Journal of Experimental Psychology
VOL 47, No. 3, 1954

'THE ROLE OF AUDITORY LOCALIZATION IN ATTENTION AND MEMORY SPAN[1]

D. E. BROADBENT

Applied Psychology Research Unit, Cambridge, England

It is known from the work of Hirsh (5) and Kock (6) that speech is more readily understood in noise when the localization of the speech and noise sources is different. These earlier results, however, do not include the case in which both sources are emitting speech, and the listener does not know in advance which source will be relevant. The importance of this selective function has already been demonstrated (1). Poulton (8) has reported some data on the effect of localization on detection of signals, and the present paper is intended to extend this work. The experiments fulfill two main purposes: Exp. I to examine the relative efficiency of a number of different types of localization, and Exp. II to examine the limits of immediate memory on two separated channels.

EXPERIMENT I

Procedure.-The task was that used in the earlier report (1). The Ss were faced with a visual display and were required to answer such questions as "S-l from GDO. Is there a heart on position Two? Over." The answer was to be Yes Or No, and was preceded by the call-sign of the person asking the question, thus: "GDO from S-1 yes. Over." Unless both the call-sign and the answer itself were correct, the response was scored as an error. There were many voices asking the questions, each of which kept the same call-sign throughout. From time to time both voices would start questions simultaneously; when this occurred only one of them would begin his message with S-l. The S was warned that when this happened he was to ignore the message which did not begin with S-l, and to answer the other. A run on this task consisted of 19 plain messages each on one channel only, interspersed with 12 occasions on which both voices started simultaneously. There were 4 sec. between each message, and the whole run lasted between 3 and 4 min.

The questions were recorded on tape and played back on a two-channel Ferrograph. This was similar to the Model C in performance, but carried two separate tracks on the same tape. Where headphones are mentioned below, S.G. Brown Type D.1 were employed: the loudspeakers were a Grampian G. 129 (the "movable" speaker) and that built into the Ferrograph (the "fixed" speaker). Where only one loud-speaker is mentioned, half the Ss used the movable and half the fixed. The level of playback was 75 db in Groups 1, 2, and 3, and 70 db in Groups 4, 5, and 6 (in which a headset was sometimes used).

Subjects.—"The 76 Ss were all Royal Navy enlisted men under the age of 31. They were divided into six distinct groups, each containing 12 Ss except for Group 1, which had 16. All groups received a training session on the day before the experiment proper, but this training was only in answering straightforward messages and there were no presentations of two messages simultaneously. Training was given with the type of equipment used in the main experiment: for groups which used two types of equipment half the Ss were trained with each type. On the day of the main experiment, each group received two different conditions of presentation. One run on the task, as described above, was given under each condition, and there were therefore 12 observations on simultaneous messages from each S on each of his two conditions. The order of presentation of conditions was rotated within each group.

Conditions.—we will term *separate* recordings those in which each voice was recorded by itself on one of the two tracks of the recorder. Other *stereophonic* recordings were made from the separate recordings by playing back the latter simultaneously over the two loud-speakers of a similar two-channel recorder and picking up the resulting sounds with two microphones. Thus both voices were recorded on both tracks. There were two subvarieties of the stereophonic recordings, *fused* and *localized*; in the former the recording microphones were each just under 3 ft. from each of the speakers so that the recording of a particular voice was simultaneous on the two tracks. In the latter, one microphone was 1 ft. from one speaker and 3 ft. from the other, while

[1] This work was supported financially by the British Medical Research Council, and thanks are due to the Royal Navy for supplying Ss and mechanical equipment.

the second microphone reversed this relation. Thus the recording of one voice was about 2 msec, earlier on Track 1 than on Track 2 while the second voice would be earlier on Track 2 than on Track 1 by the same amount. From the data of Wallach, Newman, and Rosenzweig (12) we would expect the localized recordings, when suitably presented, to produce a perception of two separately localized voices, while the fused recordings would give the same localization for both voices.

Group 1 heard separate recordings played through two loud-speakers. The fixed speaker was kept 3 ft. in front of S at shoulder height and facing him, while the movable speaker was at the same distance from S and 1.5 ft. higher from the ground. In one run the movable speaker was directly above the fixed one, while in the other it was on S's right-hand side, i.e., separated from the fixed one by 90° in azimuth. The S's head was free.

Group 2 heard stereophonic recordings through two loud-speakers, in the 90° separated positions. One run was fused and one localized.

Group 3 heard one track only of the stereo phonic recordings, one run being a fused recording and one localized. The track was played through a loud-speaker, half the Ss hearing Track 1 through the fixed speaker and half Track 2 through the movable speaker in the90° position.

Group 4 heard stereophonic recordings, one run fused and one localized, each track being led to one ear of a pair of headphones.

Group 5 heard separate recordings in one run through a divided headset with each track to a different ear. In the other run two loud-speakers were used, mounted facing each other level with S's ears on each side of him and with 2 tracks between the speakers.

Group 6 heard separate recordings in one run through the same two-loud-speaker arrange-

TABLE 1

Conditions Compared in Listening to Meaningful Messages

(Experiment I: Only messages for S-l to be answered}

Group	N	Conditions: General	Conditions: 1	Conditions: 2	% Errors: 1	% Errors: 2	t*	p
1	16	"Separate" recordings coming *through two loud-speakers*	Speakers one above the other	Speakers 90° apart	30	20	3.30	.01
2	12	"*Stereophonic*" recording coming through two *speakers 90° apart*	"*Fused*" *recording*	"Localized" recording	58	31	4.49	.01
3	12	As above, only one speaker working (*a*) Voice leading on "localized" *recording*	"Fused" recording	"Localized" recording	49	36	<1	Insig.
		(*b*) Voice later on "localized" recording	"Fused" recording	"Localized" recording	55	54	<1	Insig.
4	12	"Stereophonic" *recordings*, each track coming through one ear of a headset	"Fused" recording	"Localized" record*ing*	57	40	3.09	.05
5	12	"Separate" record*ings*	Each track coming through one ear of a headset	Each track through a speaker opposite one ear	28	16	2.71	.05
6	12	As above	Each track coming through a speaker opposite one ear	One track through an earphone, other through a *speaker opposite* one ear	19	17	<1	Insig.

*The data required an angular transformation to give normality, and the values of t are for the differences between the means of the transformed data.

ment as Group 5. In the other run one speaker was replaced by a single phone mounted alone on the headset and worn on the ear opposite to the remaining speaker.

Analysis of data--The technique of observing rely two conditions on each group of Ss, rather that of giving all conditions to all Ss by Latin-square or factorial design, was adopted because interactions of condition and order are expected to be important in this type of task. Consequently the comparisons intended by the design are all between paired conditions, both members of the pair having been given to the same group of Ss.

For each comparison the proportion of errors was determined for each S in each of his two conditions, corrected for continuity, and transformed to ø (= $\sin^{-1}\sqrt{x}$), this being the appropriate transformation for the reasons given by Fisher and Yates (4). Analysis of variance was then carried out, that due to Ss and to order of presentation extracted, and the variance between conditions compared with the remainder. (Variance due to the use of different equipment in training or, in Groups 3 and 6, to using the particular loud-speaker has of course been concluded by the design with that due to Ss, and extracted with it.) The final comparison was this always equivalent to a *t* test with *df* equal to two less than the number of Ss.

The transformed means would be somewhat unintelligible in tabular form, as they are naturally expressed in degrees or radians and have little psychological meaning. For each condition, therefore, the total number of errors is shown expressed as a percentage of the total possible. This should make interpretation easier, but from a logical point of view the t values are the important ones.

Results.—Group 1 shows that the results of Hirsh (5) and Kock (6) apply to the present case as well as to the understanding of speech in noise; i.e., a relevant message presented simultaneously with an irrelevant one is more likely to be understood when the two sources are separated in space.

Group 2 shows that the same advantage applies with apparent differences in localization produced by a stereophonic technique.

Group 3 shows no difference between conditions when listening to only one of the stereophonic channels, which serve as a control for Group 2. It implies that the two types of stereo-phonic recording did in fact give nearly equal intensities for each voice on both channels.

Groups 2 and 3 together act as a control for Group 1. As in Group 2, the intensity relations between the voices did not alter by an important amount any frequency distortions introduced by room reverberation and should have been the same for both conditions. Therefore the results from Group 1 cannot be explained by changes in such distortions.

Group 4 shows that differences in localization are still useful when produced by stereophonic means through headphones.

Group 5 shows that with this equipment two separate tracks are more easily understood through two loud- speakers than through two separate headphones. This result is, other things being equal, to be expected from Hirsh's (5) data on restricting head movement.

Group 6 showed no difference between a two-loud-speaker system and one using one speaker and a headphone. Together with the results from Group 5 this means that a mixture of two systems of presentation may sometimes be the equal of the better of the two systems.

Experiment II

Throughout Exp.I the percentage of errors was less than 50% in all conditions allowing some spatial separation (i.e., all the two-channel conditions except the fused stereophonic recordings). This contrasts with earlier findings using two voices from one loud-speaker (1) and implies that spatial separation not only favors the understanding of messages once they receive attention, but also that the correct message \s mare likely to secure attention.

A preliminary experiment was carried out using only call-signs alone, and requiring S to pick out a sign from other irrelevant ones without knowing which channel it would occupy. At a slow presentation rate differences in localization were useful and at a fast rate they were harmful. This suggested the possibility that listeners can hear two spatially separated signals successively rather than simultaneously, and the following experiment on immediate memory was designed to test this hypothesis.

Procedure.—The two-channel Ferrograph was employed, using the split headset with each track delivered to a separate ear. Recordings were played of lists of digits under various conditions described below, and S was instructed to listen to each list completely through and then write down as much of it as he could remember.

Subjects.—The 24 Ss were again Royal Navy enlisted men under the age of 31; 12 of them were assigned to Group 1, and the remainder served both as Group 2 and as Group 3 since the comparisons involved were independent. Sir Ss performed the Group 2 conditions before the Group 3 ones, and six reversed this order. No Ss had served in Exp. I.

Conditions—We shall term *conventional* lists those in which a single series of digits was heard identically by both ears, so that S heard the series in the normal manner used for determining memory span; e.g., Track 1 = Track 2 = 734215. In *binaural* lists different material was recorded on the two tracks, so that one digit arrived at one of S's ears and another digit simultaneously at the other ear, then another pair, one to each ear, and so on; e.g., Track **1** — 734, Track 2 — 215.

Group 1 heard five conventional lists, each of six digits. A warning was then given, and ten binaural lists given, each consisting of three pairs or six digits in all. A final five conventional lists followed and the whole procedure was then repeated using lists of eight digits instead of six. A 1/2sec. interval was used between digits or pairs, and the instructions (or the binaural lists were to record the digits in whatever order s chose.

Group 2 heard 24 binaural lists, each of three pairs of digits, and Ss were instructed to record the digits in the actual order of arrival; e.g., Track 1 = 734, Track 2 = 215, Response = 723145. The E marked as an error any response in which a digit appeared earlier than one which had arrived before it, but the order of writing each pair was immaterial. The rate of presentation was varied, the first three lists having 2-sec. interval between pairs, the next three 1 1/2 sec, the next 1 sec, and the next 1/2 sec. Successive groups of three lists then became slower over the same range.

Group 3 heard the eight-digit lists of Group I, and in addition heard a series of ten binaural lists, each of eight pairs. In this condition the instructions were to write down the first four digits heard on one ear and the last four heard on the other. For half the lists the first digits to be written were on the right, and for the other half, were on the left. A 1/2 sec. interval was used throughout. This condition is labeled "control for binaural" in Table 2.

Analysis of data—As the differences obtained were large, the number of Ss showing each effect was determined and *P* calculated from the binomial distribution. For the present case the significances are high even by this method, which does not involve unnecessary assumptions. In Group 2 a t test was also applied to the differences between a 1-se. rate and a 1 1/2-sec. rate, using the distribution of individual S differences. This was because five Ss showed zero differences and so made the binomial test doubtful, but the t test is of course almost equally suspect.

Results.-—Group 1 shows that almost all correct responses to binaural limits were written down in such an order that all digits presented to one ear were written before any on the other ear; e.g., if Track 1=723, Track2 = 215, the response is 72321S or

TABLE 2

CORRECTLY RECALLED LISTS OF DIGITS (EXP. II)

Group	Conditions			% Correct		
	1	2	3	1	2	3
1 (6-digit lists)	Conventional	Binaural: correctly recorded with all digits on one ear before any on the other	Binaural: correctly recorded in any other order	93	62	5
1 (8-digit lists)	Conventional	As above	As above	56	13	2
	Conventional	Binaural	Control for binaural	42	5	3

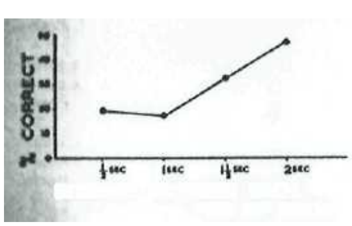
* For Group I, both lengths of list pooled, p<.001 for the difference between Cond. 1-2 and 2-3. All Ss showed these differences. For Group 3, p<.001 for 1-2 and 1-3.

15723. This confirms the hypothesis that when two different simultaneous sounds arrive at the two ears, the sound on one ear produces a response before that on the other ear. In addition, this type of task appears distinctly more difficult than a conventional memory span.

Group 2 shows that failures to record digits in the actual order of arrival are no less frequent at a 1-sec. rate than at a 1/2sec. rate, but 11 of the 12 Ss show a considerable improvement (Fig. 1) between a 1-sec. and a 2-sec. ($p<.01$). The difference between 1- and 1 1/2-sec. rates is due to only seven *Ss* ($p<.01$ by a *t* test). This seems to imply that when attention is shifted away from one channel to another and then back to the first a time interval of between 1 and 2 sec. will be required.

In addition, the low level of performance at the two fast rates shows that the results of Group 1 are not due self-instruction by *S.*

Group 3 shows that the binaural and the "control for binaural" presentations give equal increases in difficulty compared with conventional presentation. These two conditions are alike in that in both of them information to be memorized is accompanied by other sounds on the other ear, which may be "distracting"; and also in that a change of channel must be made halfway through the list, which may act as an additional item.



INTERVAL BETWEEN EACH PAIR OF DIGITS

Fig. 1. Recall of binaural spans in the actual order of arrival, as a function of presentation rate (Group 2)

But the two conditions differ in that the binaural one requires the reproduction of information which arrives while attention is directed elsewhere, while the "control for binaural" does not. This difference, which may be regarded as a case of "incidental learning," need not therefore be supposed to produce any extra difficulty.

DISCUSSION

The practical implications of these results may be summarized as follows: some degree of spatial separation is probably desirable in most cases where an individual must deal with two or more channels. But if rapid alternation between two channels is required, separation may cease to be an advantage. This in fact agrees with the results of Poulton (8) on a synthetic representation of the work of control-tower monitors. It is worth noting that the time interval required for double shifts between channels is of the order given by introspective techniques by Pillsbury (7) for double shifts of attention subjectively defined. Other relevant times in the classical work are given by the same author. The interval is also of the same order as the failures which appear in skilled tasks after prolonged performance, failures which have been ascribed on other grounds to shifts of perceptual selection away from the task and back (3). Such time intervals are important for general psychological theory (2).

In addition, the fact that information may be stored temporarily and only later give rise to selective response (attention) is of some interest. It was once well known under the name of "prior entry" (11, 13), and it allows us to harmonize the results on spatially separated speech channels with the view that the perceptual mechanism has a limited capacity (1). But to this view we must add that excess information may be stored. The results are also relevant to the controversy over continuity of learning, particularly the finding that in conditions of equal perceptual difficulty the same number of items can be recalled whether presented successively or simultaneously. It is difficult in this experiment to see any distinction between memory span and span of apprehension.

These data also place some limits on physiological theories of auditory localization. For instance, Kock (6) stated his findings on the understanding of speech in noise in terms of the insertion by the brain of a time difference between the signals arriving via the two ears. It must be noted that the findings of Exp. II and of the preliminary experiments will support this view only if we suppose that the brain can subsequently remove this time difference and insert a different one, not between future signals, but between past ones. This seems somewhat unlikely, and no doubt Kock intended his phraseology only to emphasize the importance of time as opposed to intensity as a cue, which these results support. If a physiological model is desirable, it should rather be sought in the work of Rosenzweig and Rosenblith (10) and Rosenzweig (9), in which it is shown that time differences at the ears are recorded into spatial position on the cortex. It seems reasonable that information at different cortical points should pass successively through some further mechanism of limited capacity, and that changes in the order of succession or complete neglect of one point should be common.

SUMMARY

In Exp.I Ss were required to answer messages about a visual display, and occasionally two messages were presented together. When this occurred, only one message, that starting with a particular call-sign, was to be answered and S knew that such relevant messages might come from either source. It was confirmed that spatially separated sources led to more correct answers under these conditions, and also that apparent separation produced by stereophonic techniques produced the same result. Various other methods of presentation were also compared.

In Exp. II memory span was determined for lists of digits arriving simultaneously, half on one ear and half on the other. All digits given to one ear were written down by S before any on the other; this confirmed a suspicion that spatially separated sounds may pass through the perceptual mechanism successively rather than simultaneously. As the rate of presentation was slowed up, however, it began to be possible to recall lists in the actual order of arrival. Although binaural memory span is more difficult than a conventionalone, a control experiment showed that this is due to perceptual confusions rather than to an independent failure of memory.

These results suggest that spatial separations will normally be useful unless alternation between two channels is required at a faster rate than about 1 or 2 sec; this time interval has been found important in other connections, such as the study of "fatigue."

REFERENCES

1. Broadbent, D. E. Listening to one of two Synchronous messages. *J exp.Psychol*, 1952, 44, 51-55.
2. Broadbent, D. E. Classical conditioning and human watch keeping. *Psychol*., 1953, 60,331-339.
3. Broadbent, D. E. Noise, paced performance, and vigilance tasks. *Brit J, Psychol*, 1953, 44, 295-303.
4. Fisher, R. A., & Yates, F. Statistical tables. London: Oliver & Boyd, 1941.
5. Hirsh, I. J. The relation between localization and intelligibility. *J. acoust Sac , Amer*., 1950,22, 196-200.
6. Kock,W. E. Binaural localization and masking. *J. acoust. Soc. Amer*., 1950, 22,801-804.
7. Pillsbury,W.B."Fluctuations of attention" and the Refractory Period, *J Philos. Psychol. sci, Meth*., 1913, 11, 181-185.
8. Poulton, E. C. Two-channel Listening, *J. exp. Psychol.*, 1953, 46, 91-96.
9. Rosenzweig, M. R. Representations of the two ears at the auditory cortex, *Amer.,J. Physiol*., 1951, 167,147-158.
10. Rosenzweig, M. R., Rosenblith ,W. A. Some electrophysiological correlates of the perception of successive clicks. *J. Acoust.Soc. Amer*. 1950, 22, 878-880.
11. Stone, S. A. Prior entry in the Auditory-tactual complication. *Amer. J. Psychol,* 1926, 37, 284-291.
12. Wallach, H., Newman, E. B., Rosenzweig, M. R. The Precedence Effect in Sound localization. *Amer. J. Psychol* 1949, 62, 315-336.
13. Welford, N, T. Estimation of the Position of a Tactile Stimulus in a Repeated Auditory Pattern. *Quart. J. exp. Plychol,* 1949, 1, 180-192.

(Received May 18, 1953)