16

SHORT TERM STORAGE OF INFORMATION IN VISION

E. Averbach and G. Sperling*
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey, U.S.A.

There have been numerous estimates of human ability to extract information from brief visual presentations. The greatest information intake has been observed with exposures of items having high information content as compared to items of low information content, e.g. decimal digits and letters as opposed to binary digits. The limit of performance in these experiments, called the 'span of perception' or 'span of immediate memory', seems to be a limit on the number of items recalled rather than a limit on amount of information. A subject can report only about a half dozen items regardless of the ensemble from which these items were chosen.

'Span' experiments are very gross. They treat the observer as a transmission link without indicating where or how, information is lost between exposure of the stimulus and the subject's report. This paper describes some experiments that attempt to measure the characteristics of the early stages of the perceptual process. In the first part* (by G. S.) the classical 'span' technique of dealing with brief exposures is contrasted with a 'sampling' technique. It is shown that humans are able to store rather large amounts of information for short time periods. Experimental estimates are given of the amount and duration of the visual storage and of the rate at which visually stored information can be utilized. The second part of the paper (by E. A.) presents a slightly different sampling experiment, two kinds of 'erasure' experiments and a model which relates observed storage time, true storage time, erasure and read-out rate.

PART I

The span type of experiment is an old one, and it is not surprising that it has many names. It has been called the span of apprehension, span of attention, span of perception, and span of immediate-memory experiment. In this kind of experiment the subject is briefly shown a stimulus containing a number of letters. He is asked to report as many letters as he can; that is, to make a *whole* report of the visual stimulus.

This whole report procedure was tried with a variety of different stimulus arrays[1]. *Figure 1* shows some typical stimuli. These arrays vary in the number of items, in their spatial arrangement, and in their composition; that is, some arrays have letters alone, others have letters and numbers. The various stimuli were exposed for 50 msec (1/20th sec) individually to five highly

---

* This paper is a joint presentation of two separate lines of research. The theory and methods of Part I are due to G. Sperling, those of Part II to E. Averbach.

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

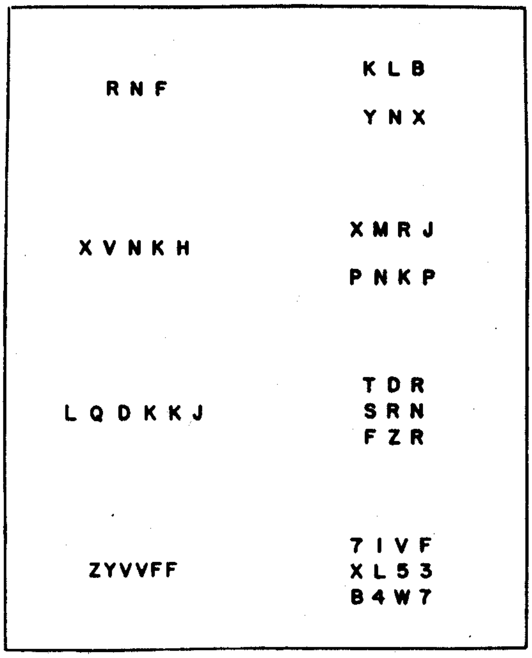

Figure 1

trained subjects. *Figure 2* shows the average number of letters that subjects were able to report correctly. The subjects reported nearly all the letters correctly so long as the number of letters in the stimulus did not exceed five. When stimuli contained five or more letters, subjects were able to report only

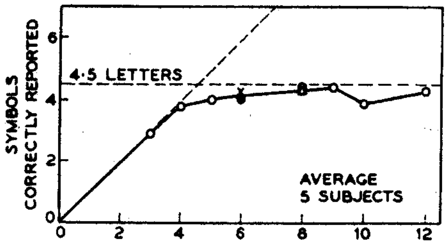

Figure 2

about 4½ letters correctly on the average. These are the classical results. We would say that the span of immediate-memory for these stimuli is about four to five letters.

In order to find out if this five letter limit was determined by the short exposure time that was used, the exposure duration was systematically varied from 15 to 500 msec at the same intensity. *Figure 3* shows the results. Apparently, within the limits of the conditions used, exposure duration does not influence the number of letters reported.

The span type of experiment may be summarized as follows: when the subjects are asked to report all the letters of a stimulus, they can report only about five letters on the average; this limit of 'immediate-memory' holds for a wide range of visual arrays and exposure durations.

In the whole report procedure observers often assert that they could *see* more letters than they were able to *report* later. They say that while they are reporting some letters, they forget others. This suggests that the immediate-memory span sets a limit on a process that is otherwise rich in available information. In other words, although an observer can correctly report only about five letters from the brief visual stimulus, he may nevertheless have chosen these five letters from a larger store of letters which were momentarily available to him. In the sampling type of experiment an attempt was made to ascertain how much information does, in fact, become available to the observer as a result of the stimulus.

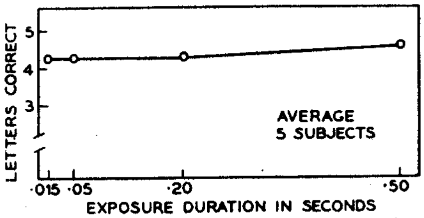

Figure 3

In this experiment a sampling procedure which does not require a whole report is used in order to circumvent the immediate-memory span. This method requires the observers to report only a part (designated by location) of a letter array exposed for 1/20th sec. The part to be reported consists of just one out of three rows of letters. It is small enough (3–4 letters) to lie within the memory span. A tonal signal (high, medium or low frequency tone) was used to indicate which of the rows was to be reported. The subject did not know which signal to expect, and the indicator signal was not sounded until after the visual stimulus had been turned off. In this way, information available for report was sampled immediately after the termination of the stimulus.

There is an important procedural difference in the two kinds of experiments. In the first kind, the observer is required to report all the letters of the stimulus. He must give a 'whole' report. In the sampling experiment the observer reports only one row of letters of the stimulus, but he does not know which row of letters will be called for until after the stimulus has been turned off. We call this a 'partial' report.

Each observer, for each set of material tested (stimuli of 6, 8, 9 and 12 symbols), gave partial reports that were more accurate than whole reports for the same material[1]. These results are illustrated in *Figure 4*. The lower curve is the same immediate-memory data as that illustrated *Figure 2*. These are obtained by a *whole* report. The upper curves represent the accuracy of *partial* reports. For example, following the exposure of stimuli consisting of 12 letters, 76 per cent of the letters called for in the partial report were given correctly by the observers. It is possible to calculate the total information available to the observer from which he draws his partial report. It is about 9·1 letters; namely, 76 per cent of 12 letters. The 9·1 randomly chosen letters are equivalent to 40·6 bits of information. This estimate of the information available in a brief exposure is considerably more than previous experimental estimates (20 to 25 bits) which were obtained using a whole report. Apparently, the subject's memory span was the limiting factor in these experiments. Further experiments with stimuli containing more than 12 letters showed that the 40-bit figure observed using partial report experiments was limited by the small amount of information in the stimuli rather than by the capacity of the observers.

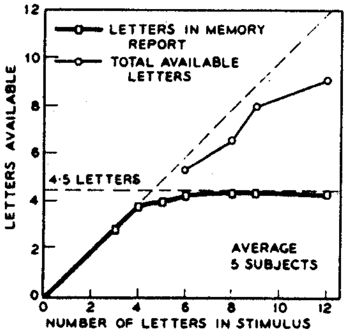

Figure 4

These results show that immediately after the stimulus is turned off, observers have available at least two to three times more information than they can give in a whole report. In order to determine how this available information decreases with time, the instruction signal which indicates the row of the stimulus to be reported was delayed by various amounts.

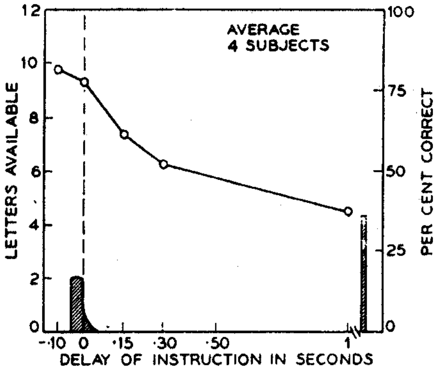

Figure 5

Figure 5 shows data obtained with stimuli having 12 letters each. The light flash is schematically indicated at the lower left. The span of immediate-memory for this material is indicated by the bar at the right. Note that information in excess of the memory span is stored for less than a second. Both the whole report and the partial report procedure give exactly the *same* estimate of the number of available letters (namely four to five) when the instruction to report is delayed by more than a second. Only when the instruction is given within a second of the exposure do the results obtained by the two methods differ. That is why the results of this experiment are different from those of previous ones; in this experiment, unlike earlier ones, the instruction to give a partial report was coded so that it could be given at a precisely determined time, *within* one second of the exposure.

One of us (E. A.) has recently completed a similar experiment, to be described in detail below, in which a visual arrow rather than a tonal instruction was used to indicate the letters to be reported. The results are quite similar to those shown in *Figure 5.* It seems to make little difference whether a tone (which calls for three letters) or an arrow (which calls for only one letter) is used. The important thing is the use of a sample, the partial report, which consists of fewer letters than the memory span.

Once it is established that subjects can retain much stimulus information for a brief time, the question naturally arises: how do they do it? The title of this paper suggests that they do it visually, i.e. that they can utilize a visual image of the stimulus, which persists for a brief time after the exposure before it fades. Subjects say, for example, that they can still 'read' the stimulus even when the instruction tone comes several hundred milliseconds after termination of the stimulus. In fact, naïve subjects sometimes think that the physical light source is a slowly fading one.

A good way to show that stimulus information can be retained as a persisting visual image is to show how the persistence depends on the kind of visual stimulation. *Figure 6* indicates two possible kinds of presentations. In each case, the stimulus is exposed for 50 msec at an intensity of 31 ft.-lamberts. In the first type of presentation the pre- and post-exposure fields are dark. This presentation is thought to favour persisting after-images. In the second type of presentation the pre- and post-exposure fields are about equal in intensity to the stimulus. An experiment using stimuli of 18 letters was performed in order to compare these two kinds of presentation. *Figure 7* shows the results for one representative subject. The persistence of stimulus information is clearly different in the two kinds of presentation. When the pre- and post-exposure fields are dark, a legible image of the stimulus persists for longer than two seconds after the exposure, whereas, in the light pre- and post-exposure presentation the persistence is definitely less than one half second.

Figure 7 also shows immediate memory data (whole report) for each of the two conditions of presentation. In these immediate-memory experiments the subject was either allowed to report 'immediately' or he was required to wait for 5·0 sec after the exposure before beginning to report. Subjects can delay an 'immediate-memory' report for much longer than 5 sec without any loss of information, but 5 sec is the largest delay plotted in *Figure 7.* This kind of experiment shows that so-called 'immediate-memory' is virtually a permanent memory. Note that despite the great differences in partial reports, the whole reports are the same for the two kinds of presentation and for the two different delays.

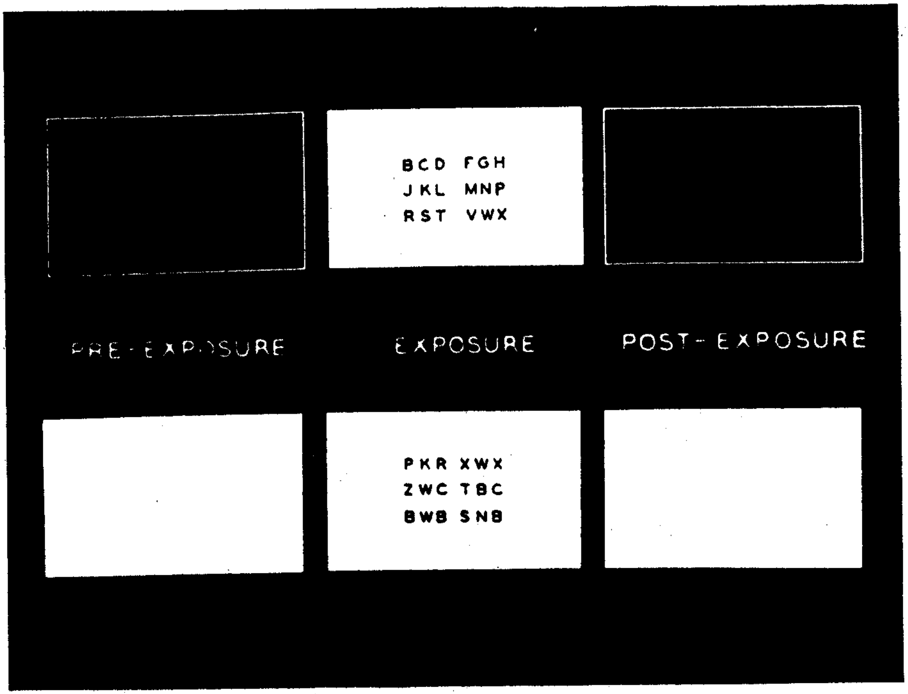

Figure 6

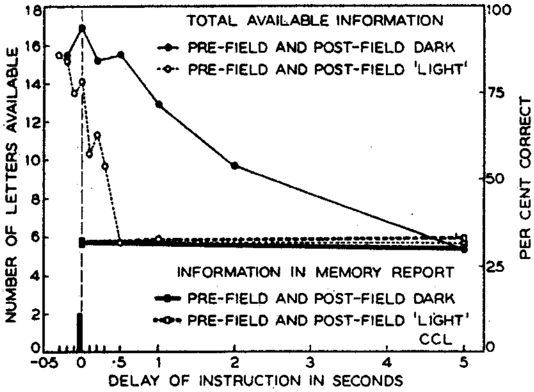

Figure 7

In *Figure 8* these data are abstracted. The rise and fall of visual information following a brief exposure is schematically illustrated. It is assumed the rise begins at the onset of the exposure, and that decay of information follows the curve shown in *Figure 7*. However, exact values are not important for the inferences that will be drawn; they obtain for any reasonable model. In fact, one such model will be presented below.

The stimulus consists of 18 letters or 77·6 bits of information, counting each letter as 4·3 bits. Since it is exposed for 1/20th sec, there are 1/20 × 77·6 or 3·9 bit-sec of stimulus information. (In *Figure 8* the contraction 'bis' is used for bit-second.)

In the presentation with dark pre- and post-exposure fields there are about 120 bit-sec. This means that, on the average, stimulus information is stored for approximately 30 times longer than it was presented. This is certainly information storage; in this case, it is 'visual information storage'.

Even in the presentation with light pre- and post-exposure fields, which is the traditionally 'good' tachistoscopic presentation, stimulus information is, on the average, visually available for about nine times longer than it is presented.

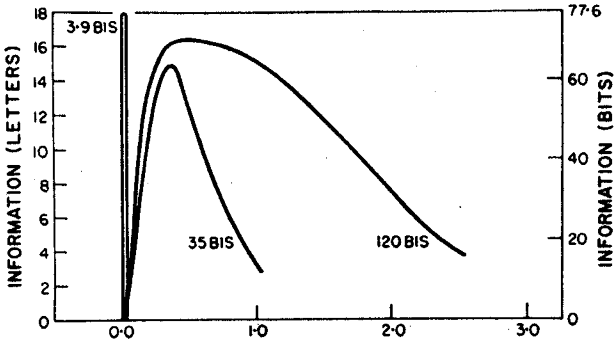

Figure 8

These considerations show that in ordinary tachistoscopic presentations the time for which the stimulus is visually available may greatly exceed the exposure duration. Therefore, if the rate at which visual information can be utilized is computed on the basis of exposure duration (as it usually is), then this computed rate is likely to be wildly erroneous.

In order to find out how fast information can be utilized or 'read-out' of visual storage, we use a masking method in which the persisting image is masked or 'erased' by a subsequent interfering field.[2] In our procedure, the stimulus exposure is followed immediately by a visual 'noise' field, which consists of parts of letters spread randomly over the field. The noise field would effectively mask the stimulus even if both fields were on simultaneously; therefore, it is assumed to stop any possible persistence of the stimulus. The subject's task is to report as many letters as possible, and the exposure duration (delay of the noise field) is varied in successive blocks of trials. Although this procedure has its pitfalls, we have found the following generalization to hold over a wide range of exposure conditions. Each 10 msec increment in exposure duration enables the subject to report about one additional letter, provided that the total number of letters reported is not greater than four. The fifth and sixth letters, if they are reported, require a relatively longer exposure. This experiment shows that information can be read-out of storage rapidly in bursts of about five letters. During such a reading burst the rate can be as high as 100 letters per sec.

PART II

In this section we will outline four experiments* aimed at elucidating some of the erasure properties of the visual storage and at estimating the visual storage time for a fixed and constant background brightness. The first experiment is essentially like the sampling experiment in Part I, except that a visual signal is used instead of a tone to designate part of a briefly exposed array.

1. Retention

A 2 × 8 array of randomly chosen letters is exposed for 50 msec. Then, after a variable delay, a black bar marker of 50 msec duration is presented either above one of the letter positions of the upper row, or below one of the letter positions of the lower row. The subject's task is to name the letter designated by the marker. *Figure 9* shows a typical array and bar marker. Also shown is a black circle that is used as a marker in the second experiment to be described later.

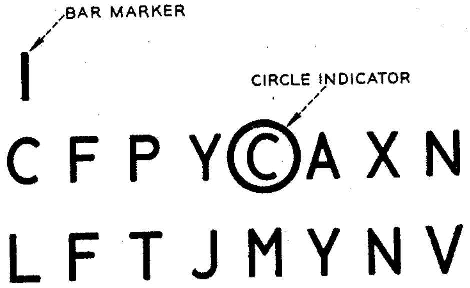

Figure 9

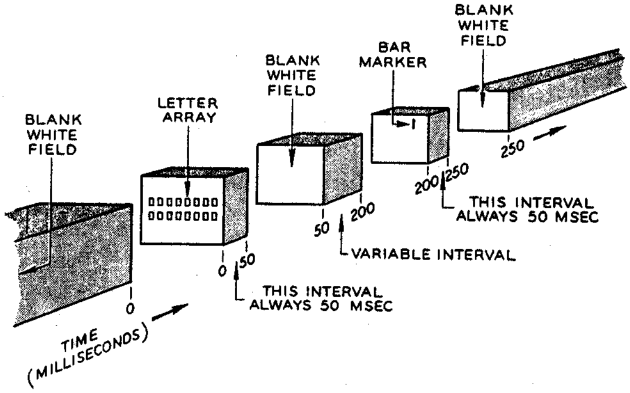

Figure 10. Sequence presented in a typical trial

The experimental sequence is illustrated in *Figure 10.* A uniform field of 70 ft.-lamberts, having a dark fixation point in the centre, was maintained constantly. The black letters and marker appeared briefly against this

---

* Mr. A. S. Coriell of Bell Telephone Laboratories assisted with the experimental work of Part II.

background. The letter size and contrast were such that single letters, exposed for 50 msec in any of the 16 letter positions, were perfectly legible when the eyes were fixed on the centre. The experiment involved three sessions during each of which 128 different arrays were exposed. The markers appeared randomly in each of the 16 array positions at each of eight time intervals between array and marker. An additional session was run later in which array and marker were presented simultaneously.

Figure 11 shows the average performance of our three well-practised subjects as a function of the time between onsets of array and marker. The vertical lines indicate the onset and offset of the array. The curve obtained is very much like that reported in Part 1 (cf. *Figure 7*) with similar exposure conditions (light after-field).

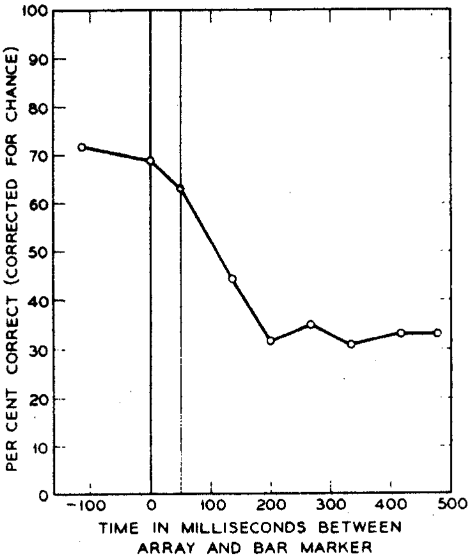

Figure 11. Retention

Although it might be assumed that this experiment yields a reasonably good description of the time decay of the short term visual storage, the curves obtained cannot be said to represent this decay, for two reasons. First, the true storage would be expected to decay to zero for long enough time intervals; these results decay to a final level of about 30 per cent indicating that the measured performance contains components of more permanent memory as well as the short term memory component we would like to measure. Second, because the process of detecting a marker and reading a letter undoubtedly takes time, the measured performance should not give a true indication of the storage decay. As will be shown, it is possible to measure and correct for these factors in arriving at an estimate of the storage time.

It is somewhat surprising that even when marker precedes array by over 100 msec, performance does not reach 100 per cent. The explanation seems to lie in the fact that letters, although perfectly legible by themselves, are not legible in the context of the array. This is shown in *Figure 12* by a plot of performance as a function of position. The numbers 1 to 8 represent, from left to right, the positions of the upper line of the array, and 9 to 16 the positions of the lower line.

The percentage is based on the pooled data of the three subjects taken across all time intervals. All subjects show the same distribution in which performance is better at the centre and ends and poorer in between. Performance on the upper line is consistently better than performance on the lower.

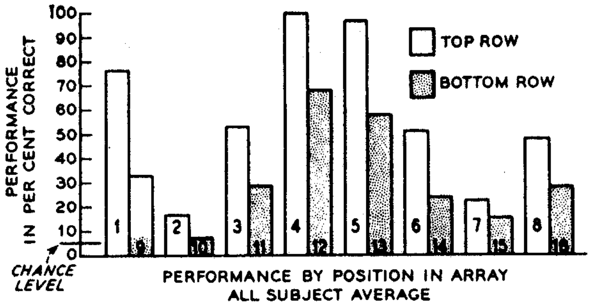

Figure 12

An estimate of the amount of information stored, based on performance measured, when bar marker follows immediately after array yielded the figure of 37 bits for one subject and 54 bits each for the other two.

2. *Erasure*

If persistence were the only property of the visual storage, it would be difficult to understand how we see at all in our normal continually changing visual environment. A storage process ordinarily involves erasure also, to assure that old information is out of the store before new information is put in. Otherwise new information and old would be inextricably merged in the store.

The visual erasure process is nicely illustrated by an experiment that is identical in every respect to the experiment just described, except for the replacement of the black bar marker by a black circle like that shown in *Figure 9.* The effect of such a circle on the letter it follows is very different from that of the bar. This effect, which we call erasure, is illustrated in *Figure 13* where performance with bar and circle may be compared. It is seen that the curves begin together and end together, with performance in the 'circle' experiment significantly poorer between.

What is the essential difference between the circle and bar that results in such a great difference in performance? It is primarily a matter of distance. If the bars were placed close to the letters, parts of the letters near the bar would be rendered illegible. In the retention experiment, the bar was carefully set far enough away to avoid this unwanted effect.

We are inclined to consider the effect of the circle on a preceding letter—space does not allow elaboration—as a quick substitution of the circle for the stored letter. The shape of the erasure curve may then be interpreted as follows:

(1) High performance when circle follows immediately after array is due to simple temporal averaging. This results in letter and circle being effectively superimposed which should not affect letter legibility.

(2) Lower performance at slightly longer delays can be attributed to the change from the superposition (averaging) condition to the erasure condition.

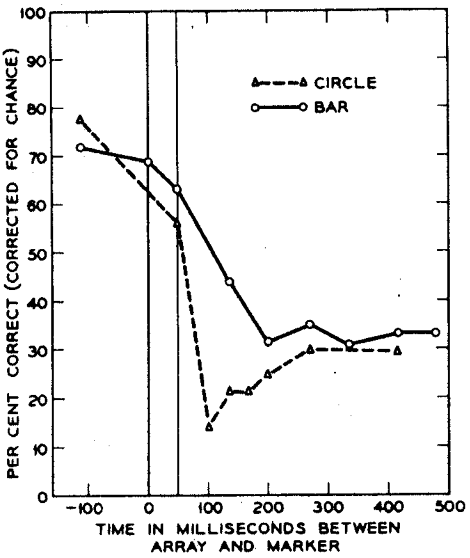

Figure 13. Erasure (first example)

(3) The slow rise from the minimum with further increases in delay of the circle is attributable to the increased time available to read the letter before it is erased.

(4) At the longest delays the circle arrives after the storage has decayed. Therefore it no longer erases, but acts only as a marker. Performance at these times simply measures the number of letters in the more permanent memory.

The suggestion that two closely-timed stimuli are perceived as superposed was verified by repeating the erasing experiment using a circle filled with grid lines. As shown in *Figure 14*, such an 'eraser' is more disturbing than an unfilled circle for times when superposition might hold (<100 msec), but not more disturbing with longer time separations.

We should mention that preliminary work has been done in which eight letters are exposed to one eye and an unfilled circle is delivered to the other. Erasure clearly occurs under these conditions, but we have not determined how binocular erasure compares with monocular.

3. *Read-out*

As we have already pointed out in discussing the bar-marker experiment, the process of detecting the presence of a marker and reading the marked letter undoubtedly takes some time. If the time required for this process can be measured, we shall have a more accurate idea of the time duration of the storage. A method for measuring this time is available provided that our conclusions about the action of the circle in erasing are correct. Suppose we present simultaneously an array and a bar marker pointing to one letter in the array, and then a short time afterwards present a circle around the marked letter. If the circle indeed removes the marked letter from the subject's storage, his performance under these conditions will measure how well he can detect the marker and read the letter given only the time interval between the onset of the array and marker, and the onset of the circle. The results of such an experiment are shown in *Figure 15.* It is seen that when circle follows by more than 100 msec, performance rises rapidly as a function of the time allowed for detecting the marker and reading the designated letter. This task clearly takes a significant amount of time, as much as 250 msec before maximum performance is attained. Performance when circle follows by less than 100 msec is very much like that obtained in the erasure experiment using circle without marker.

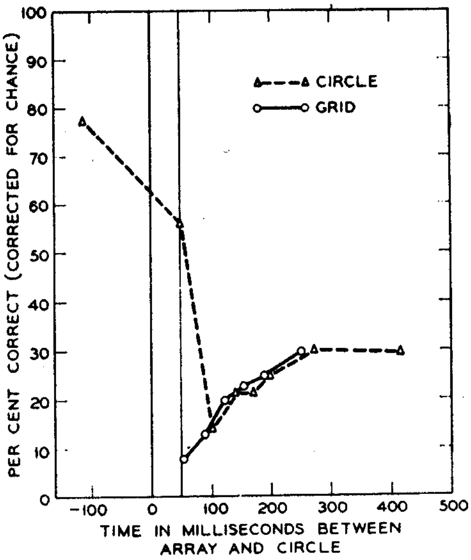

Figure 14. Erasure (second example)

4. *Storage time*

In discussing the bar-marker experiment we offered two reasons why it would not be correct to interpret the bar-marker performance curve as representing the short term storage decay: (*a*) the measured decay contains components of permanent memory, and (*b*) the curve does not take into account the detection and read-out time that we have just measured. In this section we shall return to our analysis of the bar-marker experiment and attempt, by correcting for these two factors, to arrive at an estimate of the effective visual storage time.

Performance in the bar-marker experiment is undoubtedly the result of two different kinds of performance on the part of the subject as shown by the fact that the decay curve does not fall to zero. First, there is a non-selective read-out which occurs when the array is exposed but *before* the marker is perceived; second, a selective process which occurs *after* the marker is perceived when the subject has been cued to direct his attention to the single required letter. Although we did not point it out earlier, we have already measured this 'before-marker' component, for in the erasure experiment (*Figure 13*), the circle plays the peculiar role of both 'marker' and 'eraser'; it erases the letter it marks. Performance in that experiment must represent only what has been read into the permanent memory *before* appearance of the marker (circle) because read-out from the short term memory is impossible after appearance of the marker.

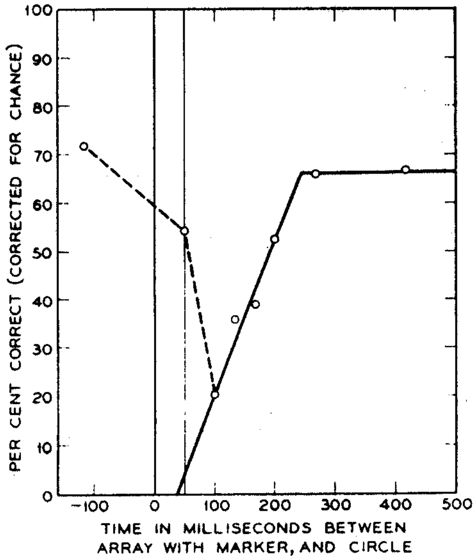

Figure 15. Read-out

The 'after-marker' component can be derived by subtracting out the 'before-marker' component from the whole (bar-marker curve). This derivation is illustrated in *Figure 16.* In *Figure 16a* the bar-marker curve is shown together with the erasing curve which has been extrapolated to zero. *Figure 16b* shows the curve derived by subtracting one from the other. The subtraction used is not a simple algebraic one. If, by chance, the subject reads the correct letter before appearance of the marker, designation by the marker cannot improve performance. Therefore it seemed more reasonable to solve for P_A (the after-marker component) from the equation $P_T = P_B + (1 - P_B)P_A$ where P_B is the before-marker performance component and P_T is the total, the performance measured in the bar-marker experiment.

If it be assumed that the derived decay curve in *Figure 16* represents the 'after-marker' component of performance, and that this component of performance is limited by the time available to detect the marker and read

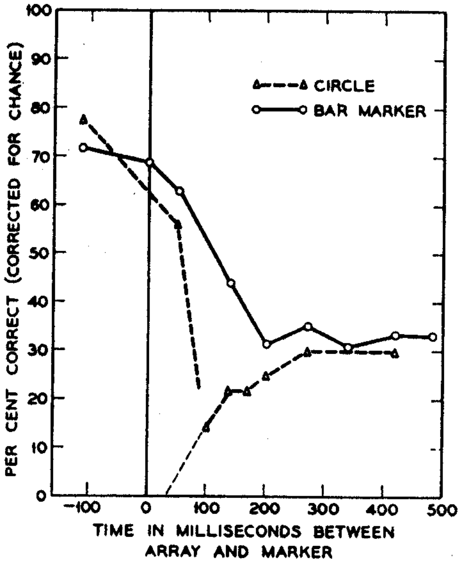

Figure 16a. Retention and erasure

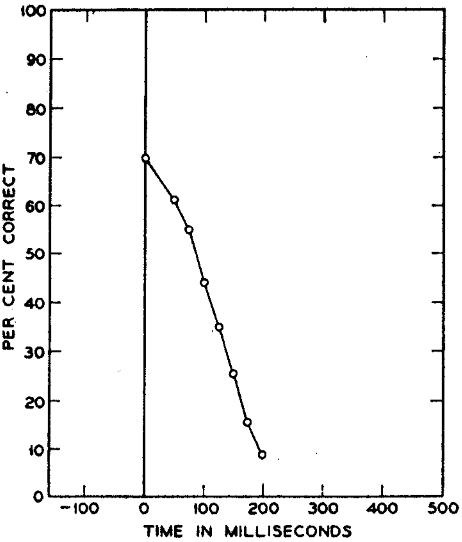

Figure 16b. Derived 'after-marker' performance

the letter before decay of the storage, the storage time can be obtained by adding the experimentally determined read-out times for each level of performance (*Figure 15*) to the times corresponding to the same level of performance in *Figure 16*. The result of this operation is shown in *Figure 17*. The points shown on the right, each of which is an independent estimate of the storage time, approximate a vertical line surprisingly well. The estimated storage time is 270 msec.

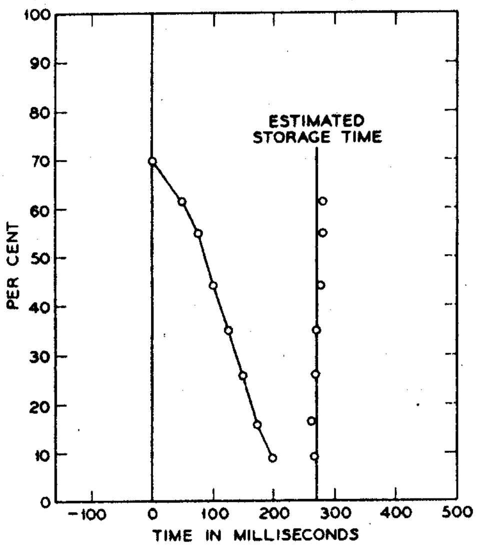

Figure 17. Storage time

SUMMARY AND CONCLUSIONS

The work described in this paper shows that the visual process involves a buffer storage of relatively high capacity that can take in information virtually instantaneously and retain it to permit its relatively slow utilization. By making use of sampling techniques and the erasure characteristics of the storage, the following properties were demonstrated:

(1) The capacity of the storage is high compared to the 20–25 bits obtained in span of immediate-memory. The experiments showed that 70 bits could be stored, but this is still only a lower bound on the capacity since different arrangements might well have produced higher figures.

(2) The decay time depends upon pre- and post-exposure conditions as well as on the exposure itself. The measured decays varied from ¼ sec to several seconds.

(3) The spatial resolution of the storage is disturbed when too many letters are put in. A 2 × 8 array is enough to demonstrate this effect.

(4) The storage is erasable; new information erases previously stored information.

(5) The storage may be read out rapidly up to about five items, the initial rate during such a reading burst being of the order of 100 items per sec.

REFERENCES

[1] Sperling, G. 'The Information Available in Brief Visual Presentations'. *Psychol. Monogr. 74,* No. 11, Whole No. 498 (1960)

[2] Baxt, N. 'Ueber d. Zeit welche nötig ist, damit ein Gesichtseindruck zum Bewusstsein kommt', *Pflüg. Arch. ges. Physiol.,* 4 (1871) 325–336

DISCUSSION

N. M. Blachman: What is the effect of eye movement in between presentation of the letters and presentation of the marker?

J. O. Ackroyd: In everyday life we can remember numbers of 5, 6 or 7 digits which have been seen or heard quite briefly, for example telephone numbers or sums of money. Can this be reconciled with the authors' finding that the number of symbols memorized is asymptotic to approximately 4·5? In their experiments they employed a larger alphabet; perhaps also the motivation was less than in real life situations.

E. Averbach in reply to Dr. Blachman: Although we did not study their effect systematically, we know from casual observation that eye movements between presentation of the letters and presentation of the marker can have a very disturbing influence on observer performance in this kind of experiment. The effect of such movements, from the observer's point of view, is to displace the marker relative to the letters. In fact, if the movement is just right, it appears to the observer that the marker is pointing to a different letter from the one designated by the experiment. We eliminated these eye movements from our experiments by having the observer fix his eyes just before each exposure on a fixation point in the centre of the screen. It is well established that trained observers can maintain fairly accurate fixation for a great deal longer than the half-second fixation required in these experiments.

G. Sperling in reply to Dr. Ackroyd: I should first like to point out that the finding that only 4 or 5 letters can be reported after a brief exposure dates back to the last century[1]. My contribution has been, I think, to show that these few letters can be arbitrarily selected from a considerably larger number of letters which are available momentarily during and shortly after the exposure. The reason that, in everyday situations, the number of items remembered seems to be larger than four is that everyday situations usually do not represent brief exposures. For example, Deininger[2] found that in order to remember a seven digit telephone number correctly, most people actually require two glances, in each of which they learn three or four digits. In the text, I pointed out that while four items can be read and remembered in a fraction of a second, the learning of additional items requires substantially more time and effort. In contrast to the extremely rapid learning of the first few items, the slow learning of additional items probably depends on complicated coding and/or association processes. These mnemonic processes are undoubtedly markedly influenced by motivation, as is all difficult learning. However, the 4·5 item limit (individual differences ranged from four to six items) seems to be quite independent of motivation.

REFERENCES

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

Erdmann, B. and Dodge, R. *Psychologische Untersuchung ueber das Lesen,* Halle, M. Niemeyer, 1898

[2] Deininger, R. L. 'Human Factors Engineering Studies of the Design and Use of Pushbutton Telephone Sets', *Bell System Tech. J.,* 39 (1960) 995–1012