XXIV. COGNITIVE INFORMATION PROCESSING*

Academic and Research Staff

Prof. W. L. Black
Prof. M. Eden
Prof. T. S. Huang
Prof. F. F. Lee
Prof. S. J. Mason
Prof. W. F. Schreiber
Prof. D. E. Troxel
Dr. K. R. Ingham
Dr. P. A. Kolers
Dr. P. H. Liss
Dr. O. J. Tretiak
C. L. Fontaine
E. R. Jensen
G. L. Wickelgren

Graduate Students

G. B. Anderson
T. P. Barnwell III
B. A. Blesser
A. L. Citron
D. P. De Wan
H. D. Evans III
A. Gabrielian
R. V. Harris III
D. W. Hartman
H. P. Hartmann
G. R. Kalan
R. W. Kinsley, Jr.
W. H. Lee
J. I. Makhoul
L. C. Ng
D. L. Peterson
G. F. Pfister
D. S. Prerau
G. M. Robbins
C. L. Seitz
D. Sheena
R. M. Strong
G. A. Walpert
J. A. Williams

A. OPTIMUM THRESHOLD LEVEL FOR THE READING MACHINE OPAQUE SCANNER

A flying-spot opaque scanner[1] has been constructed to enable computer-controlled scanning of printed characters, including books, magazines, and newspapers. In this scanner the cathode-ray tube beam is positioned at an x-y coordinate, and the output from the photomultiplier tubes is integrated for a fixed time. A comparison of the integrated output with a threshold voltage is made to determine whether the point is black or white. At present, this threshold is adjusted by a sighted operator to optimize the scanner display. As the intended application is for use in a reading machine for the blind, it is obviously desirable to provide an alternative method for adjustment of the threshold to an optimum value.

The first step toward this end was to determine that there is indeed an optimum value. Perfect printed matter would have only two possible reflectivities, a high level corresponding to sections of the page which have not been inked, and a low level corresponding to the black sections. The density function of the reflectivity of points on the page would in this case consist of two impulses. If the printing process were perfect, the optimum threshold could be anywhere between these two levels to obtain a perfect discrimination of black from white. Actual printing should contain a fairly continuous distribution of reflectivities from the darkest printing to the brightest white. Most points should be either black or white, which leads to the suspicion that there must be a relative minimum between these two maxima in the density function. If this is true, it

*This work was supported principally by the National Institutes of Health (Grant 1 PO1 GM-14940-01), and in part by the Joint Services Electronics Programs (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DA28-043-AMC-02536(E).

follows that this minimum should be an optimum threshold value because it has the minimum occurrence of all levels in its vicinity. In other words, there would be minimum ambiguity about whether a spot is black or white if the threshold were set at this level. The existence of an optimum threshold level was verified by measuring a cumulative distribution function (CDF) and deriving the density function from it. The CDF was measured by counting the number of black points in a field as a function of threshold level. Three CDF's and their associated density functions are shown in Fig. XXIV-1. Figure XXIV-1a and -1b are for two different sections of the same page of a dull stock magazine

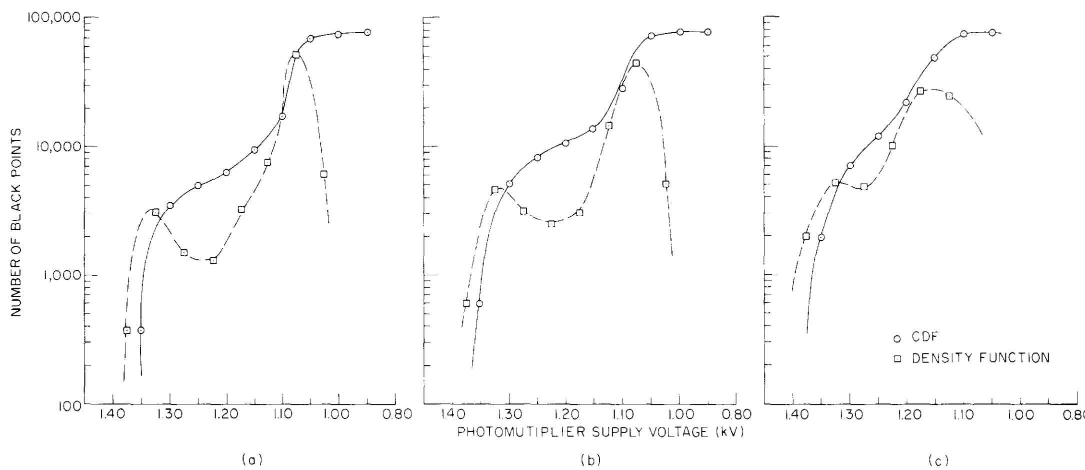

Fig. XXIV-1. Three cumulative distribution functions and their associated density functions.

style printing sample, and Fig. XXIV-1c was obtained from a sample of newsprint. The abcissa for these curves is the voltage on the photomultiplier tubes in the opaque scanner. As the sensitivity of the photomultiplier tubes varies as the twelfth power of their supply voltage, these graphs are actually plotted against the twelfth power of the threshold level. The reflectivity is inversely proportional to the threshold level. While the threshold increases to the left, the reflectivity increases to the right. All of these graphs show a characteristic relative minimum — the optimum threshold level — and the two expected maxima, the smaller being the median black reflectivity, and the larger the median white reflectivity. It is fortunate that the position of the minimum did not change for different sections of the same page (see Fig. XXIV-1a and -1b), even though the relative number of black and white points did change.

An algorithm has been devised to automate the determination of the optimum threshold level. As there is a great difference between the thresholds for the two maxima shown in Fig. XXIV-1, we assume that the range of variation of the threshold is located between these two maxima. The first step is to set the threshold at the highest possible level within its range and to determine the number of black points in the field. The threshold is decremented a standard amount A, and the number of black points is again determined. This process is repeated, and a set of second differences is calculated. The occurrence of the first negative second difference means that the optimum threshold has been passed. If the decrements of the threshold have been sufficiently small, then this last setting of the threshold is a good approximation to the optimum threshold. Application of this algorithm resulted in an optimum threshold that was within ±A of the threshold produced by the best judgment of a sighted operator.

Another algorithm, which has not yet been tried, would be to look for that threshold in which the total number of black points varies least. This variation in the total number of black points should be monotonic, with the mean-square difference between several CDF's obtained for the same threshold value. An advantage of this new algorithm would be that it does not depend directly on the change in threshold which occurs between successive measurements, as does the first algorithm that has been described.

D. E. Troxel, G. F. Pfister

References

1. C. L. Seitz, "An Opaque Scanner for Reading Machine Research," S.M. Thesis, Department of Electrical Engineering, M.I.T., January 1967.

B. READING MACHINE STUDIES

The grapheme-to-phoneme translation procedure reported in Quarterly Progress Report No. 80 (page 225) has been further refined and implemented as an integral part of the reading machine.

Written English sentences can now be used as input to the PDP-1X computer which performs the translation of the orthographic form into phoneme strings, and transmits the resultant strings to the TX-0 computer. TX-0 performs additional processing on the strings and generates sequences of control commands to operate a terminal analog speech synthesizer. A syntax analyzer is planned for the future to resolve certain kinds of ambiguities and to assist in the design of intonation contour. The current speech output derives intonation parameters from lexical stresses and punctuation marks only.

Following is a description of the translation procedure as it is now implemented.

1. Definitions

A lexeme is a word in spoken form.

A morpheme is in the spoken domain and is the smallest unit that has a meaning in the language. (Lexemes and morphemes may be symbolically represented by phonemic transcriptions.)

A free morpheme is a morpheme that can exist by itself. A bound morpheme is a morpheme that is not a free morpheme.

A lexeme is a free morpheme, or a bound morpheme combined with a lexeme, or a free morpheme combined with a lexeme. The rules of combination are called morphophonemic rules.

A lex is a word in written form.

A morph is in the written domain and is the smallest unit that has a meaning in the language.

A free morph is a morph that can exist by itself. A bound morph is a morph that is not a free morph.

A lex is a free morph, or a bound morph combined with a lex, or a free morph combined with a lex. The rules of combination are called morphographemic rules.

2. Lex-to-Lexeme Mapping

The relation between lex and lexeme is illustrated in Fig. XXIV-2. The mapping of a morph into the corresponding morpheme is performed through a morph lexicon look-up.

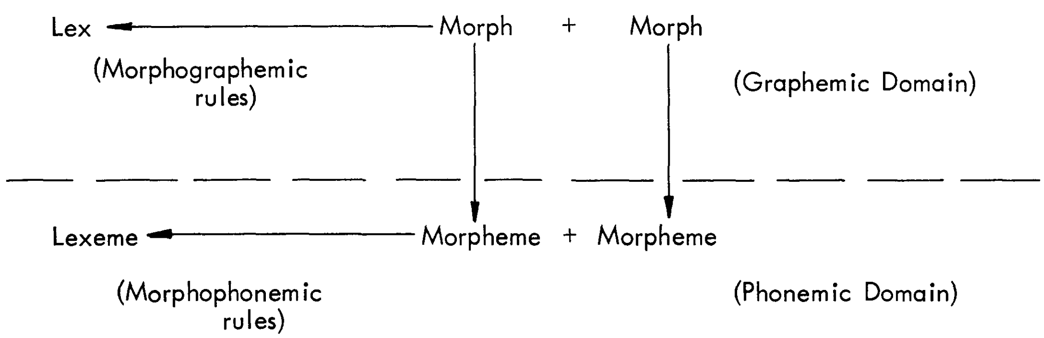

Fig. XXIV-2. Lex-to-lexeme mapping.

The morphophonemic rules are simplified to include only simple concatenation and those dealing with [ed] [s] suffixes.

The morphographemic rules are listed in Figs. XXIV-3 and XXIV-4.

$(\alpha)_a \rightarrow T$	(1)	$(\alpha)_r (\beta)_{s_c} \rightarrow (\alpha\beta)_r$	(7)
$(\alpha)_r \rightarrow T$	(2)	$(\alpha\text{'e'})_r (\beta)_{s_v} \rightarrow (\alpha\beta)_r$	(8)
$(\alpha)_r (\beta)_r \rightarrow (\alpha\beta)_r$	(3)	$(\alpha C)_d (\beta)_{s_v} \rightarrow (\alpha C C \beta)_r$	(9)
$(\alpha)_p (\beta)_r \rightarrow (\alpha\beta)_r$	(4)	$(\alpha C)_o (\beta)_{s_v} \rightarrow (\alpha C C \beta)_r$	(10)
$(\alpha)_r (\beta)_{s_t} \rightarrow (\alpha\beta)_a$	(5)	$(\alpha C)_o (\beta)_{s_v} \rightarrow (\alpha C C \beta)_r$	(11)
$(\alpha)_r (\beta)_{s_v} \rightarrow (\alpha\beta)_r$	(6)	$(\alpha\text{'y'})_r (\beta)_{\bar{p}} \rightarrow (\alpha\text{'i'}\beta)$	(12)

Fig. XXIV-3. Most often used morphographemic rules of English.

a : a free morph which never combines with others, such as [me].

d : a free morph which must have its final letter doubled when followed by a vocalic suffix, such as [hit].

o : a free morph which may have its final letter doubled when followed by a vocalic suffix, such as [model].

r : a free morph which must not have its final letter doubled when followed by a vocalic suffix, such [house].

s_v : a bound morph which is a vocalic suffix, such as [-able].

s_c : a bound morph which is a consonantal suffix, such as [-ness].

s_t : a bound morph which is a terminal suffix.

p : a bound morph which is a prefix, such as [un-].

Fig. XXIV-4. Definition of subscripts.

At present, a 3000-morph lexicon (corresponding to approximately a Fourth Grade level) is used. Lexes for which simple concatenation of morphemes would lead to serious pronunciation errors are arbitrarily moved to morph status.

F. F. Lee, T. P. Barnwell, E. R. Jensen

C. THE FIRST HUNDRED MILLISECONDS OF VISION

The experiments described here are preliminary studies of the way in which the visual system of an adult human operates when perceiving simple and complex displays of letters. Special interest is devoted to the question of what aspects of character recognition are accomplished in parallel and what aspects require serial processing. We find that the detection of from one to six letters along a row is accomplished in parallel, while identification of even a few random letters is done by a serial scan.

Tachistoscopic exposures of visual stimuli normally result in a positive afterimage that lasts at least 200 msec (Sperling[1]). Reducing the intensity or duration of the stimulus may greatly affect the clarity or distinctness of the percept, but the viewer still seems to have at least 200 msec in which the image is available for processing. In order to study the details of visual processing, better control of the processing time is clearly necessary. An effective way of controlling the processing time is to present a masking stimulus after the target (Sperling[2]). By varying the time before the mask is applied, one can analyze the stages in the development of a percept within the first 200 msec.

The interpretations reported here of the results obtained with masking are based on the assumption that perception of the mask prevents further processing of the test stimulus. In other words, we assume that at some point during the processing of the mask, further analysis of the test stimulus is impossible. All that the subject can do at that point is try to remember the results of the processing already accomplished and report them. It is not at all obvious how this assumption can be directly tested.

1. First Part

The stimuli used in these experiments were capital letters, 1/8 inch high, typewritten with a plastic ribbon on a white index card to maximize contrast. The stimuli were viewed at a distance of one meter; thus each leter subtended 10' of visual angle. There was a 5' space between the ends of adjacent letters. The mask consisted of three long rows of 0's superimposed on N's, with no spaces between adjacent letters. Luminance was approximately 5 mL.

The first experiments describe the effects of visual masking on simple character identification and on the number of letters that could be detected along a single horizontal row. A single subject was tested intensively. The test stimuli consisted of 21 cards, each containing a single horizontal row of either I's, 0's, or X's. There were either 1, 2, 3, 4, 6, 8, or 10 letters on each card. Ten 0's subtended 2.4°; 10 X's, 2.0°; 10 I's, 1.2° of visual angle. The letters on the different cards were positioned so that they had a common center.

The procedure for a single trial was typically as follows: (i) The subject looked through an eyepiece with his better eye and indicated when he was ready. (ii) When the experimenter pressed a button, a fixation field with three dots appeared for 1.2 seconds. (iii) The test stimulus then appeared for a variable amount of time. (iv) The masking stimulus was presented for 0.5 second. There were no blank intervals between stages 2, 3, and 4. The subject then wrote down a letter for every letter perceived, with the instruction that a letter should be written down even if an unidentifiable haze or smudge was seen. Under each reported letter, the subject was required to make a clarity judgment of 1-5 indicating the apparent distinctiveness of that letter. One meant unidentifiable smudge, low contrast, no distinctive features; five meant perfectly clear form, high contrast, all features very distinctive. These judgments seemed fairly easy to make.

A single test consisted of all 21 cards (1, 2, 3, 4, 6, 8, or 10 I's, O's, or X's) presented for the same test duration but randomized as to number and kind. After two days of practice, great similarity in results was found from one test to another, so only a few tests for each time condition seemed necessary.

a. 30 msec, no mask (just a dark field followed the 30-msec test stimulus). Three tests were given. There was 100 per cent accuracy in identifying the letters as either I's, O's, or X's. The clarity judgments were all fours and fives. The numbers inside the graphs (Fig. XXIV-5) describe the number of letters reported for each trial. For example, looking at Fig. XXIV-5a, for the 9 trials in which 6 letters were presented (3 trials in each of the 3 tests), 5 letters were reported once, 6 letters were reported 7 times, and 8 letters once. Figure XXIV-5a shows that accuracy in identifying the number of letters was generally quite good, with the variability gradually increasing with an increase in the number

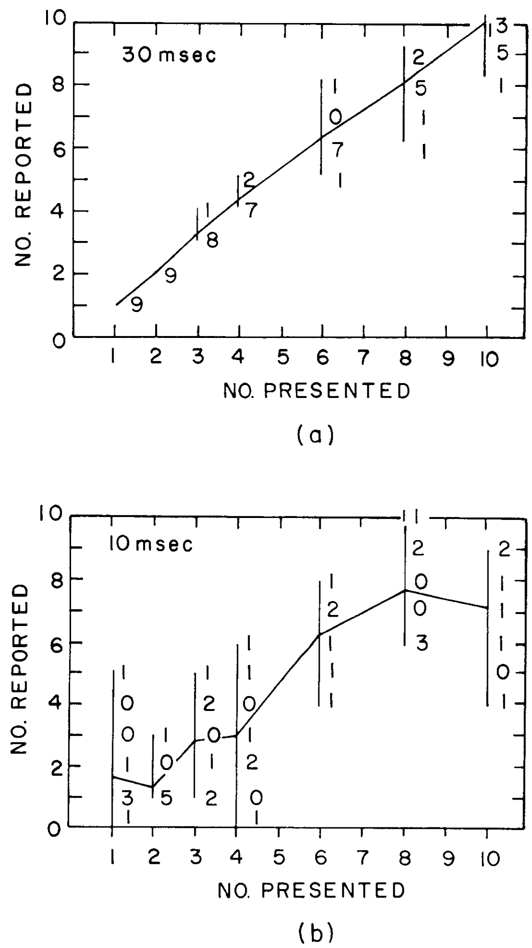

Fig. XXIV-5.

Number of letters reported as a function of number of letters presented and duration of test stimulus. No mask followed the stimulus.

presented. There were no differences between the I's, 0's, and X's.

b. 10 msec, no mask. Two tests were given. There was 74 per cent accuracy in identifying the letters as I's, 0's, or X's. The clarity judgments were usually one, sometimes two. The number of letters reported for a trial is given in the figure even if the subject misidentified the letters on that trial. Figure XXIV-5b shows that a highly variable number of letters was reported for any given number of letters presented, and in this linear graph the variability is approximately the same whether 1 or 10 letters were presented. There were many "false alarms" when only a few letters were presented. These false alarms should, I think, be called hallucinations, since the subject claimed that she was not uncertain about the presence of the smudges when reporting the perception of non-existent figures.

When 8 or 10 letters were presented, the subject reported an average of 7 or 8 letters. The exact number was highly sensitive to whether I's, X's, or 0's were presented. The top values in these conditions are for the I's, the bottom for the 0's, with the X's in between. This difference is probably due to differences in acuity, since the I's spread 1.2°, and the 0's 2.4° of visual angle.

c. 30 msec followed by the mask. Three tests were given. The results were more like the 10-msec than the 30-msec condition without the mask. There was 68% accuracy in identifying the letters as I's, 0's, or X's. Clarity judgments were almost all ones.

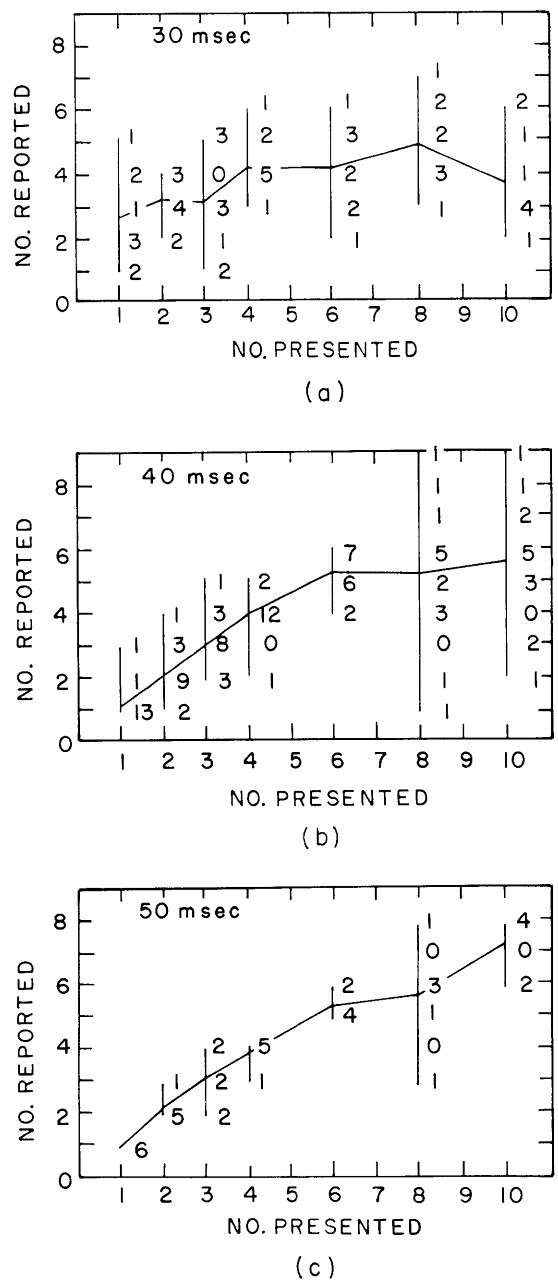

Fig. XXIV-6.

Number of letters reported as a function of number of letters presented and duration of test stimulus. Stimulus immediately followed by mask.

Figure XXIV-6a shows that the mask greatly affected the number of letters reported. Although the variability around the mean values was similar to that in the 10-msec no-mask condition, there was now almost no relation between the number of letters reported and the number presented. The average for 1, 2, and 3 letters presented was 3.0 reported, indicating many hallucinations, while the average for 6, 8, and 10 letters presented was only 4.3 reported, indicating that many of the letters were not perceived. The decrease from the 10-msec no-mask to the 30-msec mask conditions in the number of letters reported when I's were presented was somewhat greater than the decrease in the number reported when 0's were presented. Therefore the effect of the mask in limiting the number of letters perceived was not due to an augmentation of the acuity factor. It is emphasized that these misperceptions of the number of letters presented occurred despite considerably greater than chance accuracy in identifying the letters as I's, 0's, or X's (68 per cent). Almost all of the confusions were between X's and either I's or 0's; I's and 0's were confused with each other on only one trial. There was no consistent relation between the number presented or number reported and probability of error in the identification task.

d. 40 msec, with mask. Results were so interesting that 5 tests were given, and essentially the same results were found on each test. There was 85 per cent accuracy in identifying the letters as I's, 0's, or X's. Clarity judgments were most frequently two. Figure XXIV-6b shows that the accuracy in reporting the number of letters was greatly improved over the 30-msec condition when the number of letters presented was 1, 2, 3, 4, or 6. Most of the reports in these cases were perfect; the errors were relatively slight, and fewer hallucinations were reported. These differences between 30 and 40 msec with the mask seem to indicate parallel functioning in recognizing the presence of from 1 to 6 letters along a row.

The results when 8 or 10 letters were presented were very different. The average number of letters reported was the same as that for 6. The variability, however, was very great. Sometimes few letters were perceived, sometimes relatively many. The difference between the number of I's and the number of 0's reported was again less than that for the 10-msec no-mask condition. The results as a whole indicate that under these stimulus and time conditions, it is relatively easy to accurately detect up to 6 letters in a row, but very difficult to perceive more. Modifications of this conclusion are discussed below.

c. 50 msec, with mask. Two tests were given. There was 91 per cent accuracy in identifying the letters as I's, 0's, or X's. The clarity judgments were mostly three. Figure XXIV-6c shows that there was an increase in accuracy and decrease in variability in number of letters reported all across the range. The mask still limited the number that could be perceived, since the largest number reported was only 8. There were no longer any differences at all between the number of I's and 0's reported, as one would expect from the generally increased clarity and consequent reduction of acuity as a significant factor.

Conclusions

1. When stimuli are perceived very hazily (10 msec, no mask; 30 msec, with mask) there is considerable hallucinating. When stimuli become somewhat clearer, hallucinating stops (40 and 50 msec, with mask). A possible explanation for this result is that when the visual system is busy creating form out of weak signals, it is more likely to create an "expected" form, despite no signal, than if it is basing its creations on strong signals. It is doubtful that these hallucinations are simply due to random noise, since they were perceived with the same relative spacing as real letters. Thus one's expectations determine where the forms are created. A further experiment demonstrates this point much more forcefully.

Two sets of stimuli were presented under the 30-msec with mask condition. One was I's, 0's, and X's as above; the other was various numbers of random letters, each letter 5' distant from the adjacent one, just as for the I's, 0's, and X's. In the random-letter condition the subject was not asked to identify the letters but rather simply to indicate the presence or absence of more or less clear smudges. Abnormally large spacing between perceived hazes were to be indicated on the data sheet by dashes.

In general, there were no differences in the total number of smudges reported between the two conditions. The number of reported items was almost randomly related to the number presented. Clarity judgments were mostly 1, some 2.

The interesting data are those on spacing. The first time each condition was tested, no abnormally large spaces between perceived letters were ever reported for the I's, 0's, and X's. Extra spacing was reported, however, in seven out of the 21 presentations in the random-letter condition. Apparently, there is some interference between adjacent letters if they are different. The following repetition shows that this interference is primarily central and not sensory in origin.

The second time each condition was tested, the subject, as it happened, assumed that the I's, 0's, and X's were random letters, and vice versa. When random letters were presented, the subject wrote down rows of I's, 0's, and X's and put extra spacings in only 2 out of the 21 presentations. There was no question in her mind that the stimuli really were I's, 0's, and X's. This was so, despite the fact that when I's, 0's, and X's were really presented accuracy was well above chance (68%). When I's, 0's, and X's were presented, the subject now considered them as random letters and saw extra spacings in 9 out of the 21 presentations. These results seem to show that one's expectations greatly influence both the kinds of poorly defined smudges that are constructed and the presence or absence of interference between these smudges.

2. The great increase in accuracy in reporting the number of letters when the number presented was 1, 2, 3, 4, or 6, and when stimulus duration was increased from 30 to 40 msec, demonstrates that the accurate detection of from 1 to 6 letters in a row is done essentially in parallel. The accurate detection of more than 6 items in a row requires further processing, and is not yet complete at 50 msec.

In order to test more strongly the notion that it is the number of letters — and not how central they are in vision — which determines how many can be perceived at 40 msec with the mask, another experiment was performed. (We already have considerable support for this notion from the finding that approximately 5 out of 6 0's are reported, while only approximately 5 or 6 out of 10 I's are reported, despite the fact that the six 0's subtend a larger visual angle than the ten I's.) The following experiment illustrates what is perceived when 2 horizontal rows of 0's are presented.

The stimuli were two rows of 0's, 1, 3, 6, or 10 letters per row. There was only a $\frac{1}{16}$th inch space between the rows, which was the same distance as that between adjacent letters on a single row.

One test consisted of 24 stimulus presentations. Sixteen presentations included all combinations of 1, 3, 6, and 10 letters on each row. There were also 2 stimulus presentations of a single row (one top and one bottom) of 1, 3, 6, and 10 letters. Four tests were given.

The main result is that in these 4 tests the subject never reported seeing more than 7 letters. The number of letters perceived when 2 rows were presented was, on the average, no more than the number perceived when only 1 row was presented. Clearly, the limitation on the number of letters perceived was due to the brief processing time available and not upon acuity. The results also demonstrate that what is apprehended under these conditions is very dependent on exactly where one is attending (top or bottom row), as well as the number of letters presented. When the bottom row was primarily attended to, as happened naturally during the first 3 tests, up to 6 letters were fairly accurately seen on that row, and usually only 1 letter was perceived from the top row. When the subject was instructed to attend primarily to the top row, as in the fourth test, the reverse results were usually obtained.

Later tests in which the subject was instructed to attend to all parts of the field and not concentrate on any row (wide attention) produced very different results. Up to 12 0's were now reported on a single trial, 8 or 9 from one row and 4 or 3 from the other, but there was a very marked reduction in the clarity of all letters. At this stage of testing, the clarity judgments were usually 4 when the subject concentrated on a row, but only 2 in the wide-attention condition. Furthermore, there was a considerable increase in hallucinatory activity in the wide-attention condition. An average of 4.5 nonexistent letters were reported during the 24 presentations in the concentrated-attention conditions, and 17 extra letters in the wide-attention condition.

These results demonstrate that the previous findings with the 40-msec condition, in which the fairly accurate detection of up to 6 letters in a row occurred along with an inability to consistently detect any more letters, depends upon the subject concentrating narrowly upon that row with the intention of obtaining maximum letter clarity. If the attention is relaxed and more dispersed, more letters may be detected but less processing is available for each leter, thereby resulting in decreased clarity and increased hallucinations. Thus, when a limited amount of processing time is available to the visual system, it may be used in radically different ways.

2. Second Part

We have shown that a parallel process is involved when an adult detects the existence of from 1 to 6 letters in a row, but that some further processing is necessary for detecting the presence of more items. The results of Sperling[2] have indicated, however, that identifying random letters (which is different from merely detecting their presence) demands a serial scanning process requiring approximately 10 msec per letter.

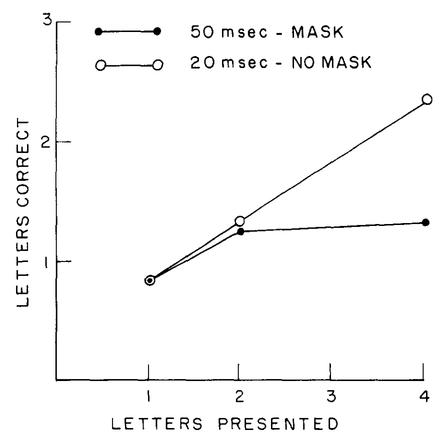

Fig. XXIV-7.
Number of letters correctly identified as a function of number of letters presented. Mask vs no mask, letter clarity approximately equal.

When studying the ability to identify random letters, either full- or partial-report techniques may be used. In a full-report condition, the subject must attempt to identify each of the letters. He knows in advance how many are presented, and is instructed to guess if uncertain. Also, if more than one letter is perceived when only one is presented, as frequently happens when the stimulus is presented very briefly, only one is reported.

In the present partial-report procedure, adapted from Averbach and Sperling,[3] a bar is presented over one of the letters in each stimulus array. The subject must report the letter under the bar. The subject was also required to guess at the letters on either side of the bar, but he was asked not to let that subsidiary task interfere with the main goal of identifying the letter directly under the bar. Pilot work indicated that, in fact, this procedure had no interfering effects.

It is necessary to know whether the improvement that one observes in masking experiments in identifying random letters with increasing durations of test stimuli is really indicative of a scanning process. It is possible that the improvement is due simply to the increased over-all clarity of the letters. To test this possibility, a comparison was made between a mask and a no-mask condition on the identification of random letters of equal clarity. Two experienced subjects were tested at both 20 msec without a mask and at 50 msec with a mask, conditions in which clarity is approximately equal, but not perfect (clarity judgments were usually 3 or 4). Either 1, 2, or 4 letters were presented and full report was called for. The results are presented in Fig. XXIV-7. Both conditions gave identical results when 1 or 2 letters were presented, but not with 4 letters. In the no-mask condition, additional letters were identified when 4 letters were presented, while in the mask condition no improvement between the 2- and 4-letter conditions was observed. It is clear that identifying more than 2 letters requires some scanning process, and the improvement that is observed by increasing the time before mask onset cannot be accounted for simply by referring to the mask's effect on stimulus clarity.

Full Report

Sperling[2] has reported the results of an experiment using the full-report technique on 2 subjects. From 2 to 6 letters were presented, and only items identified in their correct position were counted as correct. The main results were: (i) The number of letters correctly identified at a fixed stimulus duration is independent of the number of letters presented; and (ii) there seems to be a scanning rate for letter identification of approximately one letter per 10 msec for the first 3-4 letters. The curve then becomes asymptotic between 4 and 5. The asymptote is interpreted as indicating a limitation on immediate memory.

In the following experiment a single subject was used. Either 1, 2, 4, or 8 letters were presented along a single row. There were two kinds of four-letter combinations: 4M (Massed) – the spacing between the letters was the same as that for the 2 and 8 conditions (5' of visual angle); 4D (Distributed) – the spacing between the letters was such that the total width was the same as in the 8-letter combination. Results for these 4-letter combinations were combined since, despite considerable differences in accuracy for different letter positions, there were no over-all differences in terms of totals. One test consisted of 8 trials all in a single condition. A condition may be something like 8 letters, 70 msec before the mask, full report. Full- and partial-report conditions were intermixed. After one day of pretraining, all conditions were run twice. Although there was some improvement from the first to the second series, results for the two series are combined.

Figure XXIV-8 shows striking differences from the full-report results reported

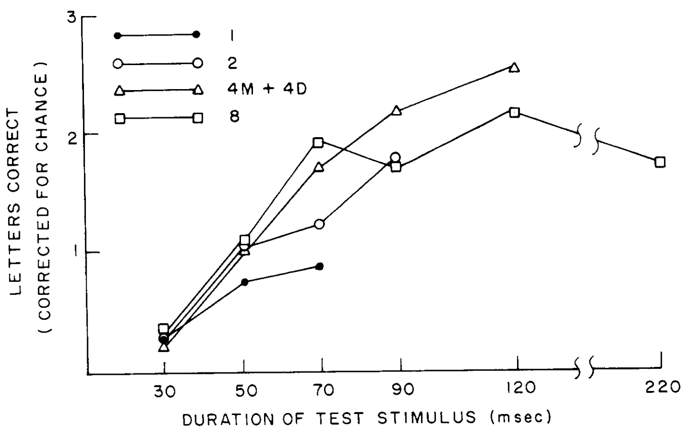

Fig. XXIV-8. Full report. Number of letters correctly identified as a function of number of letters presented and duration of the stimulus before the mask.

by Sperling. In the present experiment the maximum scanning rate for letter identification was approximately 20 or 25 msec per letter, not 10 msec. The asymptotic number of letters identified was less than 3, not between 4 and 5. The most likely explanation of these differences from Sperling's results is that I used smaller letters: Sperling's letters were 1.2° high; the letters here were 10' high. (Luminance does not seem to be the relevant variable, since Schiller (personal communication) has replicated Sperling's results with luminance similar to that used here, but with large letters.)

One may wonder whether using small, relatively hard-to-identify, letters really decreased the scanning rate from 10 to 20 or more msec per letter. It is quite conceivable that the subject in the present experiment still scanned at a "natural" rate of 10 msec per letter, with the smallness of the letters resulting in many letters being reported that were similar but not identical to the correct letters.

To check on the amount of analysis the subject gave to the letters that were not perfectly identified, I liberalized the criterion and counted as correct the 26 most frequent confusions (such as C for G or Q). When the results thus obtained were corrected for chance, all conditions showed on the average considerably less than one extra letter identified per trial. In fact, counting frequent substitutions as correct resulted in as great an improvement in the identification of a single letter as in the identification of 2, 4, or 8 letters. Apparently, close substitutions were given only for letters that were in the focus of attention. It seems, then, that a letter was typically attended to and analyzed for at least 20 msec before the next one was studied.

We thus come to the important conclusion that the visual system is able to adjust its rate of scan to the requirement of the task. If the letters are large and easy to recognize, it may spend approximately 10 msec on each letter, while if the letters are small and harder to recognize, it may spend 20 or more msec on each letter. The order in which the scan proceeds in the different conditions will be described and discussed below.

The fact that the asymptote, as well as the slope, is greatly affected by the ease of identifying the letters seems to indicate that the report depends upon visual and not only auditory memory. Presumably difficult-to-analyze letters take up more room in visual memory or are more readily interfered with by other letters than easy-to-analyze letters.

Partial Report

The partial-report task looks deceptively simple. Since only one letter must be accurately reported, memory limitations are disposed of. Despite this, in an experiment similar to the present one, using two rows of 8 letters each with the bar above or below the crucial letter, Averbach and Sperling[3] found that more than 200 msec was required before asymptote was reached. Apparently a search of approximately 15 msec per letter occurred before the bar was found. The purpose of the present partial-report experiment is to discover the conditions under which it is necessary for the visual system to search for the bar.

The results are shown in Fig. XXIV-9; the number of letters "available" (Sperling[1]), rather than per cent correct, is shown for comparison with full-report results

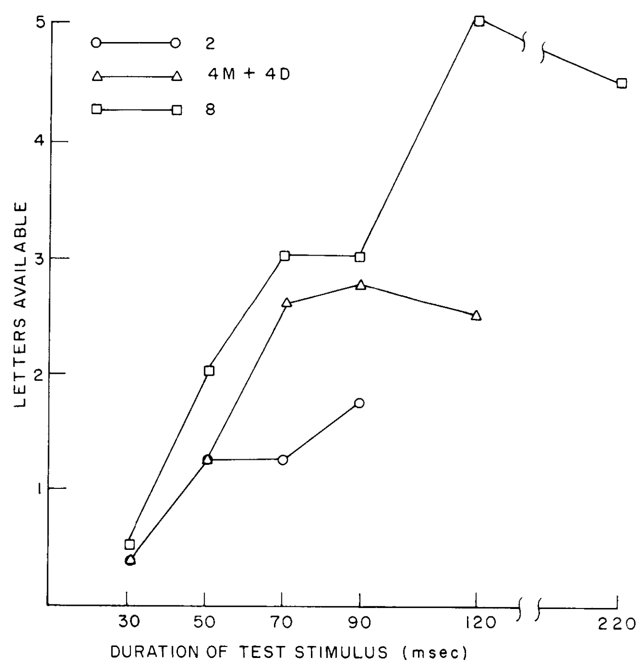

Fig. XXIV-9. Partial report. Number of letters available as a function of number of letters presented and duration of the stimulus before the mask.

(Fig. XXIV-8). (The proportion of trials correct in a condition was multiplied by 2, 4, or 8, depending on how many letters were presented in that condition.) The main findings are:

1. An advantage of partial report over full report was apparent only when 8 letters were presented. There were no differences in the two-letter conditions, and only a slight difference in the four. In these conditions the subject was not more likely to report a letter correctly, even if that was the only letter required. The subject's introspections were that only in the 8-letter condition was there a special act of attention involved in locating the bar and identifying the corresponding letter.

2. Counting substitutions and correcting for chance, 1-3 additional letters were identified in all the 8-letter conditions, except in the 30-msec presentation. There was considerable improvement with practice. In subsequent informal tests using 8-letter, 50-msec presentations, the subject correctly reported 4.5 out of 8 letters when substitutions were counted.

3. The letters that were correctly perceived in the 8-letter conditions were far from random. Figure XXIV-10 shows that for 50, 70, and 90 msec stimulus durations,

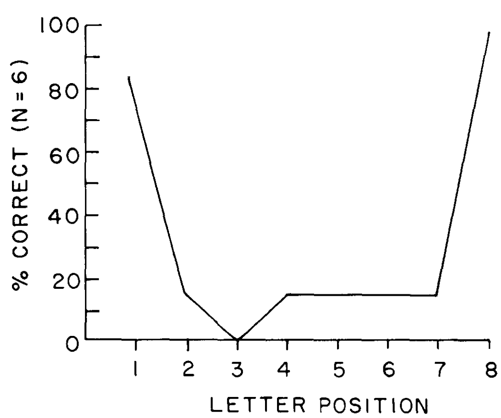

Fig. XXIV-10. Partial report. Accuracy in identifying the letters at each letter position. Eight letters presented; 50, 70, and 90 msec durations combined; both series combined.

most letters correctly reported (substitutions not counted) were at the ends. Neither end letter was correctly reported in the 30-msec condition. The fact that the subject consistently identifies correctly the letters at either end in 50-msec but not 30-msec presentations demonstrates that attention may shift from one end of the display to the other and identify an end letter within only 20 msec. This fits the subject's comments that he usually first checked the last 4 letters, and if the bar was not there he attended to the first four. The search for the bar was clearly not from one letter to an adjacent one.

To go back to the details of the scan used in the full-report conditions, it should be interesting to compare letter positions correctly reported with those just seen when using partial report. Selected data follow: (i) Eight letters, 70 msec, first series (Fig. XXIV-11a). The results were very similar to the partial report

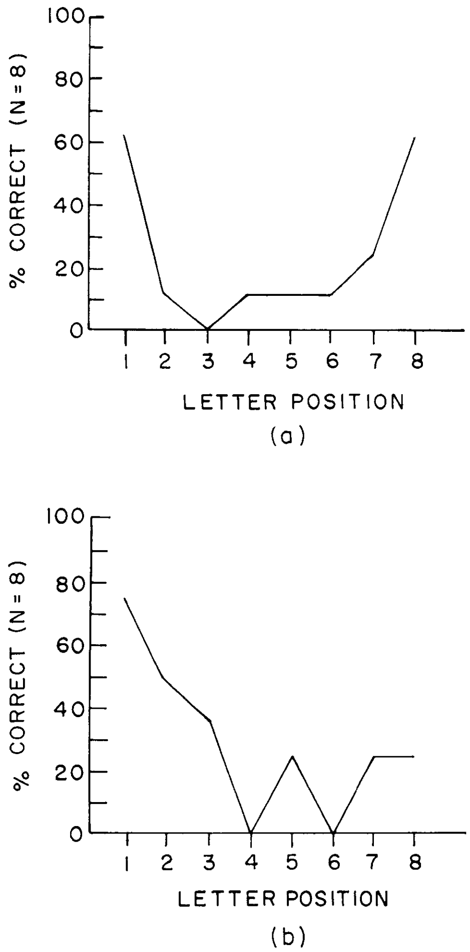

Fig. XXIV-11.

Full report. Accuracy in identifying the letters at each letter position. Eight letters presented; 70 msec duration. (a) First series. (b) Second series.

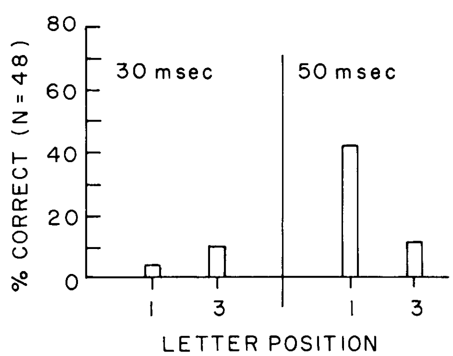

Fig. XXIV-12.

Full report. Accuracy in identifying letters in the first and third letter positions at two stimulus durations. 4M, 4D, and 8 letter presentations combined; both series combined.

results. (ii) The same condition, the second series (Fig. XXIV-11b). Now there is clear evidence of a left-to-right scan. Toward the end of the first series and throughout the second the dominant tendency for all (except the 30 msec) 8-letter conditions was to scan left to right. The third or fourth letter was rarely identified correctly unless the second or third letter was correct.

The letter-position preferences in the other conditions were as follows. There was little left or right predominance in the two-letter conditions, with full or partial report. In the 4-M conditions, the order of position preferences for both full and partial reports was 1, 4, 2, 3, while in the 4-D conditions it was 1, 2, 3, 4.

The proper interpretation of these results on letter position may be the following. Those letter positions best reported in the partial-report conditions probably were the most distinctive visually and thus the easiest to analyze. In the 4M and 8 partial-report conditions the letter on the right end was more often identified correctly than the letters in the middle, while the reverse was true in the 4D conditions. Thus massing of the letters seems to increase distinctiveness of the ends. The letter position preferences in the 4M and 4D full-report conditions show similar tendencies to the partial-report conditions. This indicates that when 4 letters are presented the order of the full-report scan was mainly from the most distinctive to the least distinctive positions. Greater option in the scan order was revealed in the 8-letter condition.

Figure XXIV-12 shows how the tendency to attend to the first letter in the full report conditions developed sometime between the 30- and 50-msec presentation times.

Conclusions

1. In attempting to identify random letters, in contrast to detecting them, the subject seems to attend to only one letter at a time. The rate at which letters are scanned depends on the difficulty of the task, with the no scan rate being the "natural" one.

2. Little or no active search for a bar over a letter or special identification of that letter seems to occur when 2 or 4 letters are displayed in a row, but active search and special identification do occur when 8 letters are displayed.

3. The attentional mechanism seems able to shift from one end of an 8-letter display to the other and identify the letter under a bar at the end within approximately 20 msec.

4. The scanning order of a 4-letter array is primarily from the most distinctive to the least distinctive letter position, while more options seem available when scanning an 8-letter array.

3. Afterword

The results described here concerning the scanning process used in letter identification should not be thought to be descriptive of normal reading. In the present tasks the subject knew that the letters were selected randomly, and thus it made good sense to attend to one letter at a time and not to allow any possible preliminary analysis of other letters to affect the identification of the attended-to letter. A preliminary experiment in which familiar words of 8 letters were randomly intermixed with stimuli of 8 random letters yielded very different results. The subject knew that both familiar words and random letters would be presented. The average number of letters correctly identified in the 30, 40, and 50-msec conditions was 0.4, 0.5, and 1.7 for the random letters and 0.4, 3.4, and 7.4 for the familiar words. There was a tendency to report familiar words and to show many letter substitutions in the random-letter 50-msec condition (ORDERED was fairly clearly seen when QKDQHBPG was presented). This shows that the processing strategy was very different from the one used when only random letters were expected. Results with the familiar words indicated that chunks of 3-4 letters at a time were analyzed in approximately 10-msec intervals. Probably, each letter was only partially analyzed, and the remainder was filled in by the subject's knowledge of the sequential letter probabilities in English.

The author would like to acknowledge the aid of Dr. Paul A. Kolers, both in the use of his tachistoscope and in helping me make this report as readable as it is.

P. H. Liss

References

1. G. Sperling, "The Information Available in Brief Visual Presentations," Psychol. Monogr., Vol. 74, No. 11 (Whole No. 498), 1960.

2. G. Sperling, "A Model for Visual Memory Tasks," IEEE Trans., Vol. HFE-5, pp. 19-31, 1963.

3. E. Averbach and G. Sperling, "Short-Term Storage of Information in Vision," in Information Theory, edited by C. Cherry (Butterworths Scientific Publications, London, 1961), pp. 196-211.

D. IMAGES AND FREQUENCY-DOMAIN NOISE

Images may be represented by samples that form a two-dimensional array. They may be equally well represented by samples of their Fourier transform, with the requirement that they be bandlimited.[1] The number of required transform samples equals the number of required image samples.

Independently adding white Gaussian noise to the real and imaginary parts of the frequency samples is equivalent to adding white Gaussian noise of the same variance to the image. This is true also for the addition of white Gaussian noise to the transform magnitude samples. The only requirement is that signal energy be the same in the frequency representation as in the image-domain representation.

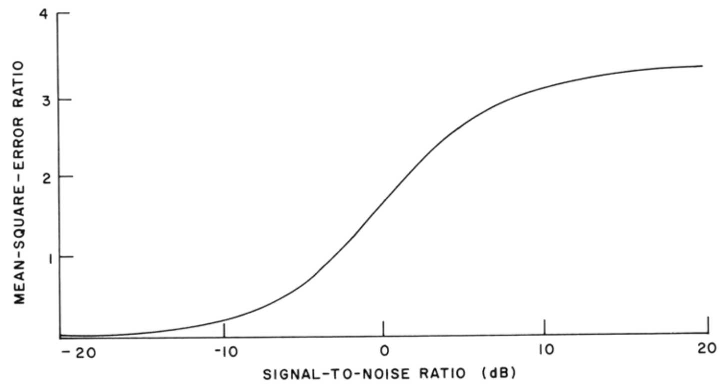

Fig. XXIV-13. Mean-square-error ratio as a function of signal-to-noise ratio.

(a)

(b)

Fig. XXIV-14. (a) Picture with image noise.
(b) Picture with phase noise.

The addition of white Gaussian noise to spatial frequency phase samples produces very dissimilar results, however. The ratio of mean-square error, caused by phase noise, to mean-square error, caused by image noise, is given by

$$\text{mean-square-error ratio} = \frac{2\pi^2}{3\sigma^2}\left(1 - e^{-\frac{\sigma^2}{2}}\right). \tag{1}$$

This expression is derived under the assumption of equal signal-to-noise ratios for both noise additions. Figure XXIV-13 is a graph of this function in the range from minus 20 db to plus 20 db.

Figure XXIV-14a is an image that has been subjected to white Gaussian noise addition in the image domain. Figure XXIV-14b results when white Gaussian noise is added to the transform phase samples. In both pictures a signal-to-noise ratio of 22 db was used for the noise addition.

G. B. Anderson

References

1. G. B. Anderson, "Images and the Fourier Transform," S. M. Thesis, Department of Electrical Engineering, M. I. T., May 1967.