Memory & Cognition
1977, Vol. 5 (1), 146-150

Recognition vs recall of visually vs acoustically confusable letter matrices

EVELYN VINGILIS, JOANNA BLAKE, and LEONARD THEODOR
Department of Psychology, York University, 4700 Keele Street, Downsview, Ontario, Canada

In an attempt to separate auditory and visual components in short-term memory, five subjects were exposed to letter matrices composed of six visually confusable letters, six acoustically confusable letters, or a mixture of the two, under two response conditions: recognition and recall. A 50-msec stimulus presentation was followed by a variable dark interval of 1, 250, 1,000, or 3,000 msec. In the recall condition, the interval was followed by a buzzer which signaled the subject to recall, in any order, as many letters as possible. In the recognition condition, the variable interval was followed by a second letter matrix which was either identical to the first matrix or differed from is by one letter. Subjects responded either "same" or "different." The results support the notion that the auditory component plays a major role in recall, whereas the visual component dominates in recognition.

A common tenet of current information processing models is that items in short-term memory (STM) are auditorily encoded (Conrad, 1964; Glanzer & Clark, 1964; Sperling, 1963, 1967), despite the fact that many authors point to the possibility of visual encoding (e.g., Neisser, 1967). Sperling's (1963) model states that items that enter visual information storage (VIS) are scanned and read out into auditory information storage (AIS) and rehearsed in AIS. The discovery of the detrimental effects of acoustically confusing stimuli on recall performance (Conrad, 1964; Conrad & Hull, 1964) has provided much of the empirical basis for models of short-term memory which incorporate an AIS.

Studies done on visual memory search tasks, however, weaken the hypothesis that STM is only auditory in nature. Neisser (1967) found that when the context letters were made highly visually confusable with the target letters, search time increased, showing a strong interfering effect of visual context. Gibson and Yonas (1966) also discovered that a highly confusable visual context significantly reduced scanning rate, but a highly confusable acoustic context played over earphones had no effect.

Tversky (1969) attempted to vary encoding modality, pictorial or verbal, by manipulation of subjects' expectations of the way material was to be used. She found that subjects were able to encode schematic faces presented pictorially or named verbally in either modality depending upon expectation.

Encoding modality may depend upon whether the task involves recall or reproduction, usually verbal, or recognition, usually nonverbal. Cohen and Granstrom (1968) studied the retention of visual figures in STM. The mode of recall and the type of material interpolated during retention was varied. Reproduction was as good as recognition when the retention interval was empty, but was inferior with an interpolated paired associate learning task involving either visual figures, auditory words, or both. The conclusions were drawn that "short-term reproductive memory is mainly verbal whereas short-term recognition memory is mainly nonverbal. This nonverbal type of memory does not exhibit properties of a fast decaying sensory visual trace and is therefore postulated to be a third type of store, over and above the brief sensory visual storage and auditory short-term memory" (p. 653). Gibson and Yonas (1966) have drawn similar conclusions from their experiments. They state that "it is possible that such encoding (auditory) may occur in a recall task, when rehearsal is attempted, but no such strategy appears to be taking place when the task is one of detection" (p. 164).

Consequently, it appears that there is a visual component in STM and that this visual form of encoding may be more prevalent in recognition tasks than in recall tasks, where the very nature of the task usually requires verbal report, and, therefore, verbal encoding. The present study compared recall of letter matrices with a recognition analogue in which subjects judged whether two consecutively displayed letter matrices were identical. The matrices, of six letters each, were designed to be either visually confusing, auditorily confusing, or a mixture (control).

The following predictions were made: (1) Following previous results with similar recall tasks, poorer performance in recall was expected with acoustically confusing matrices than with visually confusing matrices. (2a) If, in recognition, as in the AIS theory of STM claims, one is also scanning and reading letters into AIS in order to compare them to the second matrix, then

This research was supported by NRC Grant A8575 to the second author and by NRC Grant APA-300 to the third author. Reprint requests should be sent to Evelyn Vingilis, Department of Psychology York University, 4700 Keele, Downsview, Ontario, Canada.

poorer performance would similarly be expected with matrices of high acoustic confusability. (2b) However, if recognition does not require verbal encoding, and, if a visual component in short-term memory exists, as Cohen and Granstrom's (1968) study suggests, then poorer performance would be expected with visually confusing matrices.

The present study also varied the interval between the two consecutively presented letter matrices in the recognition task. Even verbal STM models would seemingly predict inferior performance with visually confusing letters at intervals of less than 1 sec, since the letters would still be in visual sensory memory (Sperling, 1963). At intervals longer than 1 sec, according to these models, a shift from inferior performance on visually similar letters to inferior performance on acoustically similar letters would be expected, since by 1 sec, the trace would usually be encoded in auditory form. On the other hand, if visual confusions were found throughout the VIS and STM stages, this would be evidence of a visual coding system in STM or a much longer VIS.

METHOD

Design

The experiment consisted of two tasks: recall and recognition. In recall, the subject was required to name as many letters as possible from the matrix he had just observed. In recognition, the subject was shown two consecutively presented matrices and his task was to decide whether the two matrices were "same" or "different." The first stimulus matrix was presented for 50 msec, followed by a variable dark interval and a buzzer to signal the subject to respond in the case of recall or a second matrix presented for 50 msec in the case of recognition. The dark intervals were 1, 250, 1,000, and 3,000 msec. One-third of the stimulus matrices contained visually similar letters, one-third contained acoustically similar letters, and one-third was a mixed set, included as a control. Four blocks of 36 trials, 12 of each matrix type, were administered at each interstimulus interval (ISI) to every subject under both recognition and recall, for a total of 1,152 trials. In each test session, subjects received one block of trials at each ISI under one of the tasks. The tasks were alternated between sessions, and the order of tasks counterbalanced across subjects. The order of ISIs at each session was randomized separately for each subject. The order of matrix type within each block of trials was random, with the constraint that successions of any type be limited to three trials. In each block of recognition trials, half of the trials for each matrix type were the same and half were different, in random order, but again with no succession of either same or different greater than three trials.

Subjects

The subjects were five university students, one female and four male, naive as to the purpose of the experiment. The ages of the group ranged from 21-25 years, with a mean age of 22.2 years. The students were all native English-speaking Canadian[1] undergraduates at York University and were paid for their participation.

Apparatus

A standard Iconix four-field tachistoscope with associated control logic was used. In recognition, the first three fields of the tachistoscope contained a fixation field, the first stimulus matrix, and the second stimulus matrix, respectively. In recall, Field 1 contained the fixation field and Field 2 the stimulus matrix. The luminance of the three fields was approximately 1.30 log fL for the fixation field, 1.30 log fL for the first stimulus field, and 1.32 log fL for the second stimulus field.

In recognition, the clock counter started at the end of the last stimulus presentation and was stopped by means of a voice key triggered by the subject's verbal response. Response latency was measured in milliseconds.

Materials

The stimuli were 7.2 cm x 4.8 cm matrices containing capital letters, one in each cell. The size of each letter was approximately 12 mm in height, subtending a visual angle of .65 deg at a field-subject distance of 864 mm. The letters were drawn in black ink with a Rapidograph pen, with the aid of a Pickett lettering guide. The visually confusing set consisted of H K L T W X Z and the acoustically confusing set consisted of B G T F S X M. Letters for the visual set were first selected on the basis of the general distinguishing feature of "angularity," that is, letters with all straight lines (Bowma, 1971; Gibson, 1965; Neisser, 1967). Then, using two of Gibson's, Neisser's, and Bowma's categories within the angular set, two subsets were defined: one including angular letters with vertical features (H T L K), and the other including angular letters with oblique features (X W Z).[2] Letters that were highly acoustically confusing were kept to a minimum. Letters for the acoustic set were selected on the basis of the shared phoneme e (Conrad, 1964; Wickelgren, 1965a). Two subsets were constructed: one containing the shared phoneme ĕ (F S X M), and the other containing the shared phoneme ē (B T G). Similarly, letters that could be considered visually confusing were eliminated as much as possible.

Three sets of matrices were made: (1) a visually confusing set containing six visually confusable letters, (2) an acoustically confusing set containing six acoustically confusable letters, and (3) a mixed set, containing three letters from the visually confusing set and three letters from the acoustically confusing set. On the mixed matrices, the letters from each set were alternated with each other in a $\begin{smallmatrix}V&A&V\\A&V&A\end{smallmatrix}$ or $\begin{smallmatrix}A&V&A\\V&A&V\end{smallmatrix}$ fashion. The matrices were generated randomly with the following constraints: (1) Each letter appeared only once in the matrix; (2) each letter appeared approximately the same number of times in each position across all trials; (3) the different element was always selected from the same set as the other letters of the matrix.[3]

Procedure

Practice session. A card with 12 stimulus letters arranged alphabetically was placed in front of the subject and remained there throughout all sessions. Subjects were instructed that the letters could be recalled in any order. For recognition, they were told that the two matrices would either be identical to each other, or that the second matrix would have one letter different from the first matrix. It was specified that (1) no letter would appear twice on an individual card, (2) all the unchanged letters on the second matrix would be in the same position as they were on the first matrix, and (3) the changed letter would occur in each position an equal number of times. Subjects were asked to respond "same" or "different" verbally as quickly as they could after the last stimulus matrix had been presented. One practice session was devoted to recall and the other practice session was devoted to recognition, three subjects beginning with the recall task and two subjects beginning with the recognition task. For each task, subjects were given 12 practice trials for each matrix type at each ISI.

Test sessions. Eight test sessions were administered, the sessions alternating between the recognition and recall tasks and the order of tasks counterbalanced across subjects. For each trial sequence, the fixation field was shown for 1 sec, followed by a dark interval of 100 msec. The first stimulus matrix then appeared for 50 msec, followed by a dark interval of 1, 250, 1,000, or 3,000 msec. In recognition, the second stimulus matrix

was then presented for 50 msec. In recall, a buzzer was presented for 150 msec. The subject was then required to respond either by saying "same" or "different" in the case of recognition or by verbally recalling as many letters as possible from the matrix just presented. The letters recalled were recorded in the order reported by the subject.

RESULTS

Recall Accuracy

Recall trials were scored as correct if all letters presented were recalled correctly regardless of order. The all-or-none scoring allowed for greater comparability of recall with recognition. The mean percentage of correct trials for each matrix type at each delay for each subject was subjected to an arc sine transformation. A 3 (matrix type) by 4 (ISI) analysis of variance was performed on the transformed mean percentages. The main effects of matrix type [$F(2,8) = 25.40$, $p < .01$] and ISI [$F(3,12) = 4.74$, $p < .05$] were significant, but they did not interact significantly. The graph of accuracy for each matrix type as a function of ISI is presented in Figure 1. Comparison of means using the Tukey HSD procedure (Winer, 1962) showed that acoustically confusing matrices (AM) were recalled more accurately than either the mixed matrices (MM) ($p < .01$) or the visually confusing matrices (VM) ($p < .01$), while the MM and VM did not differ significantly from each other. Accuracy was higher at 250 msec than at 3,000 msec ($p < .05$); no other ISI comparisons were significant.

There were significantly more within-sets intrusion errors for the AM but not for the VM and MM. For AM, 76% of the intrustion errors were chosen from the acoustic set vs 24% from the visual set ($t = 3.00$, $p < .05$). For VM, 56% of the intrusion errors were from the visual set vs 44% from the acoustic set ($p > .1$).

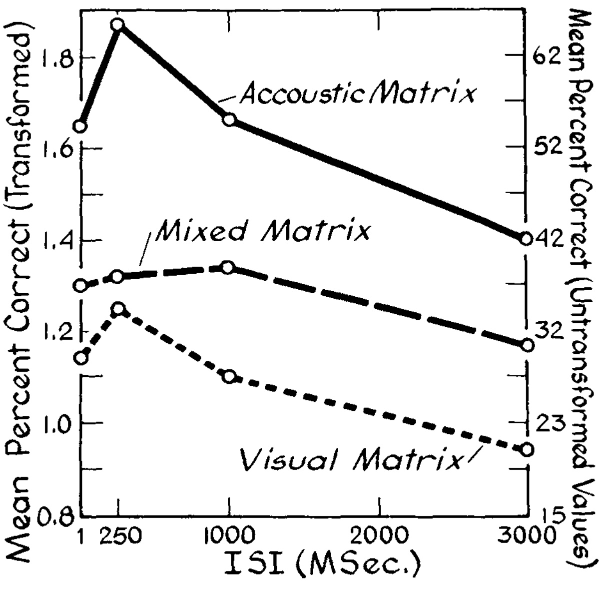

Figure 1. Transformed mean percent correct in recall as a function of ISI.

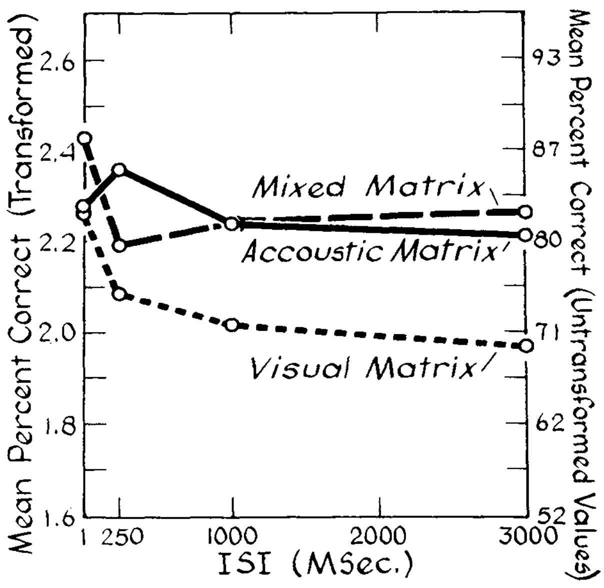

Figure 2. Transformed mean percent correct in recognition as a function of ISI.

For MM, 58% were from the acoustic set, 42% from the visual set ($p > .1$). However, for the AM, within-sets intrusion errors were not selected significantly more often from the same phonemic subset (ĕ or ē) as the original letter. Similarly, for the VM, intrusion errors did not tend to come from the same graphemic subset (diagonal or oblique) as the original letter. When intrusion errors were analyzed across as well as within sets, graphemic confusions were somewhat more frequent than phonemic confusions. This reflected the fact that there were slightly more pairs that might be potentially graphemically confusing than phonemically confusing. Both phonemic and graphemic confusions were more frequent than unrelated intrusion errors ($t = 3.90$, $p < .01$, and $t = 8.60$, $p < .001$).

Recognition Accuracy

The mean percentage of correct same and different trials for each subject for each matrix type at each delay was subjected to an arc sine transformation. A 3 (matrix type) by 4 (ISI) by 2 (same-different judgment) analysis of variance was performed on the transformed scores. The only significant effect was a main effect of matrix type [$F(2,8) = 9.63$, $p < .01$]. A graph of accuracy for each type of matrix as a function of ISI is presented in Figure 2. A comparison of means demonstrated that both the AM and MM had significantly higher scores than the VM ($p < .05$), while the MM and AM did not differ significantly from each other. There were no significant interactions.

Recognition Latency

Latencies of correct responses were transformed individually by taking $\log(X + 1)$. A 3 (matrix type) by 4 (ISI) by 2 (same-different trials) analysis of variance was performed on the transformed scores.

Table 1
Log-Latency Means for Recognition by Matrix Type and Same-Different Trials at Each Interstimulus Interval

Trials	ISI	Matrix Type: Acoustically Confusing	Visually Confusing	Mixed
Same	1	.23262	.24908	.23108
	250	.24806	.25160	.23976
	1000	.27852	.32260	.28466
	3000	.35668	.35494	.32656
Different	1	.24338	.25878	.23976
	250	.24398	.23664	.24118
	1000	.28378	.29766	.29526
	3000	.32114	.33126	.32626

The log-latency means by ISI, matrix type, and same-different trials are given in Table 1. Significant main effects were obtained for ISI [$F(3,12) = 20.79$, $p < .01$] and matrix type [$F(2,8) = 6.91$, $p < .05$]. A comparison of means showed that log latencies for the VM were significantly longer than log latencies for the MM and AM ($p < .05$), while the AM and MM did not differ significantly from each other. Log latencies at 3 sec were significantly longer than at 1 msec and at 250 msec ($p < .01$). and at 1 sec ($p < .05$), and significantly longer at 1 sec than at 1 msec and 250 msec ($p < .05$). The log latencies at the 1-msec and 250-msec ISIs did not significantly differ from each other. There were no significant interactions.

DISCUSSION

The most obvious interpretation of the results is that STM is primarily visual in nature, since VM elicited the poorest performance under all conditions (i.e., Predictions 1 and 2a were not confirmed). However, further consideration suggests a far more complex system of memory. It was predicted that acoustically similar letters would be more difficult to recall than acoustically dissimilar letters. However, in recall, accuracy was in fact highest on the AM, while the MM and VM were lowest in accuracy and did not differ significantly from each other. This would seem to suggest that acoustic similarity facilitated rather than hindered performance. In fact, Wickelgren (1965b) has argued that if STM is associative in nature, phonemically similar lists would be more difficult only in *ordered* recall, since there would not be enough distinctive associations to differentiate positions and facilitate correct order. However, "*free* recall could well be facilitated since the shared phoneme is certain to be recalled and direct associations exist from the representative of the shared phoneme in the list" (pp. 568-569). The associative factor present in Wickelgren's theory of STM for recall could account for the better item accuracy on acoustically similar letters. Furthermore, significantly more intrusion errors on the acoustic matrices were from the acoustic set, although there was no significant difference in phonemic similarity vs dissimilarity, that is, the erroneous letter did not tend to have the same phoneme, ĕ or ē, as the stimulus letter. This would suggest that subjects "caught on" to the acoustic set.

Consistent with Sperling's (1963) findings, accuracy in recall decreased with increases in ISI from 250 msec to 3 sec. The decrease was the same for all matrix types, suggesting that the same method of encoding was used across all matrix types.

There are two basic differences in accuracy under recognition as compared with recall. First, there was no significant decrease in performance over ISI, as there was in recall, suggesting that subjects were relying upon different modality codes in recognition and in recall. Second, in recognition, the VM elicited the poorest performance, while the MM and AM elicited the best performance and did not differ significantly from each other. Thus, whereas, in recall, VM did not differ from the MM (control), both being inferior to AM, in recognition, AM did not differ from the MM (control), both being superior to VM. These findings suggest that, in recognition, visual similarity *hindered* accuracy, as compared with recall, in which acoustic similarity *facilitated* accuracy.

Admittedly, the varying interpretations of the inferior performance on the VM in recall and in recognition seem contradictory. It could be argued that Wickelgren's associative theory could also explain the superior performance on the AM in recognition. However, it is not possible to explain on the basis of this theory the reason that, in recognition, the MM are not significantly different from the VM. Furthermore, previous research on recognition and recall has demonstrated that organization, high association, and high frequency (Cofer, 1967; Dale, 1967; Shepard, 1967) are all good requisites for good recall, but have no effect on recognition. In the present study, matrices with high phonemic associations elicited the highest performance in free recall. However, performance on these matrices was not superior to the MM (control) in recognition.

There was a significant increase in log latencies for correct recognition with increase in ISI. The difference occurred around the 1-sec interval, suggesting that processing after that interval is occurring at a more "central" level, that is, STM. According to an AIS theory of STM, if items are visually perceived, scanned, and then read out into AIS after 1 sec (Sperling, 1963), one would expect a drastic shift in performance on the AM from good performance in VIS at 1 msec and 250 msec to poor performance in AIS after 1 sec, but this was not confirmed. At each ISI, the longest latencies occurred for the VM, again suggesting that processing in recognition was most difficult for visually similar letters.

Thus, the results seem to support the view of Cohen and Granstrom (1970) that "different mechanisms or stores are predominant with two modes of recall, visual in the case of recognition and verbal in the case of reproduction" (p. 456). It is not claimed that verbal information is never used in making recognition judgments. If the experiment is biased enough, the use of verbal information can be just as great as it is in recall (Cohen, 1966, 1967). What is suggested, however, is that verbal information is not necessary under the kinds of conditions used in recognition or detection. Furthermore, the results also imply that visual information may be used in arriving at recall responses. The fact that the percentages of phonemic vs graphemic intrusions were not significantly different from each other, but were both significantly higher than the unrelated intrusions, would suggest that subjects were systematically making both phonemic and graphemic errors. Bahrick and Boucher (1968) consider it unlikely that the verbal store is separate from the visual store, and, in fact, the present data are not entirely inconsistent with their viewpoint. Thus, the initial encoding may be into a visual store, with verbalizing entering into the memory process only during recall, which could be thought of as a "verbal decoding from the visual store."

In any case, it seems quite clear that recall and recognition in STM involve different processes to some extent. The auditory component that seems to form an integral part of recall has not been shown to play a significant role in recognition. The exact mechanisms involved in the visual component of recognition memory are still open to inquiry.

REFERENCES

Bahrick, H. P., & Boucher, B. Retention of visual and verbal codes of the same stimuli. *Journal of Experimental Psychology*, 1968, **78**, 417-422.

Bowma, G. Confusability of lower case matrices. *Vision Research*, 1971, **5**, 9-26.

Cofer, C. N. Does conceptual organization influence the amount retained in immediate free recall? In B. J. Kleinmuntz (Ed.), *Concepts and the structure of memory*. Wiley, New York: 1967. Pp. 181-214.

Cohen, R. L. Effect of verbal labels on the recall of a visually perceived simple figure: Recognition vs. reproduction. *Perceptual Motor Skills*, 1966, **23**, 859-862.

Cohen, R. L. Interaction between a visually perceived simple figure and an appropriate verbal label in recall. *Perceptual Motor Skills*, 1967, **24**, 287-292.

Cohen, R. L., & Grandstrom, K. Interpolated task and mode of recall as variables in STM for visual figures. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 653-658.

Cohen, R. L., & Grandstrom, K. Reproduction and recognition in short-term visual memory. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 450-457.

Conrad, R. Acoustic confusion in immediate memory. *British Journal of Psychology*, 1964, **55**, 75-84.

Conrad, R., & Hull, A. J. Information, acoustic confusion, and memory span. *British Journal of Psychology*, 1964, **55**, 429-432.

Dale, H. C. A. Response availability and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 47-48.

Gibson, E. Learning to read. *Science*, 1965, **148**, 1066-1072.

Gibson, E. J., & Yonas, A. A developmental study of the effects of visual and auditory interference on a visual scanning task. *Psychonomic Science*, 1966, **5**, 163-164.

Glanzer, M., & Clark, W. H. Accuracy of perceptual recall: An analysis of organization. *Journal of Verbal Learning and Verbal Behavior*, 1964, **1**, 289-299.

Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

Shepard, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 156-163.

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

Sperling, G. Successive approximations to a model for short-term memory. *Acta Psychologica*, 1967, **27**, 285-292.

Tversky, B. Pictorial and verbal encoding in a STM task. *Perception & Psychophysics*, 1969, **6**, 225-233.

Wickelgren, W. Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology*, 1965, **70**, 102-108. (a)

Wickelgren, W. Short-term memory for phonemically similar lists. *American Journal of Psychology*, 1965, **78**, 567-574. (b)

Weiner, M. Phonemic and visual coding in short-term memory. Unpublished doctoral dissertation. Yeshiva University, 1970.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

NOTES

1. The letter Z was used in terms of British and Canadian pronunciation (zed) as opposed to the American pronunciation (zee). Consequently, Canadian subjects had to be used.

2. The visual similarity of the letters within the two subsets receives validational support from a visual confusion table constructed from subjects' ratings of visual similarity of letters (Weiner, 1970).

3. In order to determine if the 48 mixed-matrix cards generated randomly were actually a mixture of visual and acoustic confusions, the number of each type of potential confusion on each card was computed. In the computation, between-sets as well as within-sets confusions were included. Letters were considered to be visually confusing on the basis of verticality (H, K, L, T, F), diagonality (W, X, Z, M), and roundness (B, G, S). They were considered to be acoustically confusable on the basis of the phonemes ĕ (L, F, S, X, M), ē (B, G, T), s (Z, S, X), ā (H, K), and k (K, X). Across all 48 cards, there were 226 acoustic confusions (mean = 4.7 per card) and 248 visual confusions (mean = 5.2 per card). Out of 48 cards, 41 contained a fairly equivalent number of each type of confusion: Nine had an equal number of each type, 10 had one more acoustic than visual confusions, 10 had one more visual than acoustic confusions, 3 had two more acoustic than visual, and 9 had two more visual than acoustic. Of the remaining seven cards, two had three more acoustic than visual confusions, four had three more visual than acoustic, and one had four more visual than acoustic. Thus, although there were slightly more visual than acoustic confusions across all mixed matrices, on the whole, this computation indicates that there was a good mixture of both types of confusions.

(Received for publication September 24, 1976.)