a noise field. Fig. 3 illustrates the sequence in each of the two kinds of presentations. In the first, the pre-exposure field is dark. It is followed by an exposure of variable duration, which is then terminated by a noise post-exposure field. In the second kind of presentation, the pre-exposure field is noise, followed by an exposure of variable duration, which is then also terminated by a noise post-exposure field. The number of letters in the stimuli was varied from two to six in order to check for possible spatial location effects. Data were obtained with two

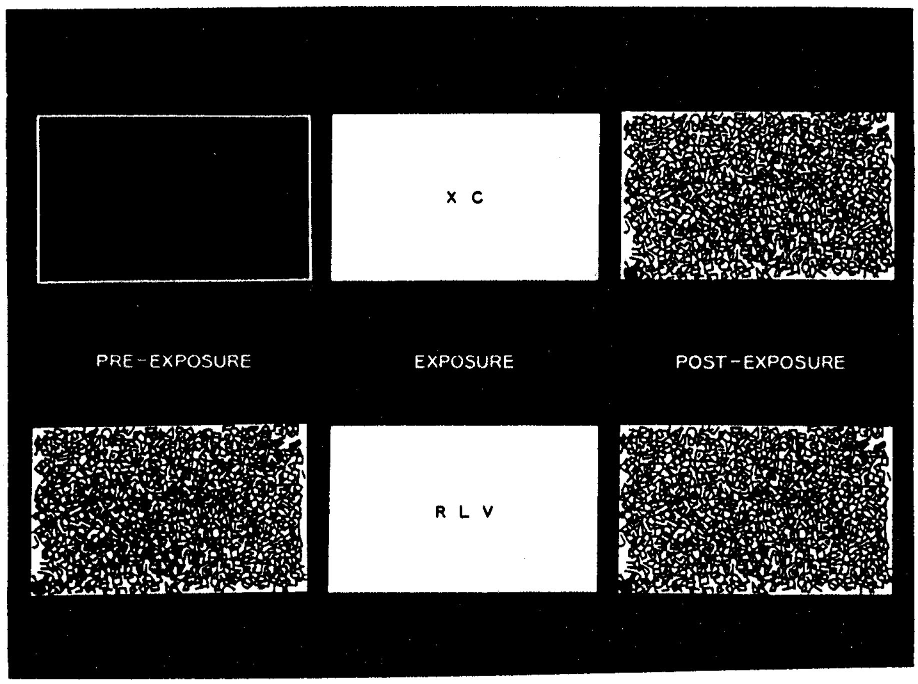

Fig. 3. Two kinds of stimulus presentations modified from the method of Baxt. The sequence of stimuli on a trial is indicated from left to right. The luminance of the lettered stimulus was 31 Ft-L.; of the visual noise, 20 Ft-L.

subjects. Fig. 4 illustrates the results for the less variable subject.

The data of Fig. 4 indicate that the subject gained information from the various stimuli at

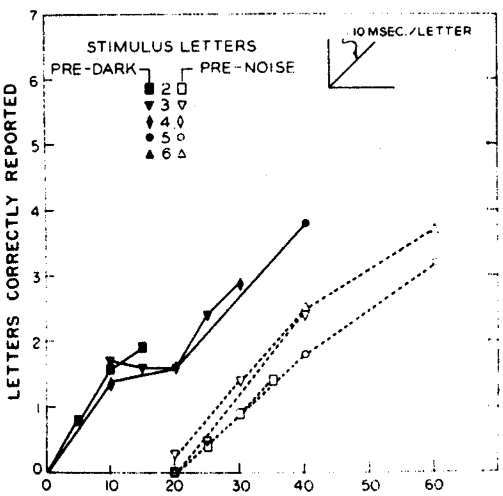

Fig. 4. The number of letters correctly reported as a function of the delay between the onset of a lettered stimulus and an interfering visual "noise" stimulus. The number of stimulus letters was also varied. The pre-exposure field was either dark or "noise."

the rate of one letter per 10 msec and that this rate was independent of the number of letters in the stimulus or of the pre-exposure field. When the pre-exposure field was dark, the subject began to gain information almost immediately upon stimulus exposure. It took this subject about 20 msec to recover from pre-exposure to visual noise. Once she was able to scan, the rate was still about one letter per 10 msec. These results support the hypothesis that pre-exposure could change the time at which a scan begins but not its rate.

Various other pre-exposure fields have also been tried. For example, with a pre-exposure field 100 times brighter than the stimulus field, scanning may be delayed for over 100 msec. Once the stimulus letters become legible, they are scanned at comparable rates. A serious problem in this kind of experiment is that it is very difficult to insure that all letters fall upon retinal locations which recover sensitivity at the same rate. This particular procedure therefore has not been pursued in detail. Preliminary data also indicated that rate of scanning was but slightly reduced when the stimulus intensity was reduced to $\frac{1}{100}$ of its prior level. Other factors such as stimulus size, geometry, and contrast have not yet been studied to see how they affect the scan rate.

The point of all these experiments is that, under a variety of conditions, random letters of good contrast are scanned at the same rate; typically, about one letter per 10 msec. However this holds true only for the first three or four letters to be scanned. Fig. 5 shows data obtained with the same two subjects viewing a dark pre-exposure field and a noise post-exposure field. One subject reported three, the other four letters in the first 50 msec of exposure. Additional stimulus exposure from 50 msec to 100 msec accounted for about one or two additional

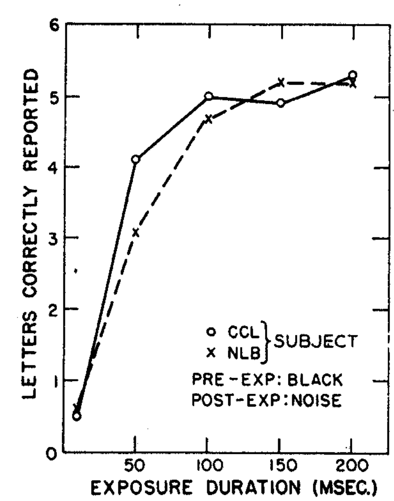

Fig. 5. The number of letters reported correctly as a function of the exposure duration or equivalently, the delay of the post-exposure visual "noise" interfering stimulus. The pre-exposure field was dark.

letters. Beyond 100 msec the rate of acquiring additional letters is so low as to be virtually indistinguishable from zero on this time scale. Additional data points would have shown the critical break in the curve to occur well before 100 msec.

This kind of experiment perhaps more clearly than any other defines an immediate-memory span for visual materials. Letters up to the immediate-memory span can be scanned at a rate of one letter per 10 or 15 msec. This is so rapid that the rate of acquiring additional letters beyond the immediate-memory span is negligible by comparison.

III. REHEARSAL AND AUDITORY INFORMATION STORAGE

In the last two sections it was shown (1) that letters are visually available to an observer during and after a brief exposure and (2) that they need not be visually available for more than 10 to 15 msec per letter in order to be reported correctly. Yet, the observer usually does not begin his report until several seconds after the exposure, long after his visual store of letters is depleted. In fact, he can delay his report for an additional 30 seconds or more without any loss of accuracy. What are the characteristics of this later, long term memory?

One important clue is obtained from the kinds of errors subjects make. In writing down symbols from a brief stimulus, subjects frequently make auditory confusions; that is, writing D for T, T for 2 and so on. On questioning, subjects assert that they indeed do "say" the letters to themselves prior to report. This kind of behavior is represented in the model by rehearsal (saying the letters to oneself) and by auditory information storage (hearing the rehearsed letters). The properties of auditory information storage (AIS) are in some respects similar to those of VIS. The main functional difference between AIS and VIS in these experiments is the possibility of feeding information already in AIS back into storage again by rehearsal.

The various components of the model and their interactions are represented in Fig. 6. The model illustrates that a visual stimulus is first stored in VIS, then scanned and rehearsed. The rehearsal component has two inputs: items scanned from VIS and those heard in AIS. Only one input is rehearsable at any one time and it probably takes a short time to switch from one input modality to the other (cf. Broadbent, 1956, 1958). The rehearsal component produces a verbal response "audible" in AIS or a vocal response audible to another person and indirectly also in AIS.

It will be suggested below that rehearsal cannot proceed faster than about ten syllables per second. Scanning itself may initially be ten times faster, 100 items per second. The difference between the rates of scanning and rehearsal poses a real problem. To account for this difference, it tentatively will be assumed that

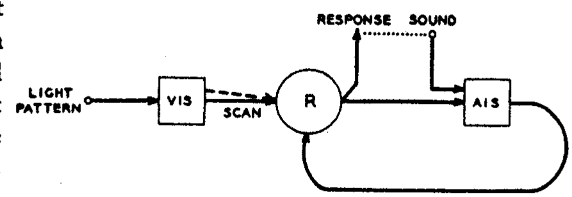

Fig. 6. A schematic model for short-term memory tasks. VIS = visual information storages R = rehearsal component, AIS = auditory information storage. See text for details.

items may accumulate very briefly in the scan path or rehearsal component until they can be rehearsed.[5] Two scan arrows are drawn from VIS to Rehearsal to represent the assumed temporal overlap in the scan of successive items.

The rehearsed items are maintained in AIS at

[5] Another possible assumption is that there is an error in the interpretation of scan experiments using a Baxt type of stimulus presentation. Two attempts were made to measure directly the stimulus persistence (availability). In the first attempt, the partial report procedure was applied to a Baxt presentation. It was not successful because subjects' performances tended to deteriorate in all respects. In the second procedure, the apparent duration of the visual stimulus was compared to the duration of various simultaneous acoustic signals. For example, a Baxt stimulus presentation from which four letters could usually be read (letter duration = 50 msec) was matched to acoustic stimuli of roughly comparable duration. In such comparisons, it is quite obvious that the apparent duration of the visual letters is a small fraction of a second. Although this is a subjective judgment, it has strong face validity. The result only affirms that the scan rate can be much more rapid than rehearsal.

least for several seconds during which time they are usually rehearsed again. Finally, when a response is required, the letters are either rehearsed again as they are being written or, if they are spoken aloud, the feedback loop may be closed externally as well as internally—the spoken sound re-entering AIS.

In summary, there are at least three sources of information for auditory storage: (1) an actual acoustic stimulus, (2) rehearsal of information already in AIS, (3) the scanning process—it is assumed that observers hear themselves make a verbal response as they scan.

Evidence

In this model, the auditory components assume the major burden of short term visual memory. Data relevant to the auditory part of the model are difficult to obtain because audition involves higher order processes. There are some data, however, which come from three kinds of experiments: (1) experiments that measure AIS directly by using auditory stimuli, (2) experiments that seek to measure rehearsal by instructing subjects with regard to rehearsal, and (3) experiments that use interfering stimuli to interfere selectively with some aspects of the process and thereby reveal others. I shall consider these approaches in order.

(1) *Auditory information storage.* Our common experience leaves little doubt of the fact that there is auditory information storage. If, for example, one is spoken to while occupied with another task, one may continue working for a few moments and yet still be able to recover the message later. Two kinds of recent experiments illustrate this type of auditory storage especially well. In the first kind of experiment the subject is presented with several stimuli at once. Foi example, Broadbent (1597a) produced a spoken sequence of digits at the subjects' ears, with different digits being spoken simultaneously into each ear. When digits were spoken rapidly, subjects could not shift their attention back and forth from one ear to the other with each pair of spoken digits. Rather, they could only report all the digits heard at one ear first and then attempt to report the digits—if any—that were still audible at the other ear. This ingenious procedure excludes the possibility of any simultaneous recoding of the inputs at both ears. It thereby forces the subject to rely directly upon AIS.

The second kind of experiment (Anderson, 1960) was a direct application to audition of the partial report procedure previously used to measure visual storage. A list of 12 letters in three groups of four each was read to subjects. At various times later, an instruction signal was given the subjects to indicate the particular group to be reported. Anderson's results can be interpreted in terms of auditory storage to mean that initially subjects had about 10 letters available. Accuracy of report deteriorated most rapidly during the first five seconds after the presentation and slowly thereafter. Other factors, such as unlimited opportunity for rehearsal and differential recall for certain groups of letters indicate that more than simple AIS was probably involved. All in all, both experiments demonstrate an auditory information storage capable of retaining information before it has been acted upon and whose contents decay during a period of many seconds.

(2) *Rehearsal.* The role of rehearsal in memory is a problem that psychologists—with a few exceptions—have hoped would solve itself if only they ignored it sufficiently. Rehearsal is not quite so intractable to measurement as many of us were brought up to believe. For example, subjects may be asked to say a particular sequence of letters as fast as they can and to use a stop-watch to time themselves for the task. They may also be asked to say the letters to themselves without producing any sound and to time themselves as before. Subjects understand this second instruction without difficulty. Elapsed times are about three syllables per second for unfamiliar letter sequences spoken aloud. Silent rehearsal is slightly faster. The maximum rate for highly familiar sequences is about 10 syllables per second.

More controlled and somewhat similar studies of rehearsal were conducted by Brown (1958), Peterson and Peterson (1959) and by Murdock (1961). In the Petersons' experiment, subjects were initially presented with a sequence of letters. They were then told to rehearse them out loud in time with a metronome at two letters per second. Other subjects were given an equivalent time duration that could be used for silent rehearsal. The length of time allowed for rehearsal was varied. Following the initial rehearsal period, the subjects were put to a task intended to interfere with further rehearsal, such as counting backwards. Those subjects who had most time for initial rehearsal—aloud or silent—recalled best, accuracy being proportional to rehearsal time. Spoken rehearsal was more effective than silent rehearsal, a result that will be encountered again below.

These experiments indicate that a short period of rehearsal after an auditory stimulus presentation facilitates later recall. An interpolated period of non-rehearsal interferes with recall as does the competing auditory stimulus of hearing oneself saying irrelevant material.

(3) *Interference.* Brown (1958) sought to interfere with rehearsal by requiring his subjects to read digits aloud. Peterson and Peterson (1959) made their subjects count backwards between stimulus presentation and recall. These interpolated tasks produced a decrease in accuracy of recall which was interpreted as due to the interference of counting with rehearsal. The longer the period of counting, the less accurate the recall, presumably because of the fading of the rehearsed letters in auditory storage and the accumulation of interfering auditory stimuli. Brown had previously noted that interpolated silence (which permits rehearsal) does not have a comparable destructive effect on recall.[6]

To a certain degree, even when he is counting backwards, the subject's cooperation is needed for interference with his rehearsal. He is in effect asked to devote himself fully to the task of counting and to neglect completely the task of recall. It would be desirable to find a method of interference whose success did not depend upon the subject's cooperation—a method comparable to Baxt's method of eliminating visual persistence. Towards this end, I constructed two different acoustic stimuli which might interfere with rehearsal by providing a competing input to AIS. The first of these was a loud noise at 90 dB SPL, analogous to Baxt's interfering field. The second interfering stimulus was a recording of the subject's own voice speaking letters at a rate of 6/sec, also at 90 dB. This stimulus is analogous to visual noise. At each trial, either silence or one of these extremely loud stimuli was played into the subject's ears, beginning before stimulus exposure and continuing until after report.

The number of letters subjects reported correctly was the same in noise and in silence, and only very slightly lower for the speech interference although the difference was statistically significant (Sperling, 1962). Two supplementary observations were made which elucidate the failure of the interpolated stimuli to interfere with recall. First, the subjects described the task as one of selective listening, that is, of hearing the speech interference as localized at the ears and their rehearsal as localized inside the head. The speech interference accentuates the subjective aspect of listening to rehearsal so that even subjects who are not normally cognizant of the auditory aspect become so.

The second observation was that normally silent subjects, when listening to speech interference, frequently mouthed or even spoke the letters out loud during the period between the stimulus exposure and writing down the response. This overt behavior reflects an effort to accentuate and emphasize the rehearsal. The conclusion therefore is that speech and perhaps also noise do interfere but that subjects can accentuate their rehearsal in order to cope successfully with the interference.

Summary of Hypotheses about Auditory Information Storage (AIS) and Rehearsal

(1) The primary input to AIS is sound. The contents of AIS probably depend on sound in-

[6] Brown, J. *Immediate Memory.* Doctoral thesis, University of Cambridge, 1955. Cited in Broadbent (1958), p. 225. See also Averbach and Sperling (1960).

tensity, complexity, background, etc.[7] A secondary and possibly less effective input is by rehearsal.

(2) AIS is locally one-dimensional and directional. That is, a string of letters is stored in sequence and it can be recovered only in that sequence. However, one string of letters may be localized at the left ear and another simultaneous string of letters may be localized at the right ear. Thus, although each stored string of letters is individually one-dimensional, the simultaneous imbedding of many strings in storage implies a very complex overall structure for AIS.

(3) At least 10 non-rehearsed items can be maintained momentarily in AIS. They are probably grouped into several strings.

(4) The contents of AIS decay slowly relative to VIS. Rehearsal of items appears to lengthen their subsequent decay times but it also brings into play more complex mechanisms for long term memory via association which are beyond the scope of this paper.

(5) The contents of AIS, VIS, and of any comparable system in other modalities may be rehearsed. That is, a verbal response is made to items in the storage and this verbal response, which may or may not be audible to another person, is an input to AIS. When a stimulus consists of a number of clearly identifiable items, then the characteristics of the AIS–Rehearsal loop will probably be the limiting factor in the immediate-memory span for this stimulus.

(6) Rehearsal usually proceeds at a rate of about 3 syllables per second and cannot proceed faster than about 10 syllables per second.

(7) Only one item may be rehearsed at a time. The interval between the rehearsal of successive items is longer if they are not taken from the same sense modality.

[7] Wundt (1912) provides an insightful and provocative analysis of auditory memory for patterns of metronome beats. Wundt could maintain in his consciousness 8 groups of two beats each or 5 groups of eight beats each. This material was presumably not rehearsed verbally. Wundt's introspections would suggest that the contents of AIS are limited most directly by subjective grouping (itemizing) and only to a lesser extent by their duration or complexity (information content).

(8) Subsequent acoustic stimuli can interfere with the contents of AIS beyond what might be expected from passive decay. So far, only self-produced verbal stimuli have shown the effect. The interference is most effective against items that have not been rehearsed and diminishes with the number of rehearsals. If rehearsal is permitted during the presentation of interfering stimuli (recorded, in this case) then the interference is barely measurable.[8]

DISCUSSION

The model is perhaps more useful for organizing data and experiments than for precise prediction because it consists of complex and only partially specified components. In this respect it reflects the complexity of its human subject matter and the limitations of its human designer.

Transient versus Steady-State Information Flow

The model was derived only from the "transient" response of a human to information; that is, the situation in which he is given a single burst of information and much time to process it. The responses to a "steady-state" or continuous information input undoubtedly involve further complications and new processes. For example, following a brief stimulus exposure, the auditory feedback loop from AIS through Rehearsal

[8] Broadbent (1957b, 1958) has suggested several models for immediate-memory. While a brief discussion cannot do justice to his ingenious and stimulating proposals, it is worthwhile to indicate some of the major similarities and differences. The obvious, basic similarities are the possibility of storage of unrecoded information and the recognition of a rehearsal loop. The differences lie in the nature and subdivision of these functions and are undoubtedly due in part to the different sets of data which each account seeks to explain. In the present model, items (or stimuli) circulate in the rehearsal loop while Broadbent specifically states that information (not items) circulates. The present account assumes one rehearsal loop and one store for rehearsed material, not several of each. Here VIS and AIS are identified in detail from introspective and direct experimental data while in Broadbent's models, storage of unrecoded information is of a more inferential and hypothetical nature—it can even be by-passed entirely. Broadbent is especially concerned with the case of simultaneous or conflicting inputs. He does not consider the special problem of scanning in vision nor make comparable distinctions between the processing of visual and of acoustic stimuli; etc.

was of primary significance in recall. Visual processes in recall may come to be of importance only with longer exposures. Perhaps this is because following a brief stimulus exposure, most people cannot create in themselves a visual feedback loop, but they can in effect do so during a continuous stimulus exposure by looking repeatedly at certain aspects of the stimulus. This example emphasizes another difficulty in the understanding of steady-state observation conditions. Namely, the sequential flow of information into the observer may depend on his own actions so that every complexity and nuance of human judgment and decision making must be considered in a complete steady-state model. On the other hand, some complex tasks have been devised so that a continuous verbal protocol is available—the subject thinks out loud so-to-speak (e.g., Newell and Simon, 1961). In such tasks, the knowledge of how many and for how long the subject's past utterances are available to him in AIS may well be of significance in the understanding of the process of decision making or "thinking."

Keeping in mind the restriction of the model to "transient" responses, it is still possible to make some hypotheses about the mechanisms underlying humans' response to complex displays and some tentative suggestions on how to improve their response.

(1) *Independence of the immediate-memory span.* Because of verbal rehearsal, humans transmit items, not bits. The limit on the number of items depends on such factors as the code (e.g., syllables per item) and upon the individual characteristics of the AIS–Rehearsal loop (e.g., AIS decay time). In so far as these are independent of the input modality of the stimulus, the information limitations in human responding will be independent of sense modality.

(2) *High information requirements.* Even in a brief exposure a human can take in much more information than he can ultimately transmit. By scanning and rehearsal he effectively samples an uncoded form of the input in VIS and AIS. In many practical situations one cannot dictate to the observer what aspects of the stimulus to sample. For example, in order to provide customer satisfaction, a television picture may have to contain information to match the enormous capacity of VIS and not merely the small amount of information transmitted in a sample from it. On the other hand, in other contexts it may be economical to keep the displayed information at the transmission (and not the storage) level.

(3) *Rapid reading.* Observers can extract small numbers of items very quickly from visual stimuli; 10 to 15 msec per letter is a typical rate for the first three or four items. In the absence of interfering stimuli, even a microsecond flash, if it is seen clearly, will usually be visually available for many times the length of time needed to extract the items up to the immediate-memory span. To transmit the maximal information from a brief exposure of a visual display, therefore, the display should be coded into about four symbols (e.g., digits) to take advantage of the rapid scan capability. Of course, other considerations such as the kind of errors may indicate other kinds of displays. In steady exposures, the rate of item utilization is usually not limited by visual factors but by the rehearsal process to a rate of 10 syllables per second or slower.

(4) *Non-susceptibility to auditory interference.* One useful if obvious result is that properly motivated subjects can perform visual memory tasks during extreme auditory interference and show almost no performance decrement. Specifically, subjects can perform without loss in 90 dB noise and almost without loss when barely tolerable levels of speech (recorded letters spoken by themselves) are being "shouted" into their ears. For practical purposes, therefore, it may be assumed that the number of letters reported from a brief exposure will be the same under optimal or adverse conditions.

(5) *Auditory simplification.* In order to facilitate performance and to avoid errors in visual monitoring tasks, telephone dialing, etc., it is necessary to consider the task from the auditory point of view as well as the purely visual. All possible outcomes should be given names that are easy to rehearse and not likely to be confused. The *laissez-faire* practice of allowing the operator

to develop his own terminology is not recommended. When only numerals are involved, there is no problem as these are quite distinct. But we have found, for example, that subjects required to memorize sequences of the letters B,P,D,T, etc., do not do as well as when confronted with an equivalent sequence of letters which do not sound so much alike. Frequently, complex tasks which do not appear to involve audition can be simplified and made less susceptible to errors by means of an appropriate code that will ease the auditory memory load (in syllables) and help to avoid auditory confusions.

REFERENCES

Anderson, Nancy S. Poststimulus cuing in immediate-memory. *J. exp. Psychol.*, 1960, **60**, 216–221.

Averbach, E., & Sperling, G. Short term storage of information in vision. In C. Cherry (Ed.), *Information theory*. London: Butterworth, 1961, pp. 196–211.

Baxt, N. Ueber die Zeit welche nötig ist, damit ein Gesichtseindruck zum Bewusstsein kommt und über die Grösse (Extension) der bewussten Wahrnehmung bei einem Gesichtseindrucke von gegenbener Dauer. *Pflüger's Arch. ges. Physiol.*, 1871, **4**, 325–336.

Broadbent, D. E. Successive responses to simultaneous stimuli. *Quart. J. exp. Psychol.*, 1956, **8**, 145–152.

Broadbent, D. E. Immediate memory and simultaneous stimuli. *Quart. J. exp. Psychol.*, 1957a, **9**, 1–11.

Broadbent, D. E. A mechanical model for human attention and immediate memory. *Psychol. Rev.*, 1957b, **64**, 205–215.

Broadbent, D. E. *Perception and communication.* New York: Pergamon Press, 1958.

Brown, J. Some tests of the decay theory of immediate memory. *Quart. J. exp. Psychol.*, 1958, **10**, 12–21.

Cattell, J. McK. Ueber die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben. *Philos. Stud.*, 1885, 2, 635–650.

Erdmann, B., & Dodge, R. *Psychologische Untersuchung über das Lesen.* Halle, M. Niemeyer, 1898.

Exner, S. Ueber die zu einer Gesichtswahrnehmung nöthige Zeit. *S.B. Akad. Wiss. Wien.*, 1868, **58**, 601–633.

Gilbert, L. C. Speed of processing visual stimuli and its relation to reading. *J. educ. Psychol.*, 1959, **50**, 8–14.

James, W. *The principles of psychology.* New York: Holt, 1890.

Javal, L. E. Essai sur la physiologie de la lecture. *Ann. Oculist. Paris*, 1878, **82**, 242–253.

Külpe, O. Versuche über Abstraktion. In *Bericht über den I. Kongress für experimentelle Psychologie*, 1904. Leipzig: Barth, 1904, pp. 56–68.

Ladd, G. T. *Elements of physiological psychology—a treatise of the activities and nature of the mind.* New York: Charles Scribner's Sons, 1887.

Lindsley, D. B., & Emmons, W. H. Perception time and evoked potentials. *Science*, 1958, **127**, 1061.

Miller, G. A. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

Murdock, B. B., Jr. The retention of individual items. *J. exp. Psychol.*, 1961, **62**, 618–625.

Newell, A., & Simon, H. A. Computer simulation of human thinking. *Science*, 1961, **134**, 2011–2017.

Peterson, L. R., & Peterson, Margaret J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

Pollack, I. Assimilation of sequentially encoded information. *Amer. J. Psychol.*, 1953, **66**, 421–435.

Sperling, G. Bistable aspects of monocular vision. *J. opt. Soc. Amer.*, 1960a, **50**, 1140–1141. (Abstract.)

Sperling, G. The information available in brief visual presentations. Psychol. Monogr., 1960b, **74**, No. 11 (Whole No. 498).

Sperling, G. Negative afterimage without prior positive image. *Science*, 1960c, **131**, 1613–1614.

Sperling, G. Auditory interference with a visual memory task. Paper read at Eastern Psychological Association, Atlantic City, New Jersey, April, 1962.

Volkmann, A. W. Das Tachistoscop, ein Instrument, welches bei Untersuchung des monentanen Sehens den Gebrauch des electrischen Funkens ersetzt. *S.B. Kgl. Sächs. Ges. Wiss. Lpz., math-phys.*, 1859, **11**, 90–98.

Wundt, W. *An introduction to psychology.* Tr. from 2nd German ed., 1911, by R. Pintner. New York and London: Macmillan, 1912; reprinted London: Allen & Unwin, 1924.