Letter recognition: Effects of interitem similarity and report requirements*

LEON K. MILLER

University of Illinois, Chicago Circle, Chicago, Illinois 60680

The effects of auditory similarity, visual similarity, and position of letters within the visual field upon letter recognition were investigated in two experiments. In Experiment 1, significant differences associated with visual, but not auditory, similarity were found when Os were asked to report one, three, or all items from five-item linear arrays. As report requirements increased, left-right differences in report accuracy increased, although items in the center of the field were reported most accurately. In Experiment 2, either one or five items were requested from the arrays presented in Experiment 1 after a delay of 2 or 4 sec. Significant effects associated with both auditory and visual similarity were found, together with a tendency for items to the left of fixation to be reported most accurately. These results were discussed in terms of recent theories dealing with the encoding of visually presented material.

When an O is asked to report the contents of briefly presented sets of discrete items, several factors typically are found to affect report performance. Items to the left of the fixation point are reported with greater accuracy than those on the right, and identification of items at either end of the array is more accurate than that of items in the center (Heron, 1957; Harcum, 1967). Second, when the set consists of alphanumeric characters, the greater the acoustic similarity of the various items of the set, the poorer the performance (Conrad, 1964). Visual similarity, on the other hand, appears to have little, if any, effect upon performance (Cimbalo & Laughery, 1967; Laughery & Harris, 1970). Effects associated with item position are usually interpreted in terms of O's preferences for encoding and rehearsing the content of linear arrays; items on the left are scanned first, are more likely to be encoded, and, when encoded, are rehearsed longer. Items at either end of the array are not as subject to interitem interference as those in the middle (Harcum, 1967). Interpretations of the effects of acoustic vs visual similarity likewise implicate storage and rehearsal by suggesting that the encoding of visual material involves a translation of the visual stimulus into its auditory equivalent (Sperling, 1963; Conrad, 1964). Subsequent rehearsal is then in terms of the auditory equivalents of the items seen. Since storage and rehearsal are auditory in nature, similarity along these dimensions, rather than along a visual one, is more important.

*This research was supported by a grant from the Graduate College, University of Illinois, Chicago Circle. The assistance of Robert Horwitz and Tomi Henek in conducting the research is gratefully acknowledged.

In the investigations mentioned above, the task required of the O has usually necessitated the temporary storage of successively presented material and/or the report of all items shown. When Os are asked to report only a specified item from a tachistoscopically presented array, hemifield preferences become less pronounced and items in the center of the array are identified with relatively high levels of accuracy (Averbach & Coriell, 1961; Crovitz et al, 1966; Haber & Standing, 1969). Furthermore, Keele and Chase (1967) found letter confusions in a partial report task to be more highly correlated with visual similarity than with acoustic similarity. Hershenson (1969) has suggested that when Os are not required to report the identity of all elements in a display, performance may be more directly related to what was registered within the visual system. An increase in report requirements, on the other hand, results in an increasing memory requirement, and factors associated with storage rather than reception of material assume greater importance for performance.

The present investigation was

designed to assess the hypothesis that the effects of item position and auditory vs visual similarity depend upon the report requirements of the task. Os were shown arrays of items and required to identify only part of the array or the complete array. It was expected that under minimal report requirements, visual rather than acoustic similarity would be a more important determinant of report performance and items in the center of the array would be identified most accurately. An increase in report requirements was expected to lead to an increase in the importance of auditory similarity for performance and an increase in accuracy for letters to the left of the fixation point relative to those in other positions.

EXPERIMENT 1

Apparatus and Materials

Stimuli were presented in a three-channel tachistoscope, the first channel containing the fixation marker, the second containing the letters to be reported, and the third containing the indicators for the various report conditions. The illumination of each field was approximately 25 fL.

The stimuli consisted of four sets of letters representing orthogonal comparisons of high and low visual and auditory similarity. In an attempt to obtain some empirical basis for assigning letters to similarity conditions, similarity ratings for all letter combinations were found in a pilot study. These ratings were obtained by asking 14 Os to rate each of the 325 pairs generated by the (capital) alphabet on a 5-point scale representing similarity either in terms of visual or, in a second group of Os, auditory characteristics.¹ These scores were then used as the basis for assigning letters to the high and low visual and auditory similarity groups. Letters used for each of the four conditions, together with the mean similarity scores, are given in Table 1. Os tended to give a greater range of

Table 1
Composition and Similarity Ratings of the Letter Sets

Condition	Letters	Rating			
		Auditory		Visual	
		Mean*	SD	Mean*	SD
High Auditory High Visual	G C B P D	2.16	.31	2.82	.65
High Auditory Low Visual	P C Z T E	2.19	.38	3.52	.47
Low Auditory High Visual	F I L J T	4.11	.23	2.65	.58
Low Auditory Low Visual	F G A R X	3.92	.43	4.07	.47

*Means based upon the similarity ratings for the 10 pairs of letters in the set.

Fig. 1. Recognition as a function of item position and report condition: Experiment 1.

scores for auditory than for visual similarity, and it was impossible to construct sets whose items were judged equally distant along auditory and visual similarity dimensions. The letters contained in the high-auditory-, high-visual-similarity set and the bigrams generated by the two high-auditory-similarity sets occur less frequently in language samples, according to the Underwood and Shultz (1960) norms. Consequently, the letter sets constructed favored the emergence of differences associated with auditory rather than visual similarity.

Twenty stimulus cards, each containing five items, were constructed for each of the four sets of letters. The letters (Paratype No. 11207) each subtended 20 min of visual angle when seen through the viewing hood and appeared in a linear array, with items at a point slightly above the fixation marker and at 1 and 2 deg to either side. Letters were selected randomly from the appropriate set for each card, with the constraint that no letter appeared more than twice on any single card.

Ten indicator cards were made for each of the three report conditions (see below). Each indicator card consisted of horizontal bars corresponding to one, three, or five of the letter positions represented on the stimulus cards. When viewed in the tachistoscope, these markers appeared at positions .5 deg directly below the letters. Across each set of 10 indicator cards, each position and, for the partial-report groups, combination of positions was represented equally often.

Subjects

Forty-eight students participated in the experiment as part of an introductory psychology course requirement. Os were assigned

randomly to one of three conditions, the conditions being defined by the report requirements of the task. Os in Partial Report 1 (R1) were required to report only one of the five items appearing on each trial. Os in Partial Report 3 (R3) were required to report three items on each trial, and Os in a full-report group (R5) were required to report all items in the array.

Procedure

Each O was given 20 trials, 10 at each of two exposure durations, 50 and 100 msec, under each of the four sets of letters. The order in which the two exposure durations were administered was randomized, and the order of administering letter sets was counterbalanced across Os. Each O was seated in front of the tachistoscope and asked to fixate the dot appearing in the center of Field 1. He was then told that an array of letters would appear very briefly and would be followed by indicator bar(s), his task being to write down the letters which had appeared above each bar. Os were given response sheets containing five spaces corresponding to the five positions in the array and responded by writing in the appropriate letter at the positions requested. After 10 practice trials with a neutral set of letters, the 80 trials were administered. Before each new set of letters, Os were shown the letter set from which the stimuli for the next 20 trials would be drawn. This letter set remained available for referral throughout the trial block.

Results

Each letter correctly identified in the appropriate space on the answer sheets was designated a correct response. This scoring procedure resulted in a different number of total possible correct for each report condition; consequently, separate analyses of variance for each group were conducted.

These analyses revealed significant effects associated with visual similarity for each report condition [$F(1,15) = 10.79, 11.07, \text{ and } 22.0$, $p < .01$ for Groups R1, R3, and R5, respectively], high visual similarity in each case being associated with lower levels of performance. Auditory similarity, on the other hand, had no reliable effect upon performance, nor did it interact with the other conditions.

Errors were also examined to determine whether they were related to the auditory or visual similarity between the letters actually appearing and O's response. Since there were marked differences in performance between the high and low visual confusability sets, the 20 pairs of

possible confusions in each set were considered separately. Correlations between the number of confusion errors for each pair across Os and the visual-similarity rating were negative for all groups and letter sets, indicating that confusions between letters judged to be less visually similar occurred less frequently. These correlations were significant only for the 20 highly visually confusable pairs in Groups R1 and R5 ($r = -.46, -.43$, $df = 18$, $p < .05$), however. Similar analyses using the auditory confusability index revealed no consistent trend. Bigram frequency (Underwood & Shultz, 1960) also proved to be unrelated to the distribution of errors found.

Performance differences associated with item position were significant for each report condition [$F(4,60) = 22.07, 13.25, \text{ and } 22.40$, $p < .01$ for Groups R1, R3, and R5, respectively]. As is seen in Fig. 1, Os in all groups reported items appearing in the center of the array most accurately. Trend analyses (Winer, 1962) indicated that for each group the quartic component of variance associated with item position was greatest, reflecting highest levels of performance for items at the center and at either end of the display. Of the remaining components, quadratic trends were more pronounced for R1, while linear trends were greater for the R3 and R5 groups.

Further examination of responses in Group R5 revealed consistent individual tendencies to report items either in the center array or to the left most accurately. For nine Os in this group, highest levels of performance were found for the letter in the center of the display for all four sets of displays. For the remaining seven Os, accuracy was consistently highest for the items on the far left. No such tendencies were found for Groups R1 and R3. Separate analyses of variance conducted for the Os showing "center" and "left" preferences in Group R5 revealed that for each, differences in visual, but not auditory, similarity were associated with reliable differences in performance.

Levels of recognition performance attained for the various letter similarity and exposure conditions are presented in Table 2. Accuracy was significantly lower at the shorter exposure duration in all groups. Several interactions involving exposure duration and the other variables were also found. For Os in R1, the effects of visual similarity were greatest at the shortest duration. Examination of differences contributing to a significant interaction among exposure duration, visual similarity, and item position suggested that the tendency for errors to be distributed

Table 2
Mean Recognition Performance as a Function of Letter Similarity,
Exposure Duration, and Report Requirements

	Group R1		Group R3		Group R5	
	50 Msec	100 Msec	50 Msec	100 Msec	50 Msec	100 Msec
High Visual Similarity	.41	.55	.33	.38	.36	.41
Low Visual Similarity	.53	.59	.40	.45	.41	.48
High Auditory Similarity	.48	.57	.36	.40	.40	.45
Low Auditory Similarity	.45	.57	.36	.43	.37	.45

symmetrically around the central position was greatest for the visually confusable items shown at 50 msec. Finally, a significant interaction among auditory similarity, visual similarity, and exposure duration was found for the R5 group. Effects of exposure duration here were greatest for the letter set low in both auditory and visual similarity.

EXPERIMENT 2

Unexpectedly, visual similarity remained an important factor across variations in report requirements in Experiment 1, while differences attributable to auditory similarity never appeared. Similarly, items in the center of the array elicited the highest levels of performance in all three report conditions, although there was a tendency for differences in performance for the left vs the right half of the array to increase as report requirements increased. Thus, there was little support for the prediction that the effects of interitem similarity and item position would depend upon report requirements.

There is some question, however, if the report manipulation used here necessarily increased the storage or memory requirements for the O. Previous investigations reporting acoustic confusability effects for visually presented material have presented lists of items serially, necessitating some storage of material until report is possible. In the present experiment, all items were presented simultaneously, Os were allowed to respond as soon as the display terminated, and short-term memory in any of the conditions may have been only minimally involved. Keele and Chase (1967) found that even when delays were introduced between the offset of arrays of visually presented items and the request for report of array contents, letter confusions were not correlated with an index of auditory similarity. Items were displayed in a circular array around the fixation point, however, and Os may have found it difficult to organize material in terms of conventional preferences. The second experiment was designed to examine the effects of report delay under conditions presumably more favorable to encoding preferences. It was expected

that the introduction of these delays would require Os to store and rehearse array contents, thereby increasing the importance of acoustic similarity and strengthening the tendency for items to the left of the fixation point to be reported with greater accuracy than those to the right.

Subjects

Thirty students participated in the experiment as part of a course requirement in introductory psychology.

Procedure

Stimuli were shown in a three-field tachistoscope modified to permit the addition of a variable dark interval between the offset of the stimulus field and the onset of the field containing the bar markers. Os were assigned randomly to a partial-report (R1) or a full-report (R5) condition. Within each condition, Os were given a total of 80 trials, representing 20 trials for each of the four sets of letters. For one block of 10 trials within each letter set, a 2-sec dark interval appeared after the offset of the stimulus field, while in the remaining trials, the dark interval was 4 sec. Illumination of the fixation, stimulus, and marker fields was the same as that in Experiment 1, and exposure duration for the field containing the letters was 100 msec.

Results

Responses were scored as in the previous study, and analyses of variance were again conducted separately upon scores in the partial- and full-report groups. As is seen in Table 3, both auditory and visual similarity appeared to play a role in letter identification. These differences only approached conventional levels of statistical significance in the

partial-report group ($p < .10$) but were highly significant for the full-report group [visual confusability: $F(1,14) = 32.31$, $p < .01$; auditory confusability, $F(1,14) = 9.03$, $p < .01$]. When incorrect, Os tended to report letters that were both visually and acoustically similar to those actually appearing. Significant negative correlations involving auditory similarity and confusion frequency were found for both letter sets in the full-report group and for the 20 letter confusions possible in the sets low in visual similarity in the partial-report group ($r = -.41$, $-.40$, and $-.38$, $df = 18$, $p < .05$). Significant correlations with visual similarity were found for the sets rated high in visual similarity in both the partial- and full-report groups ($r = -.53$, $-.49$, $df = 18$, $p < .025$).

Figure 2 portrays the distribution of correct responses as a function of item position across the delay conditions examined in Experiment 2. Requiring a delay before response generally resulted in errors associated with position to be distributed in a fashion frequently found in tachistoscopic recognition. For the partial-report group, significant linear [$F(1,14) = 16.29$, $p < .01$] and quartic [$F(1,14) = 28.97$, $p < .01$] trend components were found, reflecting better recognition to the left and in the center of the array. As in Experiment 1, differences associated with item position seemed best described by a quartic function. For the full-report group, on the other hand, linear trend components accounted for most of the variance due to item position [$F(1,14) = 110.04$, $p < .01$], although significant cubic [$F(1,14) = 8.10$, $p < .01$] and quartic [$F(1,14) = 32.86$, $p < .01$] trends were also found.

While delays were introduced in order to increase memory requirements and presumably the difficulty of the task, overall performance in these conditions was superior to the groups examined in Experiment 1. Within the delay groups, some tendency was found for recognition to be less accurate for the 4-sec than the 2-sec delay ($p < .10$). Finally, an examination of Table 3

Table 3
Mean Recognition Performance as a Function of Letter Similarity
Delay and Report Requirements

	Delay 2		Delay 4	
	One Item	Five Items	One Item	Five Items
High Visual Similarity	.63	.51	.61	.51
Low Visual Similarity	.69	.62	.65	.56
High Auditory Similarity	.64	.55	.60	.51
Low Auditory Similarity	.68	.59	.66	.56

suggests that differences associated with visual similarity decreased as delay increased. However, none of the interactions involving delay and letter similarity reached significance.

DISCUSSION

Under minimal report requirements, the relation between accuracy and letter position was best described by an inverse U function. These results generally suggest that previous failures to find retinal sensitivity an important factor in tachistoscopic recognition (e.g., Crovitz et al, 1966) were due to the type of report requested. Partial report such as that used here and by Haber and Standing (1969) generally results in error distributions expected on the basis of retinal sensitivity. Full or delayed report of array contents, on the other hand, leads to the emergence of hemifield differences more easily explained in terms of preferences for encoding material.

Figural similarity exerted a greater influence upon recognition performance than did phonemic similarity, even when full report of array contents was required, but when an additional delay before report was required, auditory similarity did emerge as a factor in performance. In view of past studies dealing with the effects of letter similarity upon tachistoscopic perception, the consistently greater importance of figural than auditory similarity was somewhat surprising. This difference was found, even though letters were rated as more dissimilar in terms of their auditory than their visual characteristics.

The present study differed from most previous research dealing with visual vs auditory factors in tachistoscopic recognition in several respects. First, letter similarity was determined by independent ratings rather than on an a priori analysis of physical characteristics. Correlations between rated auditory and visual similarity and shared phonemic features for the auditory ratings and shared grapheme features (Gibson, 1969) for the visual ratings were significant but only moderate in strength ($r = .30$). Discrepancies between the present and previous research could reflect differences in the criteria used for determining item similarity. Second, items were presented in a linear array rather than sequentially at a single locus in the visual field. In this respect the present design is similar to that of Bryden (1968), who found report of groups of symmetrical letters to be less accurate than groups of asymmetrical letters. In terms of the ratings used here, the symmetrical letters used were more visually similar than the asymmetrical letters.

Fig. 2. Recognition as a function of item position and report condition: Experiment 2.

Most current conceptions of the processes involved in recognition of visual material postulate a fast decaying visual trace (Sperling, 1963) or icon (Neisser, 1967) which maintains the spatial and figural integrity of the array briefly after stimulus offset. As noted previously, the prevailing view has been that for iconic information to be retained and reported, it must first be translated into some form using auditory features. One interpretation of the present results is that the perceptual processing occurring used primarily figural rather than auditory features. Recent evidence has suggested that such "visual" processing can occur under certain situations. Posner et al (1969), in a series of studies, has compared reaction times for matching judgments of successively presented letters where the basis of identity could be phonemic or figural. Faster phonemic identity judgments were found when the basis of similarity was figural and phonemic than when it was phonemic only. Moreover, phonemic identity judgments were much faster when phonemic and figural identity were perfectly confounded than when they were not. This suggested that when Os find it more efficient to do so, they will encode characters in terms of their visual rather than auditory characteristics. Such encoding would preserve the figural, and perhaps the spatial, features of the display beyond the decay of the visual trace.

The structure of the present task may have optimized the chances for figural rather than phonemic encoding. The letter population used on each trial was quite small, and these populations were presented in a manner which probably emphasized interitem similarity. For example, Os were shown a card containing the letters which would be seen on subsequent trials, and the lists were

constructed so that they contained highly confusable or highly dissimilar items. Since the task required accurate report of an item at a given location, spatial characteristics of the items had to be retained. Indeed, since the letter population was so small, the task may be viewed as essentially letter localization, Os having to remember the location of each letter in the set. Both Klemmer (1964) and White (1969), in tasks requiring the O to identify an item at a particular location in an array, have reported relations between accuracy and item position quite similar to those found in Experiment 1. Finally, the response required replicating the items from the array rather than, for example, naming the letters seen.

An attempt was made to encourage the use of auditory features in reporting material in an additional group of 12 Os. The task was the same as that in the R5 group in Experiment 1, with the additional instructions that Os report the items orally from left to right and concentrate upon the sounds of the letters when performing the task. While this procedure resulted in a linear increase in error rate from the far left to the far right position, visual, but not auditory, similarity remained significantly related to accuracy.

However, the occurrence of visually related errors is not direct evidence that material was stored primarily in terms of visual characteristics or that report was directly from some persisting visual trace. Rather, it means that at some point in the processing of information, the visual characteristics of items were preventing successful encoding. The difference obtained could have reflected, for example, difficulty in a preliminary stage of visual encoding. Subsequent processing in terms of auditory characteristics could still have occurred, and such encoding would have preserved visual confusions made earlier.

This suggests that task parameters were arranged so that preliminary registration of materials within the visual system was far from perfect. For example, Glanzner (1966) and Mackworth (1963) have shown that at short exposure durations, the limiting factor in performance may be the clarity with which items are registered within the visual system, while at longer exposure durations, such registration is assured. Overall, viewing time was kept below that required for foveal viewing of all items. In Experiment 1, there was some evidence that factors associated with the visual clarity of the items were more important at the shortest exposure durations. Finally, the appearance of the illuminated marker field directly after the array in

Experiment 1 probably caused some visual masking, again resulting in poorer registration of material. The absence of visual masking may also explain why accuracy was greater in Experiment 2, even though the delays used would presumably increase task difficulty. All of these characteristics of the task would be especially effective in reducing clarity in the more peripheral parts of the field, where acuity is poorer, and for sets of items containing many shared contour features.

Whichever interpretation is more appropriate, it appears that the variables affecting spatial information extraction are quite different from those affecting the processing of figural or contour information. The effects of retinal position varied widely across report conditions, suggesting that differences in clarity associated with retinal position quickly dissipate in the face of increased storage requirements or that preferences in spatial ordering of material are so strong that they frequently override other factors. The registration of contour information, on the other hand, appears to be much more dependent upon the manner in which items are displayed than upon the requirements of the task or the

preferences of the O.

REFERENCES

- AVERBACH, E., & CORIELL, A. S. Short-term memory in vision. *Bell Systems Technical Journal*, 1961, 60, 309-328.
- BRYDEN, M. P. Symmetry of letters as a factor in tachistoscopic recognition. *American Journal of Psychology*, 1968, 68, 513-524.
- CIMBALO, R. S., & LAUGHERY, K. R. Short-term memory: Effects of auditory and visual similarity. *Psychonomic Science*, 1967, 8, 57-58.
- CONRAD, R. Acoustic confusion in immediate memory. *British Journal of Psychology*, 1964, 55, 75-84.
- CROVITZ, H. R., SCHIFFMAN, H., LIPSCOMB, D. B., POSNICK, G., REES, J., SCHAUB, R., & TRIPP, R. Identification and localization in the letter span. *Canadian Journal of Psychology*, 1966, 20, 455-461.
- GIBSON, E. J. *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts, 1969.
- GLANZER, M. Encoding in the perceptual (visual) serial position effect. *Journal of Verbal Learning & Verbal Behavior*, 1966, 5, 92-97.
- HABER, R. N., & STANDING, L. G. Location of errors with a post-stimulus indicator. *Psychonomic Science*, 1969, 17, 345-346.
- HARCUM, E. R. Parallel functions of serial learning and tachistoscopic pattern perception. *Psychological Review*, 1967, 74, 51-63.
- HERON, W. Perception as a function of retinal locus and attention. *American Journal of Psychology*, 1957, 70, 38-48.
- HERSHENSON, M. Stimulus structure,

- cognitive structure and the perception of letter arrays. *Journal of Experimental Psychology*, 1969, 79, 327-335.
- KEELE, S. W., & CHASE, W. G. Short-term visual storage. *Perception & Psychophysics*, 1967, 2, 383-386.
- KLEMMER, E. T. Perception of linear dot patterns. *Journal of Experimental Psychology*, 1963, 65, 468-473.
- LAUGHERY, K. R., & HARRIS, G. J. Visual and auditory intrusion errors in short-term memory. *Journal of Experimental Psychology*, 1970, 83, 101-107.
- MACKWORTH, J. F. The duration of the visual image. *Canadian Journal of Psychology*, 1963, 17, 62-81.
- NEISSER, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.
- POSNER, M. I., BOIES, S. J., EICHELMAN, W. H., & TAYLOR, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology*, 1969, 79, (1, Pt. 2).
- SPERLING, G. A. A model for visual memory tasks. *Human Factors*, 1963, 5, 19-31.
- UNDERWOOD, B. J., & SHULTZ, R. W. *Meaningfulness and verbal learning*. New York: Lippincott, 1960.
- WHITE, M. J. Identification and localization within digit and letter spans. *Psychonomic Science*, 1969, 14, 279-280.
- WINER, B. J. *Statistical principles of experimental design*. New York: McGraw-Hill, 1962.

NOTE

1. These matrices are available from the author upon request.

(Received for publication July 24, 1971.)