Memory & Cognition
1973, Vol. 1, No. 2, 164-168

The effects of auditory shadowing on recognition of information received visually*

HOWARD A. ROLLINS and ROBERT THIBADEAU†
Emory University, Atlanta, Georgia 30322

When Ss attend to one auditory message, they have no permanent memory for a second auditory message received simultaneously. Generally, it has been argued that a similar effect would occur crossmodally. This hypothesis was tested in the present experiment for messages presented to visual and auditory modalities. All Ss were tested for recognition of information presented either while shadowing or while hearing but not shadowing a passage of prose presented to one ear. One group heard a list of concrete nouns in their other ear. Three other groups received (1) printed words, (2) pictures of objects easily labeled, or (3) pictures of objects difficult to label. The shadowing task produced a decrement in recognition scores for the first three groups but not for the group receiving pictures of objects difficult to label. Further, the shadowing task interfered more with information received auditorily than with any form of visual information. These results suggest that information received visually is stored in a long-term modality-specific memory that may operate independently of the auditory modality.

Students of perception and memory have long been interested in whether or not the human organism can attend to and process more than a single incoming message at one time (cf. Broadbent, 1958; Treisman, 1969). Cherry (1953) and Moray (1959) have shown that Ss cannot process simultaneous *auditory* messages. If Ss are forced to attend to one auditory message by shadowing (i.e., by repeating back each item in that message as they hear it), they have no memory for the contents of a second auditory message presented simultaneously. It is generally assumed that the two messages compete for access to a limited capacity processor. Since this processor is engaged in the processing of the shadowed message, the second message is not processed and decays rapidly in a short-term sensory buffer (cf. Glucksburg & Cowan, 1970; Norman, 1969).

Treisman (1969) has questioned whether the same blocking effect would occur crossmodally. Clearly, two auditory inputs must share the same processing system throughout. Thus, these two messages may compete for access to any one of several limited capacity or serial processing subsystems. However, at least a part of the processing of two messages presented to separate sensory modalities would occur in sensory-dependent, and therefore independent, processing systems. At one extreme, the two sensory systems may process and store messages independently of one another with no system overlap or interaction. If this is true and if the two systems can operate simultaneously (in parallel), then shadowing in one modality should not interfere with the processing of a message in the other. At the other extreme, the two sensory systems may share certain serial processing subsystems. If this were true, crossmodal blocking should be as severe as intramodal blocking.

Both Mobray (1964) and Kroll, Parks, Parkinson, Bieber, and Johnson (1970) have shown that auditory shadowing produces poorer recall of a second auditory message than an equivalent visual message. However, these studies did not examine the degrees of blocking relative to no-shadow controls.

In the present experiment, then, crossmodal blocking effects were assessed for messages presented to visual and auditory modalities. All Ss were tested for recognition of one message presented either while shadowing or while hearing but not shadowing a passage of prose presented to one ear. One group heard a list of concrete nouns in their other ear. This group served as a control for crossmodal blocking and as a replication of Cherry (1953) and Moray (1959). All other groups received a second message visually.

The degree of system overlap between these two modalities, and therefore the level of crossmodal blocking, may well depend upon the type of information presented. Verbal information may be routed to a single sensory-independent system regardless of source of input. For example, Ss may process a visual or an auditory version of a word by encoding and storing it in an auditory memory (Sperling, 1963; Wickelgren, 1966). For verbal information, then, one might expect the same degree of blocking across modalities as within a modality. On the other hand, pictures of objects may be stored directly in some permanent visual memory (cf. Bower, 1970). Here one would predict little crossmodal blocking, particularly if the objects used have no verbal labels. In order to examine the effects of auditory shadowing on different types of visual input, groups of Ss in the present experiment received three types of information visually: (1) printed words, (2) pictures of

*This research was supported by Grant 4235 from Emory University to the first author. Reprint requests should be sent to H. Rollins, Department of Psychology, Emory University, Atlanta, Georgia 30322.

†Now at the University of Virginia, Charlottesville, Virginia 22904.

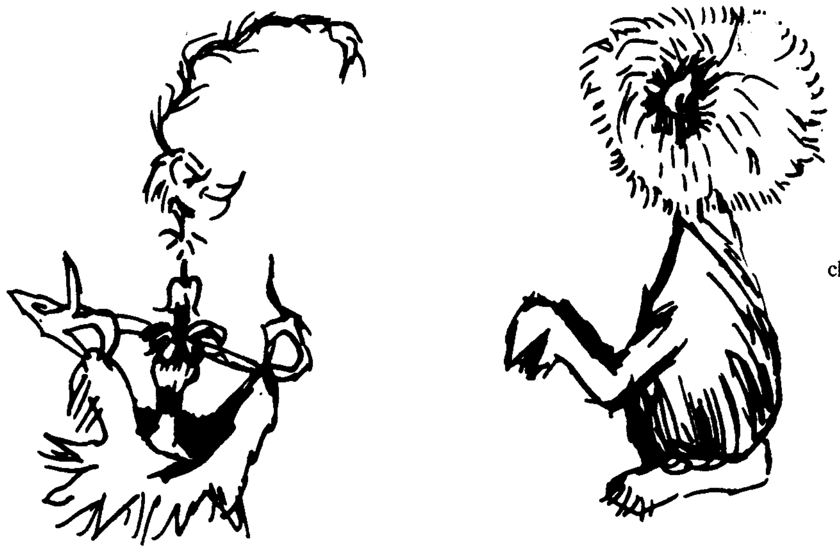

Fig. 1. Two examples of fictitious-character stimuli.

objects labeled by these printed words, or (3) pictures of objects without labels.

METHOD

Design

Each S received, in order, a passage of prose to one ear accompanied by a list of to-be-remembered (TBR) items, a recognition test for the TBR items, a second passage of prose accompanied by another TBR list, and a second recognition test. The S shadowed one of the prose passages but not the other. Four groups of Ss received one of the following types of TBR items: (1) words auditorily, (2) words visually, (3) pictures of common objects, or (4) pictures of fictitious characters. The shadowing condition was, therefore, a repeated measure factor and input condition was a between-groups factor.

Subjects

The 64 Ss were among those students at Emory University enrolled in an introductory psychology course. Students earned course credit for their participation. Sixteen Ss were assigned randomly to each of the four groups. Half (8) of the Ss in each group were male.

Apparatus and Materials

Auditory stimuli were presented to S with a Sony (Model TC-540) stereo tape recorder and a pair of Realistic-Pro stereo headphones. Passages of prose selected from the *United States Articles of Confederation* were recorded on one track of the tape. Highly redundant phrases such as "in the congress assembled" were either deleted or rephrased. On the other track of the tape, E recorded lists of 24 nouns (names of concrete objects) at a rate of one noun every 2.6 sec. Four lists of words, matched for frequency of occurrence (A or AA) in the Thorndike-Lorge count were used. Two lists were recorded on tape and served as TBR items; all lists were also printed on index cards for recognition testing. A switch built into the earphone line permitted presentation of the passage of prose alone or both prose and word list.

Visual stimuli were presented by an Anscorama (Model 970) slide projector. The projector was set to present one slide every 2.6 sec. Each slide was in view for 1 sec. Three sets of visual stimuli were constructed by printing or drawing stimuli on white 3 x 5 in. white index cards. These stimuli were then photographed using Kodak Pantomic X film for direct negative reversal.

The first visual set consisted of printed versions of the words in the auditory set. These 96 words were printed in ½-in. block type. Each of the 96 words in the above set labeled a common object. The second visual set, then, was constructed by drawing stylized pictures of these objects. The final set of visual stimuli consisted of 96 drawings of fictitious characters modeled from various Dr. Seuss books such as *On Beyond Zebra, If I Ran the Zoo,* and *If I Ran the Circus.*[1] These stimuli were selected so as to be highly discriminable but lacking ready verbal labels. All characters were drawn freehand, and numerous alterations were made to render the characters more difficult to label. Two examples of these stimuli are shown in Fig. 1.

Procedure

The Ss were given detailed instructions about the tasks they were to perform. They were told that they would hear a passage of prose in one ear and simultaneously hear a list of words in the other ear (words auditory), see a list of words (words visual), or see a series of pictures (common objects or fictitious characters). The shadowing task was described, and S was given 3 min of practice shadowing a description of the game of baseball found in the *Lincoln Library* (26th ed., 1963, Frontier Press Co.). Each S was told that he would be expected to recognize as many of the TBR items as he could and that he would not be tested for recall of the passage of prose. This procedure was necessary since each S was tested twice and would thus know what to expect on the second test.

Each S received a list of 24 TBR items (under one of the four input conditions) followed by a recognition test containing the 24 TBR items and 24 distractor items. A second, different list of TBR items (same input condition) and a recognition test with 24 new distractors followed the first. Each list presentation began 15 sec after initiation of the prose passage. For one of the two lists, S shadowed the passage of prose; for the other, S received but did not shadow the prose. List (1 or 2) and order of shadow condition were systematically counterbalanced across Ss within groups.

Recognition testing began immediately after the last TBR item was presented. The E presented 48 stimuli, 1 at a time, from one of six well-shuffled decks of 48 index cards (words auditory and words visual groups received the same recognition test). The S responded "yes" or "no" to each card, depending upon whether or not he recognized the item. The rate of card

Table 1
Proportion of Hits and False Alarms and d′ Values as a Function of Shadow and Input Condition

	No-Shadow Condition			Shadow Condition			
Input Condition	Mean Hits	Mean False Alarms	d′	Mean Hits	Mean False Alarms	d′	d′ Difference
Auditory Words	.80 (.84)	.06	2.39 (2.52)	.37 (.39)	.22	0.44 (0.49)	1.95 (2.03)*
Visual Words	.78 (.83)	.08	2.14 (2.32)	.56 (.54)	.16	1.14 (1.09)	1.00 (1.23)*
Common Objects	.93 (.94)	.03	3.38 (3.44)	.70 (.78)	.06	2.05 (2.29)	1.33 (1.15)*
Fictitious Characters	.54 (.57)	.16	1.12 (1.18)	.42 (.44)	.17	0.77 (0.80)	0.35 (0.38)†

*$p < .01$ †$p > .05$

presentation was paced by S and averaged 2-4 sec per card. Responses were scored for hits and false alarms.

RESULTS

The proportions of items correctly recognized (hits) and the number incorrectly labeled as in the TBR set (false alarms) were used to determine d′ values for each condition, following Gourevitch and Galanter (1967). The proportions of hits and false alarms and the corresponding d′ values are shown in Table 1 for all input groups under shadow and no-shadow conditions.

In order to determine the effects of shadowing on the various types of input, an analysis was performed comparing all eight d′ values in Table 1 (cf. Marasciulo, 1970).[2] These d′ values were reliably different [$\chi^2(7) = 350.76$, $p < .01$]. Specific contrasts were made (holding the significance level constant) between shadow and no-shadow d′ values for each input condition. As indicated in Table 1, the shadowing task interfered with the processing and storage of words whether the words were presented auditorily or visually and also with the processing of pictures of common objects. In contrast, the decline in d′ value observed for the group receiving pictures of fictitious characters was not reliable. These results are strengthened by a simple count of the number of Ss (of 16) in each input condition having higher d′ values when shadowing than when not shadowing. For groups receiving words auditorily, words visually, and pictures of common objects, there were none, one, and two Ss, respectively, recognizing more items under the shadow condition. Each of these frequencies differs from chance by sign test ($p < .01$). However, six Ss receiving fictitious characters had higher d′ values when shadowing ($p > .22$).

These data suggest that attending to an auditory message interferes with the processing and storage of any information whether visually or auditorily presented when that information can be verbally labeled. The absence of interference for the objects without labels suggests that information can be stored in a visual memory system that operates independently of processing in the auditory system.

The relative degree of interference produced by auditory shadowing is also of theoretical importance. For example, if words presented visually are transformed into a verbal equivalent and stored in an acoustic or auditory memory, then one would predict that the shadowing task would interfere as much with recognition of words presented visually as with words presented auditorily. This was clearly not true in the present experiment. Under the no-shadow condition, Ss receiving words auditorily recognized as many items (d′) as did Ss receiving words visually. Yet the interfering effects of the shadowing task were much larger for words presented auditorily than for words presented visually ($p < .05$) using a contrast procedure (Marasciulo, 1970).

It is possible that the above results are in part caused by differences in the amount of information retained in short- rather than in long-term memories. For example, Ss shadowing an auditory message may more easily hold information in some short-term memory when the information is received visually. However, any such effects should occur primarily for items presented near the end of the list, since the duration from input to output for early items probably exceeds the time limits of these short-term memories (cf. Neisser, 1966). The hit and d′ values in parentheses in Table 1 indicate recognition accuracy over the first 16 items of input (false alarm rates based upon all 24 foils). The d′ values for the first two-thirds of input are nearly identical to those for total input. Analyses on the d′ values for Items 1-16 yielded identical results to those reported above. Thus, the results of the present study are clearly a function of long- rather than of short-term storage processes.

Previous research with a recognition paradigm (Moray, 1959) has shown that there is no long-term memory for auditory items received while shadowing a second auditory message. In the present experiment, Ss receiving words auditorily had d′ values significantly higher than zero when shadowing ($Z = 4.63$, $p < .01$) even when the last eight input items were deleted ($Z = 4.08$, $p < .01$). Further, only 3 Ss of the 16 in this group had negative d′ values ($p < .01$). Thus, it would appear that Ss may process and store at least a few of the auditory items received while shadowing a second auditory message.

A second dependent variable that corrects for guessing (hits minus false alarms) was also analyzed. The results of an analysis of variance with this variable were identical to analyses on d′. The interaction of input and shadow conditions was reliable [$F(3,48) = 9.03$, $p < .05$]. Auditory shadowing produced reliable decrements in corrected recognition scores for groups receiving auditory words, visual words, and common objects ($p < .01$) but not for the group receiving fictitious characters. Further, the shadowing decrement for the group receiving auditory words was larger than the decrement for all visual input groups ($p < .05$).

DISCUSSION

The present study replicates and extends Cherry (1953) and Moray (1957). Clearly, Ss have considerable difficulty processing and storing one verbal message while attending to a second verbal message presented to the same sensory mode. This result occurs in spite of the fact that Ss were explicitly told that they would be tested for recall of the nonattended message.

On the other hand, Ss are able to process and store a larger portion of the contents of an equivalent visual message under the same conditions. The latter result poses some problems for many models of selective attention and memory. Both Broadbent (1971) and Treisman (1969) assume that a perceptual selection mechanism operates to process fully one of two competing verbal inputs while attenuating the processing of the other. This attenuation model permits processing and therefore some storage of nonattended verbal messages. However, if, as Broadbent (1971) and Treisman (1969) suggest, this filter operates on verbal inputs from all sources and if filtering occurs prior to the processing of the content of inputs, then nonattended visual inputs should be attenuated as much as nonattended auditory inputs. Of course, the present data are not consistent with this line of reasoning.

The attenuation model could be modified to account for at least some of the present data. One could assume either that the various nonattended visual inputs are attenuated less than auditory inputs by auditory shadowing or that visual and auditory feature analyses occur in parallel. In the latter case, one must also assume that nonattended visual words compete with the auditory message for access to an auditorily based verbal memory (cf. Sperling, 1960; Wickelgren, 1966), whereas nonverbal information received visually is stored in a separate visual memory. In this way, auditory shadowing could interfere with the storage of verbal input but not interfere with nonverbal input received visually.

However, there is some evidence (cf. Bower, 1970; Atwood, 1971; Brooks, 1968) that verbal inputs are not stored solely in an auditory memory, but rather that they may be stored visually and auditorily. In fact, a parsimonious explanation of all of the present data may be derived from a model that assumes independent and parallel processing of visual and verbal inputs with modality-specific but interacting memories. Atwood (1971) and Bower (1970) have proposed models having these assumptions. Atwood (1971) suggests that verbal information may be stored in two independent but interacting memory systems. One, the verbal-auditory system, stores an auditory version of verbal information (i.e., a label). The other, the visual system, stores a visual image of the verbal information. Atwood suggests that the visual system stores a visual image of the object named by a word or phrase, although such a system might also store an image of the word itself. According to Atwood (1971), these two systems operate in parallel with little crossmodal interference. However, the two systems may also interact. That is, a word presented auditorily may be transformed for storage in the visual system or a word presented visually may be transformed for storage in the verbal-auditory system. Thus, under normal conditions, a particular word or phrase could be stored in both systems regardless of input modality. However, if the auditory system were blocked, by the shadowing task for example, verbal information might still be stored in the visual system. Recall in the latter case would be reduced, since the information has been stored in one rather than two systems, but still be above chance.

The results of the present study are consistent with this interpretation of Atwood's model. Under the shadow condition, the auditory system is effectively blocked; no information is stored by this system. Further, Ss are unable to transform information presented to the auditory modality for storage in the visual system. Thus, Ss have little memory for words presented auditorily. On the other hand, words presented visually may be stored as visual images (either images of the words or the objects they label) even when S is shadowing an auditory message, and thus Ss have considerable memory for words presented visually. However, their performance is poorer here than under the no-shadow condition, since Ss may store items in both visual and auditory systems in the latter case.

The results of the two groups receiving pictures of objects are more difficult to interpret since, under the no-shadow condition, these two groups differed from each other and from the groups receiving words. However, the effects of auditory shadowing on memory for objects presented visually are clearly consistent with the explanation offered above. Under the no-shadow condition, Ss may store both a visual image of and a label for an object that can be labeled. If auditory shadowing blocks storage of the label, one would predict (and we, in fact, found some, but not total) blocking under the shadow condition. In fact, the performance decrements for groups receiving words visually (mean d′ decrement = 1.00) and objects labeled by words (mean d′ decrement = 1.33) were nearly identical.

One would also predict that objects that are difficult to label are stored only in the visual system under either the shadow or no-shadow condition. If this were the case, then auditory shadowing should not produce a decrement. The results for this group tend to support this prediction. The decrement produced by shadowing was small in absolute value and not reliable even with the very sensitive tests used. Further, 6 of the 16 Ss in this group had higher d′ values under the shadow condition than under the no-shadow condition.

Brooks (1968) and Atwood (1971), among others, report data consistent with this interpretation of the present study. Brooks (1968) has shown that recall of verbal information is readily disrupted by concurrent vocal activity but not by spatial-visual activity, while recall of spatial-visual information is disrupted more by spatial activity than by vocal activity. Further, Atwood (1971) has shown that, when Ss are instructed to encode verbal information as a visual image, the processing of auditory information interferes less with retention than does processing of visual information. However, when instructed to encode information auditorily, auditory processing interferes more than does visual processing.

Taken together, these studies provide convincing evidence that the visual and auditory systems operate independently and in parallel and that the visual system has a separate memory for long-term storage of information.

REFERENCES

Atwood, G. An experimental study of visual imagination and memory. Cognitive Psychology, 1971, 2, 290-299.

Bower, G. H. Analysis of a mnemonic device. American Scientist, 1970, 58, 496-510.

Broadbent, D. E. *Perception and communication.* London: Pergamon Press, 1958.

Broadbent, D. E. *Decision and stress.* New York: Academic Press, 1971.

Brooks, L. R. Spatial and verbal components of the act of recall. Canadian Journal of Psychology, 1968, 22, 349-368.

Cherry, E. C. Some experiments on recognition of speech with one and with two ears. Journal of the Acoustical Society of America, 1953, 25, 975-979.

Glucksburg, S., & Cowan, G. N., Jr. Memory for nonattended auditory material. Cognitive Psychology, 1970, 1, 149-156.

Gourevitch, V., & Galanter, E. A. A significance test for one parameter isosensitivity functions. Psychometrika, 1967, 32, 25-33.

Kroll, N. E. A., Parks, T., Parkinson, S. R., Bieber, S. L., & Johnson, A. L. Short-term memory while shadowing: Recall of visually and of aurally presented letters. Journal of Experimental Psychology, 1970, 85, 220-224.

Marasciulo, L. A. Extensions of the significance test for one-parameter signal detection hypotheses. Psychometrika, 1970, 35, 238-243.

Mobray, G. H. Perception and retention of verbal information presented during auditory shadowing. Journal of the Acoustical Society of America, 1964, 36, 1459-1469.

Moray, N. Attention in dichotic listening: Affective cues and the influence of instructions. Quarterly Journal of Experimental Psychology, 1959, 11, 56-60.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1966.

Norman, D. A. Memory while shadowing. Quarterly Journal of Experimental Psychology, 1969, 21, 85-93.

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

Treisman, A. M. Strategies and models of selective attention. Psychological Review, 1969, 76, 282-299.

Wickelgren, W. A. Short-term recognition memory for single letters and phonemic similarity of retroactive interference. Quarterly Journal of Experimental Psychology, 1966, 18, 55-62.

NOTES

1. Published by Random House, New York.

2. The ns used in these analyses were the total number of Ss per group times the number of TBR or foil items. All d′ analyses have as assumptions that signal and noise distributions are normally distributed with equal variances. Additionally, it is to be noted that these d′ analyses are based upon group data rather than upon individual S data, since several Ss in each group had either perfect hit rates and/or zero false alarm rates. Of course, d′ is not defined for these values.

(Received for publication October 17, 1972; revision received December 19, 1972.)