

Accuracy of recognition memory for common sounds*

DAVID M. LAWRENCE

Claremont Graduate School, Claremont, Calif. 91711

and

WILLIAM P. BANKS†

Pomona College, Claremont, Calif. 91711

Memory for 194 common sounds recorded in high fidelity was tested with a confidence-rated "yes-no" recognition task. Recognition d' was about 1.74 for both individual and group presentation and test. Performance was equivalent to between 85% and 89% correct in two-alternative forced-choice testing, a level considerably greater than the 69% for sounds found by Miller & Tanis (1971). Memory capacity for sounds is probably not exceeded with 194 items.

The memory capacity for complex, meaningful, pictorial stimuli has been found to be very great. Experiments by Haber (1970), Nickerson (1965), Shepard (1967), and Standing, Conezio, & Haber (1970) report well over 90% accuracy in two-alternative forced-choice (2AFC) recognition memory for photographs when very large sets of pictures are held in memory. On the other hand, Miller & Tanis (1971) found 2AFC correct recognition for a memory set of 200 common sounds to be only about 69%.

The Miller and Tanis finding does not, however, imply that the capacity or accuracy of memory for sounds is less than that for pictures. Their equipment was of poor fidelity,¹ and Miller and Tanis suggest that their study was comparable to recognition memory experiments that use blurred pictures. The question of how accurate recognition memory is for sounds recorded in reasonably good fidelity was not answered by their study.

Another problem with the Miller and Tanis study was that they presented their sounds very briefly (0.5-4.8 sec). Such short durations are near the lower limit for efficient storage of pictures (Potter & Levy, 1969), and sounds probably require more time than pictures to form meaningful units. The information in a photograph can be scanned quickly, whereas the information in sounds is spread over time and the listener can only gather information about the sound at the rate at which it comes to him. Miller and Tanis come to a similar conclusion: "It is likely that auditory stimuli must be of rather long duration or extremely familiar before they can be differentiated and stored in memory

*This research was supported by a Pomona College research grant to W.P.B. and a National Science Foundation Grant (No. GU4039) to the Claremont Graduate School.

†Requests for reprints should be sent to William P. Banks, Department of Psychology, Pomona College, Claremont, California 91711.

as well as pictures of familiar objects and scenes [can be] [1971, p. 308]."

While no strong theoretical conclusions about memory have been affected by the seeming evidence that memory for sounds is worse than memory for pictures, it is unfortunate that the only published report on the matter used what must be described as degraded stimuli. The present study intends to remedy this situation. Here recognition memory for sounds recorded in high fidelity and for adequate durations will be tested.

METHOD

Subjects

The Ss were 18 students from California State University, Los Angeles. Ten Ss (Group I) both heard the inspection tape and were tested individually with earphones; they were paid for participating. Eight Ss (Group II) heard the inspection and test tapes as a single group through loudspeakers; they were not paid.

Materials

The inspection tape contained 194 sounds, recorded in stereophonic sound from several sound-effects albums. High-fidelity stereo components, Memorex tape, and high-quality earphones and loudspeakers were used in recording and playback. Fifty sounds from the inspection tape were randomly selected from all but the first and last 20 serial positions and combined with 52 similar new sounds to create the test tape. Great pains were taken to assure that all sounds were recorded at uniform volume and in uniformly high fidelity.

Categories of the sounds on the inspection tape and an example of each category are as follows: *machinery* (29%)—Teletype machines, newsroom; *nonspeech human* (18%)—baby sneezing, laughing, playing; *animals* (13%)—horses neighing in a stall; *miscellaneous* (12%)—water running in a sink; *vehicles* (9%)—jet warmup and taxi; *musical* (9%)—tap-dance routine; *war* (7%)—torpedo tubes being fired; and *weather* (3%)—rain and thunder. The duration of each selection was great enough, as judged by the Es, for the objects or events depicted by the sounds to have some chance of being identified or at least distinguished from other similar sounds on the tape. Thus, noisemaking objects were allowed to go through a wide range of sounds before the selection was terminated, and events that were depicted by a succession of sounds were allowed to run to completion.

Durations averaged 29.5 sec, with a standard deviation (SD) of 14.8 sec. Approximately 5 sec of blank tape intervened between selections.

Procedure

The instructions and tapes for Groups I and II were identical. All Ss were told the purpose of the study, namely, that it was designed to test their ability to recognize common sounds. At the beginning of the inspection tape, the sound was adjusted to a comfortable level, and Ss in Group I had control of the volume throughout the experiment. All Ss were given two 5-min rest periods during the inspection tape and one between the inspection and test tapes. Total experimental time for each S, including the three rest periods, was about 3½ h.

In the recognition test, Ss heard the test tape and wrote yes (old) or no (new) on numbered blanks on a response sheet after hearing each test item. They then rated their confidence in their response on a 1 (least confident) to 10 (most confident) scale.

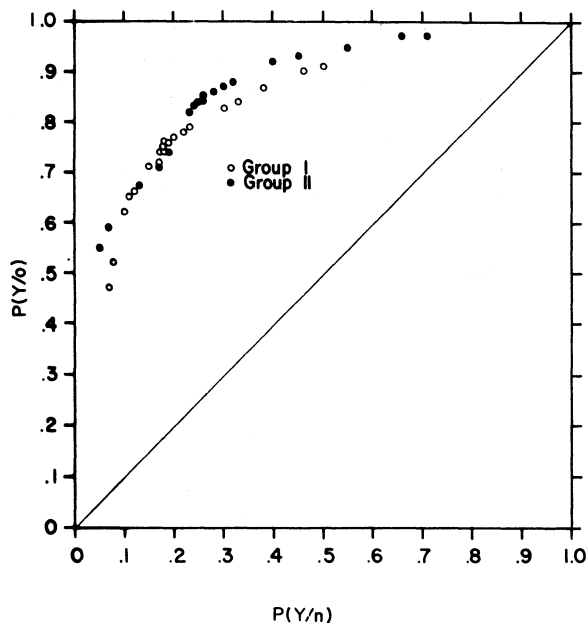


Fig. 1. Memory operating characteristics for the two groups.

RESULTS

Signal detection analysis of the recognition memory data showed virtually identical performance for the two groups. For example, when d' was computed only at the "yes-no" criterion, i.e., with the confidence ratings excluded, then Group I had an average d' of 1.73, with an SD of .35 and Group II had virtually the same mean d' (1.75), with an SD of .33. These d' values indicate that correct recognition in a 2AFC task would occur 89% of the time (Elliott, 1964) for both groups.

Operating Characteristics

Figure 1 shows the memory operating characteristics (MOCs) generated from the confidence ratings (cf. Banks, 1970). The two MOCs are virtually coincident and, thus, any measure of recognition performance would show the two groups to be nearly identical in their performance. The area under the MOC, a nonparametric index of recognition performance and interpretable directly as the proportion Ss would get correct in a 2AFC task that used the same targets and distractors (Green & Swets, 1966), is .845 for Group I and .862 for Group II.

The serial position curve was almost perfectly flat, perhaps because the first and last 20 items were not tested.

Recognition Within the Categories of Sounds

All categories except *weather* were represented on the test tape as both targets and distractors. The following list gives *within-category* d' values (computed at the "yes-no" criterial cutoff) for each of the categories.

These are termed within-category d' s because the hit rates for the items in each category are compared to the false alarms only for the distractors within that category in computing d' . Thus, the d' values reported below reflect the discriminability between old and new vehicle sounds, old and new musical selections, etc. Since all but a few of the new sounds have the same verbal label as one (or more) of the old sounds in the same category, these d' values reflect, to some extent, discrimination between old and new sounds with the same name. In this list the number of old and new items in each category on the test tape is given in parentheses after the category name in the order (old, new): *Nonspeech human* (14,15), $d' = 2.16$; *machinery* (12,10), $d' = 1.53$; *vehicles* (2,7), $d' = 1.67$; *music* (4,3), $d' = 2.27$; *war* (2,2), $d' = 1.02$; *miscellaneous* (8,8), $d' = 1.24$; *animals* (8,5), $d' = 1.30$.

DISCUSSION

Recognition memory for common sounds appears to be extremely good when the sounds are recorded for adequate durations and in high fidelity. This conclusion holds for both Group I and Group II and, given the relatively small variance in d' scores, for most of the individual Ss as well. Converting the d' scores to percent correct in a 2AFC recognition test, we find that accuracy of recognition is considerably better than the 69% Miller and Tanis got for sounds, approximately the same as they got for printed words (85%), but a bit worse than the 90% to 95% generally reported for pictures.

In comparing these recognition scores, it is tempting to draw conclusions about the relative memory capacity of the various modalities. It is essential to note, however, that accuracy of recognition cannot be interpreted directly as an index of memory capacity. As pointed out by Bernbach (1967) and Banks (1970), the detection score for recognition depends upon two distinct factors: (1) strength or availability of the memory trace and (2) the possible discriminability between the new and old test items. The recognition d' is not an absolute measure of memory strength. Because d' is an interval scale quantity whose zero point is the mean of the new item distribution, it varies as much with changes in the similarity between targets and distractors as with changes in the degree to which the targets are in storage.

Because both of these factors determine recognition performance, neither the present results nor those of Miller and Tanis can be interpreted unequivocally in terms of either capacity (of memory storage) or accuracy (discriminability in recognition). However, it can be concluded that Miller and Tanis did not obtain the maximum recognition accuracy possible when 200 common sounds are held in storage, and they probably underestimated the memory capacity for sounds.

REFERENCES

- Banks, W. P. Signal detection theory and human memory. *Psychological Bulletin*, 1970, 74, 81-99.
- Bernbach, H. A. Decision processes in memory. *Psychological Review*, 1967, 74, 462-480.
- Elliott, P. B. Tables of d' . In J. A. Swets (Ed.), *Signal detection and recognition by human observers*. New York: Wiley, 1964.
- Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.
- Haber, R. N. How we remember what we see. *Scientific American*, 1970, 222, 104-112.
- Miller, J. D., & Tanis, D. C. Recognition memory for common sounds. *Psychonomic Science*, 1971, 23, 307-308.
- Nickerson, R. S. Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology*, 1965, 19, 155-160.
- Potter, M., & Levy, E. Recognition memory for a rapid sequence

of pictures. *Journal of Experimental Psychology*, 1969, 81, 10-15.

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

Standing, L., Conezio, J., & Haber, R. N. Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 1970, 19, 73-74.

some experiments on memory for sounds using a more recent and, we are told, improved model of the Language Master device that Miller and Tanis used. The fidelity of this device is not simply poor, it is abominable. Wow and flutter is as bad as 10%, S/N ratio is very poor, dynamic range is extremely limited, and frequency response is approximately 100-5000 Hz. A few informal experiments revealed that the sex, but not the identity, of a speaker could be reliably identified through this device, and common environmental sounds with sharp transients (automobile sounds, playing cards being shuffled, breaking glass, etc.) were muffled and extremely difficult to identify.

NOTE

1. The authors began this study after attempting to perform

(Received for publication January 26, 1973.)