Evidence that short-term memory is not the limiting factor in the tachistoscopic full-report procedure*

GEORGE WOLFORD and SAMUEL HOLLINGSWORTH
Dartmouth College, Hanover, New Hampshire 03755

An experiment was conducted to determine whether acoustic short-term memory is the factor which limits performance in the tachistoscopic full-report paradigm. Many Es have demonstrated the existence of phonemic encoding in short-term memory. The confusion errors from a tachistoscopic full-report task were, therefore, analyzed for the presence of acoustic confusions. Absolutely no evidence for acoustic confusions was found; visual confusions, however, were abundant. It was concluded that acoustic short-term memory is not the limiting factor in the full-report paradigm.

The measurement of the span of apprehension (or single look perceptual capacity) has occupied a place in the literature over a considerable time span. The basic finding was that when Ss were asked to report all of the letters from a tachistoscopic display, the number of reported letters increased with increases in display size up to about four letters. Increases in display size beyond that point produced no further increment in the number of letters reported. This was believed to represent the limit on the amount of information that human observers could perceive in a single look at a visual display.

Sperling (1960) questioned the nature of the performance limitation in this task. He felt that Ss could "see" many more than four items in a brief visual display but that their report was limited by a bottleneck further upstream. Sperling and others have postulated that this bottleneck is an overloading of short-term memory (STM) (Estes & Taylor, 1964; Sperling, 1960). Various investigators, in dealing with visual information processing, have postulated a number of different types of storage, often with a different set of names per investigator. Two of those postulated stores are highly relevant to the development of this paper and need to be carefully distinguished. Most theories start the processing with a sensory store which preserves a fairly iconic representation of the stimulus, has a large capacity, and is not under the direct control of S. The second store in the system has a fairly small capacity relative to the sensory store, is generally assumed to have an acoustic representation, and Ss can preserve information in it via rehearsal (Atkinson & Shiffrin, 1968). This second store will be referred to as STM throughout this paper.

At least two different techniques have been developed which measure processing limitations under conditions which place less of a burden on STM. Sperling (1960) and Averbach & Coriell (1961) developed techniques in which the S only had to report a limited amount of the information from a display as indicated by an appropriate marker subsequent to the display. In Sperling's partial report procedure this marker was a tone which signaled which row of a matrix was to be reported and in Averbach and Coriell, the marker was a visual stimulus which indicated spatially which letter to report. Estes and Taylor (1964) introduced the detection paradigm in which S had to determine which member of a predetermined set of targets a particular display contained. All of these procedures greatly reduced the burden on STM, and all of them yielded much higher estimates of perceptual capacity. The success of these procedures was taken as evidence for STM as the culprit in the full-report paradigm. These studies, however, do not offer direct evidence for STM overloading in the full-report paradigm. It is possible that these procedures altered the "perceptual task" in ways which led to the improved performance (Rumelhart, 1970). Others have suggested that different decisional processes might be involved in the different paradigms (Gardner, 1973).

There is some evidence that the asymptote of performance in the full report paradigm is reached prior to the point where the capacity of STM becomes a problem. Sperling and Speelman (1970) estimated the capacity of STM to be 7.51 letters for an acoustically dissimilar alphabet and 5.26 for an acoustically similar alphabet. They used both auditory and simultaneous visual presentation modes in estimating STM capacity. Most studies using the full-report procedure use the entire set of consonants to construct their displays (this might be considered an acoustically neutral alphabet). This alphabet should yield a value somewhere in between the two above values and one which is considerably higher than the asymptote of 4.2 letters obtained by Sperling (1960) in his earlier studies using the full-report procedure.

*The project reported herein was performed pursuant to a grant from the U.S. Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein, however, do not necessarily reflect the position or policy of the U.S. Office of Education, and no official endorsement by the U.S. Office of Education should be inferred. This article was written while the first author was visiting the Center for Human Information Processing at the University of California, San Diego. The center (under partial support from NIMH Grant MH-15828) was generous in its assistance.

The purpose of the present paper is to obtain more direct evidence for the role of STM in the full-report procedure. The logic of the experiment rests on the finding that Ss in STM experiments tend to make a substantial number of acoustic confusion errors (Conrad, 1964; Sperling, 1963; Wickelgren, 1965). Two properties of these confusion errors are also of relevance: The conditional probability of an acoustic confusion given an error is highest at minimal delays (Conrad, 1967); and, the presence of acoustic confusions is not restricted to a particular input or output mode (Sperling & Speelman, 1970). In this experiment, then, Ss were presented with tachistoscopic arrays of letters; they were asked to report as many of the letters as possible; and their errors were examined for the presence of acoustic confusions. There is a problem, however, with the standard full-report procedure for present purposes. When an error is made in the standard full-report procedure there is no way of determining what the correct letter should have been, because position information is not retained in the response. In this experiment, therefore, Ss had to report all of the letters presented and had to report them in the correct positions if possible.

METHOD

Subjects and Apparatus

Five persons from the Dartmouth community were paid \$2 per hour for their participation in the experiment. All had normal or corrected normal vision and were naive with respect to the experiment.

The stimulus materials were presented in a Scientific Prototype three-channel tachistoscope (Model GA) that was modified with a rapid card changer on two of the channels. Character strings were presented along the horizontal median of a lighted rectangular field which subtended a visual angle of 7.82 deg in width and 1.68 deg in height. One field (the fixation field) contained a circular black fixation point measuring 0.073 deg in diam and centered with respect to the rectangle described above. The luminance of all of the fields was adjusted to 10 fL as measured with a Macbeth illuminometer. The fixation field was visible at all times except during the presentation of the stimulus field. The total illumination in the laboratory was provided by two 7-W bulbs shielded from S. The responses of the Ss were entered by E onto a Portacom portable teletypewriter that was linked to the Dartmouth time-sharing system.

Stimulus Materials

The 160 five-letter character strings were generated at random without replacement from the 20 consonants (excluding Y). The character strings were typed in royal bulletin typeface on white 5 x 8 in notecards. The characters were 0.29 deg high and 0.14 deg wide. The intercharacter space was 0.073 deg. The entire five letter array subtended a visual angle of 1.07 deg. The arrays were positioned on the cards so that the leftmost letters would begin 0.073 deg to the right of the fixation point and would be located on the same horizontal median.

Procedure

Each S received two presentations of the 160 stimuli over a 2-day period. A new random order was used for each S and for each session. A trial sequence began with the fixation field in

view and E saying ready. S then initiated a stimulus exposure by depressing a handheld microswitch. The stimulus field appeared with no delay for a predetermined exposure followed by a return to the fixation field. Following each stimulus exposure, S was instructed to report orally all five consonants in a left-right order guessing if necessary. He was told that only those responses that were output in the correct position would be scored as correct. The Ss were aware of the set from which the letters were drawn and of the absence of repeated letters. The intertrial interval was approximately 11 sec. Feedback was given to the Ss after every 20 trials as to the percent correct overall and the percent correct on the leftmost letters. The stress on the leftmost letters was intended to insure left-right processing. The exposure durations were altered on a continual basis for each S to maintain performance at approximately 60% overall. Final exposure durations ranged from 15-25 msec. An experimental session lasted approximately 60 min.

RESULTS

The overall performance is depicted in Table 1. Each point is based on 1600 observations (5Ss by 320 trials). The overall data were scored in two different ways. In the position relevant method, an item had to be output in the correct serial position to be counted as correct. In the position irrelevant method, a letter in the response was scored as correct if it appeared anywhere in the string. Unless explicitly mentioned otherwise, only the position relevant method will be used in future analyses. The probability correct averaged across serial positions was .63 for the position relevant method and .76 for the position irrelevant method. The serial position curves were U-shaped as in many tachistoscopic experiments.

The types of errors which were made in the experiment are presented as a confusion matrix in Table 2. That confusion matrix was scaled using a technique developed by McGee (1966) referred to as nonmetric elastic multidimensional scaling and the matrix was clustered using Johnson's nonmetric hierarchical procedure. The results of those analyses are summarized in Fig. 1. Visual confusions appear to predominate. The primary axis separates the letters into ones with curved components and ones without. Two problems become readily apparent in trying to locate acoustic confusions from the multivariate analyses presented: Firstly, the fact that visual errors might predominate is not surprising given the nature of the task. Our purpose, however, was to detect the presence not the predominance of acoustic confusions. Secondly, acoustic and visual confusions are not necessarily mutually exclusive. Several letter pairs which are close together visually are also close acoustically (i.e., MN and GC).

To circumvent those two problems, we constructed two mutually exclusive alphabets: one visual, and one acoustic. To construct the alphabets we used data collected by Wessel¹ who had 47 Ss sort the letters of the alphabet twice: once on the basis of visual similarity and once on the basis of acoustic similarity. The data

Table 1
Mean Percentage of Correct Responses

Scoring Method	Serial Position				
	1	2	3	4	5
Position Relevant	92	80	56	34	54
Position Irrelevant	94	90	76	54	67

from those sorting tasks were clustered using Johnson's nonmetric procedure. Any letter which was clustered with another letter above a diameter of 12 was placed on the same row of the alphabet as the other letter. The two alphabets are presented in Table 3. A confusion was defined as one which came from the same row of the alphabet as the correct letter.

The alphabets were made mutually exclusive by deleting any pair as a confusion which appeared in the same row in both alphabets. Therefore, outputting M when N was correct would not be scored as a confusion error, if however, W was output for N it would be scored as a visual confusion. Using this procedure, the errors can be partitioned into three mutually exclusive, exhaustive categories: visual confusions, acoustic confusions, and other. It is also possible to derive the probabilities that a purely random response will fall into one of the three categories given that an error was made. These chance conditional probabilities are 0.16 for visual confusions, 0.17 for acoustic confusions, and 0.67 for other. The latter value should not be confused with the value of 0.25 for the ratio of acoustic to other confusions shown in Table 4. Table 4 shows the observed conditional probabilities for the various error types as a function of serial position. The data for "other" are not presented as they can be derived from the presented data.

A *t* test comparing the mean probability of a visual confusion with chance yielded a $t(4df) = 3.59$, $p < .05$. The acoustic confusions were not significantly different from chance ($t(4df) = -1.07$). The standard error of the

Fig. 1. A summary of the two space nonmetric multidimensional scaling and clustering solutions of the experimental errors.

visual means in Table 4 was 0.024 and for the acoustic means was 0.015. An analysis of variance revealed that neither of the serial position curves for the two confusion types was significant. When the full visual confusion alphabet was used (i.e., without deleting the overlapping pairs), the U-shape of the serial position curve became more pronounced and was significant ($F(4,16) = 6.49$, $p < .01$).

The tests presented above were based on the null hypothesis that Ss were making random error responses. Since visual confusions occurred much more frequently than chance would predict, the estimate of chance for acoustic confusions might have been inaccurate. To circumvent this problem the ratio of acoustic to "other" errors was compared with the predicted chance ratio. This chance ratio would remain invariant as a function

Table 2
Confusion Matrix for Experimental Data

	B	C	D	F	G	H	J	K	L	M	N	P	Q	R	S	T	V	W	X	Z
B	**	4	17	4	5	4	3	0	5	2	6	10	5	8	14	2	3	4	3	1
C	3	**	10	9	22	2	2	1	3	2	4	5	6	4	13	2	1	3	3	4
D	7	5	**	5	3	6	3	1	6	4	10	3	19	4	8	1	6	2	4	3
F	3	5	7	**	1	6	2	3	11	2	5	8	3	8	5	20	2	5	3	3
G	7	12	2	2	**	3	2	1	3	2	3	4	13	3	15	10	3	6	3	5
H	5	2	4	4	3	**	3	3	10	11	25	6	2	4	4	4	5	4	1	3
J	5	8	5	3	3	10	**	1	20	2	8	2	5	3	2	6	6	8	1	1
K	4	3	2	4	2	4	3	**	7	3	6	5	6	10	5	3	5	12	15	4
L	4	8	5	6	1	10	9	3	**	3	10	2	1	3	11	9	4	5	4	1
M	0	2	3	2	2	19	3	2	5	**	25	4	2	3	3	4	2	16	1	4
N	5	2	4	1	1	18	3	2	3	16	**	2	3	6	4	5	3	16	2	2
P	5	3	7	15	0	9	9	2	3	1	8	**	4	10	2	11	2	3	5	2
Q	2	9	20	2	9	1	1	1	3	2	5	4	**	11	9	6	2	4	5	4
R	5	4	8	7	1	10	3	1	2	2	8	11	8	**	8	7	4	4	4	4
S	12	9	5	5	7	3	4	1	6	1	7	4	7	4	**	7	8	3	5	3
T	4	1	3	7	2	9	3	4	15	1	7	5	2	3	9	**	10	8	2	4
V	12	3	4	6	0	2	1	2	8	3	4	4	4	4	5	10	**	16	9	1
W	3	1	5	4	2	5	2	5	4	18	13	3	4	3	5	8	8	**	5	3
X	3	1	5	4	0	5	3	5	10	1	3	4	2	4	13	10	6	4	**	16
Z	4	7	4	9	5	5	1	0	6	5	6	5	1	10	4	14	4	3	4	**

Table 3
The Two Mutually Exclusive Confusion Alphabets

Visual	Acoustic
SJ	HJK
MN WKX VZ	QW
FHLT	MN L
CG Q	FSX
DBP R	R
	BPD TVZ CG

of the percentage of visual confusions. These ratios are presented in Table 4. The predicted chance ratio was 0.254. The standard error for the mean ratios was 0.013 and the overall mean ratio was not significantly different from chance ($t(4df) = -0.99$). From the various statistical tests and the data presented in Table 4 it is evident that while visual confusions are prominent there is absolutely no evidence for the existence of acoustic confusion errors.

The final analysis performed on the data was to break the errors into two types: transposition errors (a letter which was misplaced in the response string) and nontransposition errors or intrusions. These data were then subjected to the confusion analysis described above. The reason for doing this analysis is based on the finding of Estes (1972) that the conditional probability of an acoustic confusion given an error was higher for transposition errors than for intrusions in a STM experiment. The analysis was not broken down by serial position. The probability of a visual confusion was 0.24 for intrusion errors and 0.20 for transposition errors. The former value was significantly above the chance value of 0.16 ($t(4df) = 4.04$, $p < .01$) but the latter value was not significant. The probability of an acoustic confusion was 0.13 for intrusion errors and 0.15 for transposition errors. Both values are below the chance value of 0.17 for acoustic confusions. Intrusion errors outnumbered transposition errors about 2 to 1 in the data. The ratio of acoustic errors to "other" errors was 0.19 for intrusions and 0.23 for transposition errors; both below the chance value of 0.254.

DISCUSSION

The results described above seem quite conclusive in demonstrating an absence of acoustic confusions in our tachistoscopic experiment using the full-report procedure. Our conclusion from these data is that the

Table 4
The Conditional Probability of Visual Confusions, Acoustic Confusions, and the Ratio of Acoustic to Other Confusions

Confusion Type	Serial Position				
	1	2	3	4	5
Visual*	.27	.22	.22	.23	.25
Acoustic**	.16	.15	.14	.16	.13
Acoustic/Other†	.28	.24	.22	.26	.21

*Chance = 0.16 **Chance = 0.17 †Chance = 0.25

asymptote of performance in the full-report procedure is reached *prior* to the potential influence of an overload in STM. At least two issues need to be discussed with reference to this conclusion: What are the possible flaws in our design and logic; and, if STM is not the culprit in the full-report procedure, what is?

There are several possible flaws in our conclusion. The first problem is that we may have distorted the ordinary full-report paradigm in such a way as to produce results which were different from those obtained with the ordinary procedures. One way to evaluate this problem is to compare the summary statistics of our data with those of Sperling (1959). The asymptote in Sperling's experiment was reached at a display size of about five and equaled 4.2 letters correct. In our experiment the average number of letters correct per trial was 3.3. A large part of the gap between the two figures was produced by the difference in scoring procedures. If our data are scored in a similar fashion (i.e., the position irrelevant method of Table 1), the average number of letters reported correctly jumps to 3.9 per trial, much more in line with Sperling's data. In fact, with a display size of 5, Sperling's Ss output an average of 4.0 letters. The similarity between the summary statistics of the two experiments plus the failure to find acoustic confusions when the errors were broken down leave us with no compelling reason to believe that our confusion errors would differ substantially from other studies using the full-report procedure.

A minor potential problem was that in constructing the two confusion alphabets to be mutually exclusive, we may have left the acoustic alphabet with weaker clusters. One way to test this is to compare the diameters of the remaining clusters in the two alphabets. With Johnson's clustering procedure (maximum method) large diameters represent more prominent clusters. Using the clustering results from the sorting tasks, the mean diameter for the remaining visual clusters was 0.45 and the mean diameter for the remaining acoustic clusters was 0.60. It would appear, then, that the acoustic alphabet did not suffer from weaker clusters.

Another major problem is the possibility of nonacoustic forms of STM. As mentioned in the introduction there are several postulated stores in the visual system which are thought to be nonacoustic, but we are referring specifically to the store with limited capacity and in which rehearsal can be used to maintain information. Several investigators have produced evidence of acoustic coding in STM that was not dependent on input and output modes. Estes (1973), however, demonstrated that it is possible to design a STM experiment in which no acoustic confusions are made. The primary ingredient seems to be the suppression of the opportunity to rehearse or vocalize (even subvocally) the items. In our experiment, the Ss were not prevented from rehearsing in the interval from the offset of the display until they initiated their report. While no attempt was made to measure latencies, Ss often spent considerable time pondering their answers

prior to responding. Our procedure, then, would be somewhat similar to the "early pause, rapid display" condition in Experiment 2 of Estes (1973). A condition which did produce significant acoustic confusions.

Why are Ss only able to report four items in the full-report procedure when the other procedures outlined in the introduction yielded much higher estimates of perceptual capacity? It seems quite possible that the full-report paradigm differs from detection paradigms in the number of features which need to be extracted to insure correct performance (Rumelhart, 1970). It is clearly possible to design a detection task in which the detection of a single feature is sufficient to distinguish the targets from one another and from the noise elements (e.g., V and H as targets and O's as noise elements). In the full-report procedure, however, as many as seven features per letter are required to insure correct letter identification (Norman & Lindsay, 1972). It may be that the asymptote in the full-report paradigm is produced by a limit on the number of features which can be extracted per look. A detection task with this same feature limit could produce a much higher estimate of perceptual capacity with the estimation procedures used in those experiments (Rumelhart, 1970). In addition Ss may use different decisional strategies in the various paradigms which may put the full-report paradigm at a disadvantage (Gardner, 1973).

REFERENCES

- Atkinson, R. C. & Shiffrin, R. M. Human memory: A proposed system and its control processes. In Spence, K. W. & Spence, J. T. (Eds.), *The psychology of learning and motivation*. New York: Academic Press, 1968.
- Averbach, L., & Coriell, A. S. Short-term memory in vision. *Bell System Technical Journal*, 1961, 40, 309-328.
- Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, 55, 75-84.
- Conrad, R. Interference or decay over short retention intervals? *Journal of Verbal Learning & Verbal Behavior*, 1967, 6, 49-54.
- Estes, W. K., & Taylor, H. A. A detection method and probabilistic model for assessing information processing from brief visual displays. *Proceedings of the National Academy of Sciences*, 1964, 52, 446-454.
- Estes, W. K. An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.) *Coding processes in human memory*. Washington, D.C.: Winston, 1972.
- Estes, W. K. Phonemic coding and rehearsal in short-term memory for letter strings. *Journal of Verbal Learning & Verbal Behavior*, 1973, 12, 360-372.
- Gardner, G. T. Evidence for independent parallel channels in tachistoscopic perception. *Cognitive Psychology*, 1973, 4, 130-155.
- McGee, V. E. The multidimensional analysis of "elastic" distances. *British Journal of Mathematical & Statistical Psychology*, 1966, 19, 181-196.
- Norman, D. A. & Lindsay, P. H. *Human information processing*. New York: Academic Press, 1972, p. 120.
- Rumelhart, D. E. A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, 1970, 7, 191-218.
- Sperling, G. Information in a brief visual presentation. Unpublished doctoral dissertation. Harvard University, 1959.
- Sperling, G. The information available in brief visual presentations. *Psychological Monographs*, 1960, 74 (11, Whole No. 498).
- Sperling, G. A. model for visual memory tasks. *Human Factors*, 1963, 5, 19-31.
- Sperling, G. & Speelman, R. G. Acoustic similarity and auditory short-term memory: Experiments and a model. In Norman, D. A. (Ed.) *Models of human memory*. New York: Academic Press, Inc., 1970.
- Wickelgren, W. A. Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology*, 1965, 70, 102-108.

NOTES

1. The sorting task data were provided by David Wessel at Michigan State University, East Lansing, Michigan. We are grateful for his assistance.

(Received for publication February 22, 1974;
revision received May 6, 1974.)