Monitoring simultaneous auditory messages

L. H. SHAFFER[1] AND JANE HARDWICK[2]
UNIVERSITY OF EXETER

If a S is asked to monitor two simultaneous auditory speech messages and to report only on the occurrence of target words appearing at random in either message, then it is shown that he will fail to detect all of them but will detect significantly more than half. The targets used in these experiments were immediate repeats of text words. The results reject theories that part of the sensory input is blocked or that all is recognized. Detection performance was a function of rate of speech and of intertarget interval; there was a small, not significant, effect of instruction to recognize message content.

If two verbal messages are presented simultaneously to a S, one to each ear over headphones, then he can, with minor exceptions, report on only one of them. The exceptions occur if the messages are brief (Broadbent, 1958) or if they are correlated up to a temporal displacement (Cherry, 1953; but see Treisman, 1964). He may recognize his name in the alternate message (Moray, 1959), or that the contents of the messages have abruptly switched (Treisman, 1960). Although he can say little about the content of the alternate message, he can report on some of its physical properties (Cherry, 1953). Why is so little of the alternate message recoverable?

Note that this question misses a possibly more interesting one, which is, how much information a S could extract from two or more messages under optimal conditions. The experimental technique used in most of the above studies requires him to repeat aloud (shadow) one of the messages and attempt to remember the other. It is not an optimal condition and cannot be commended as a natural mode of communication. Loss of the second message may occur as a failure of memory, or because it is dominated both by the primary message and by the S's own voice repeating it, or because he is having to speak while listening.

In order to overcome the memory factor, Treisman and Geffen (1967) introduced into the shadowing task verbal signals at random in either message that required a tapping response as soon as they occurred. About 86% of signals on the primary ear and 8% on the alternate ear were detected.

We have simplified the procedure by eliminating the shadowing task and retaining the task of monitoring verbal signals. Neither message is now primary; the S must monitor both messages continuously in order to detect all signals. His response load is reduced to pressing a button whenever he hears a signal and, in case it helps, he is given a different button for each message. The signal used is the immediate recurrence of a word within a message.

METHOD

Apparatus

The messages were recorded with a Grampian DP4/N microphone onto a Ferrograph stereophonic tape recorder at a tape speed of 7½ in/sec and were played back over Amplivox headphones.

In each pair of messages, one was spoken by a male and the other by a female voice, and the voices were presented at different phones. The messages were prose extracts, one from an autobiography, the other from a travel article, and the signals were produced by the speaker repeating immediately certain words. It was necessary for the speakers to have extended practice speaking monotonously with the minimum of pause and not giving articulatory prominence to the repeat words.

Factors like signal distribution and length of signal words had been manipulated in several pilot experiments. It was found that (a) detection probability increased, from about 70% to 85%, with length of signal word and (b) varying signal density over a wide range had little effect on detection performance.

For the main experiment, two sets of messages were recorded, differing only in speed of reading, the speeds being 145 and 175 words per minute. The text words repeated as signals were chosen at random with a probability of 0.1, with the constraint that they should be only one or two syllables long, and, in fact, there were 125 signals in each message of about 1300 words.

The S had to press one or the other of two Morse keys that activated a pen recorder.

Design

There were 12 conditions; different Ss were tested in each, and a S was tested once only. In half the conditions, the recordings were played normally and in the other half they were played backwards. The speech could be fast (F) or slow (S).

The S might be required to monitor both messages (Monitor 2), or hear both messages but monitor only one of them (Monitor 1), or hear and monitor only one message (Hear 1). The choice, or combination, of message and ear was randomized across Ss within a condition.

The S might be instructed that apart from detecting word repetitions he would be later questioned (Q) on the contents of the messages (questions were in fact asked at the end) or left unquestioned (U) and therefore did not need to monitor content.

The complete set of conditions is shown in Table 1.

Subjects

The Ss were 92 students in the University.

Instruction

The S was told to report the occurrence of a repeat word as soon as possible by pressing the left or right key to indicate left or right message. He was given some practice on a tape recording similar (in speed and signal density) to the one he was tested on,

Table 1
Mean and Standard Deviation (Bracketed) Probabilities of Detection, d, and of false positives, fp

		Monitor 1		Monitor 2		
Forward Speech						
SU	d	0.948(0.030)	SU	0.684(0.062)	SQ	0.644(0.072)
	fp	0.0006(0.0006)		0.0007(0.0006)		0.0010(0.0010)
		n = 6		n = 10		n = 10
FU	d	0.956(0.020)	FU	0.592(0.054)	FQ	0.570(0.063)
	fp	0.0011(0.0010)		0.0008(0.0007)		0.0015(0.0008)
		n = 6		n = 10		n = 10
Backward Speech		Hear 1		Monitor 1		Monitor 2
S	d	0.651(0.124)		0.594(0.159)		0.345(0.092)
	fp	0.0061(0.0027)		0.0101(0.0069)		0.0100(0.0042)
		n = 4		n = 10		n = 10
F	d	0.532(0.059)		0.489(0.113)		0.256(0.059)
	fp	0.0096(0.0075)		0.0123(0.0048)		0.0169(0.0051)
		n = 4		n = 6		n = 6

but if he heard the test passages backwards, he heard it played forward in the practice period and was told that the test recording would be backwards. The listening conditions, Q and U, were specified as above.

RESULTS

The basic scores for a S were the numbers of signal detections and false positives (detection responses in the absence of signals). It was necessary to choose a criterion response latency in order to decide, for a given response, whether it was a detection or a false positive. Analysis of the printouts was done by hand and given a preliminary sampling of a S's response latencies; a cutoff point was selected. In fact, there was seldom ambiguity because response latencies within a S's data tended to have a narrow distribution.

The basic scores were converted to probability scores and their means and SDs are shown in Table 1. The probability of a false positive was estimated as the ratio of number of false positives to the total number of nonsignal words. This is reasonable with normal speech but with backward speech, in which the phonetic sound pattern does not have natural word boundaries, it is defensible only as being the least arbitrary procedure. This must be borne in mind in assessing the results.

Table 1 shows that detection scores were higher on forward than on backward speech, on monitoring one message rather than two, on slow rather than fast speech, and, if the instruction was to monitor signals only, rather than signals and message content. All these differences, except the last, were statistically significant, using Mann-Whitney U tests and a fixed significance level of $p < 0.05$.

There was generally an inverse relationship, among the conditions, between detections and false positives so that a consistent inference about signal detectability can be based upon detection scores alone. There were few false positives on forward speech, but a considerable number on backward speech.

Listening conditions were not uniform in time since word-and-repeat might overlap a pause in the alternate message. For each word-and-repeat, the tape record was drawn past the soundhead of the recorder several times to decide whether none, or part of, or all of the word pair was accompanied by silence in the alternate message. If we categorize the overlap of word-and-repeat with a pause as None, Partial, or Full, then we obtain the detection probabilities in Table 2, based upon data from the four conditions of dichotic monitoring of normal speech.

It is clear that signals are more detectable if they and/or the words they repeat occur during pauses in the alternate message, also that the detection probabilities in Table 1 would nevertheless be little modified by excluding these cases.

One can also examine sequential properties of detection. A given signal may be detected, d, or not detected, nd; the previous signal may have been detected or not detected and may have occurred in the same, s, or alternate, a, message. We can distinguish the events: d | d = a detection following a detection, and nd | nd = a nondetection following a nondetection; d | d,s (or d | d,a) = a detection following a detection in the same (alternate) message, and nd | nd,s (or nd | nd,a) = a nondetection following a nondetection in the same (alternate) message.

Table 2
Detection Probability as a Function of the Overlap with Pauses in the Alternate Message

	Slow	Fast
None	P = 0.621 (n = 179)	P = 0.567 (n = 222)
Partial	0.728 (n = 38)	0.654 (n = 14)
Full	0.803 (n = 33)	0.739 (n = 14)

n = number of signals in each category

Table 3
The Conditional Probabilities of Detection and Non-Detection for Signal Pairs at Different Intersignal Intervals

		P(d|d,s)	P(d|d,a)	P(d|d)	P(nd|nd,s)	P(nd|nd.a)	P(nd|nd)
	0-1 Sec	0.69	0.38	0.54	0.21	0.17	0.19
	1.1-2	0.70	0.70	0.70	0.23	0.31	0.27
SU	2.1-3	0.65	0.75	0.70	0.30	0.25	0.28
	over 3	0.74	0.75	0.75	0.27	0.34	0.30
			P(d) = 0.68			P(nd) = 0.32	
	0-1	0.74	0.29	0.51	0.52	0.17	0.35
SQ	1.1-2	0.67	0.66	0.66	0.19	0.28	0.24
	2.1-3	0.64	0.64	0.64	0.31	0.44	0.37
	over 3	0.69	0.79	0.74	0.32	0.24	0.28
			P(d) = 0.64			P(nd) = 0.36	
	0-1	0.68	0.29	0.49	0.42	0.24	0.33
FU	1.1-2	0.59	0.63	0.61	0.40	0.38	0.39
	2.1-3	0.60	0.61	0.60	0.37	0.29	0.33
	over 3	0.66	0.70	0.68	0.37	0.34	0.36
			P(d) = 0.59)			P(nd) = 0.41	
	0-1	0.70	0.26	0.48	0.57	0.23	0.40
FQ	1.1-2	0.64	0.55	0.59	0.41	0.33	0.37
	2.1-3	0.64	0.53	0.58	0.42	0.40	0.41
	over 3	0.59	0.65	0.62	0.48	0.32	0.40
			P(d) = 0.57			P(nd) = 0.43	

The conditional probabilities of these events were estimated and the unconditional probabilities, P(d) and P(nd) = 1 − P(d) were available from Table 1. They are shown in Table 3 for the conditions of dichotic monitoring of normal speech. This table includes a breakdown of results into four categories of intersignal interval.

Note that if successive events were independent then all conditional probabilities should be close to the unconditional probabilities, P(d) and P(nd).

The major results are the following:

(1) At the shortest intersignal intervals, P(d | d) falls close to P = 0.5. At longer intervals, it is nearly always greater than P(d), $\chi_1^2 = 24.9$, $p < 0.05$, and tends to increase with intersignal interval, $\chi_2^2 = 6.7$, $p < 0.05$.

(2) Decomposing P(d | d) at the shortest intersignal intervals shows that P(d | d,a) is very low, whereas P(d | d,s) is larger than at longer intersignal intervals, except in Condition SU. Correspondingly, one finds a low value of P(nd | nd,a) and a high value of P(nd | nd,s). These results are not tested statistically, but they indicate that at any moment in time detection is dominant in one message and that dominance alternates typically more slowly than once per second.

Other statistical analyses found that there were no consistent ear or message preferences. We also examined whether or not there might be a classification among signal words that correlated with detection rate, and found none.

DISCUSSION

The results show that under conditions approaching optimal, the S can recognize only 59%-68% of the two messages, depending upon speech rate. It is, however, more than has previously been demonstrated and is certainly more than the 50% predicted on the assumption that he can listen to only one message at a time. We assume that the per cent signals detected is a measure of the amount of the message words recognized and shall now examine this assumption.

In principle, any discriminative feature of a sensory input may be used to define a source of information. In practice, ear of arrival, spatial location, voice pitch, and intensity provide the most useful bases for selecting an auditory message, but sequential syntactic and semantic features can also be used with some success (Treisman, 1964). In the monitoring task, the words that repeat themselves are not predictable within the messages and they have no discriminative acoustic properties.

If signals were detected as the recurrence of an acoustic pattern, then performance should have been the same in forward and backward speech. That detection is possible in the latter condition shows that acoustic pattern can provide a basis for detection, but the level of performance shows that this is not adequate to account for the result in forward speech. Perhaps words were recognized in the momentarily dominant message and acoustic pattern in the other. This should be testable by averaging detection scores in the Monitor 1 conditions of forward and backward speech (see Table 1). Unfortunately, we cannot do this because of the high rate of false positives in backward speech: we do not know how to estimate this false positive rate properly, nor do we know the proper tradeoff function relating P(d) and P(fp).

To help settle the issue, we have carried out another experiment that replicated the conditions of the first except that the signals were Christian names rather than word repeats. The names occurred at random, mainly out of context of the prose, in either message. The results were essentially similar to those presented here, yet there was little chance of recognizing signals at the acoustic level; we suppose, therefore, that recognition occurs at the word level.

Turning to theories of attention, the results reject the hypothesis that a filter of the sensory input excludes all but one message at a time at the level of cognition (Broadbent, 1958). On a simple interpretation, the tactic of alternating between messages should guarantee more than 50% detection.

The position is not affected by the ancillary postulate, in Broadbent's theory, of a sensory store. If the S can select information from an input or from a store, then, since these are to be regarded as alternative sources, he must surrender the inputs while retrieving from a store.

Since Ss made false positive responses, their detection scores may be inflated by guessing. If, by hypothesis, the information from a channel is or is not sampled, we can estimate the contribution of guessing as:

$$\text{P(correct guess)} = \frac{\text{P(signal)}}{1 - \text{P(signal)}} \cdot \text{P(false positive)}$$

The probability of a signal was 1/10, so that the probability of a correct guess was 1/9 of the false positive rate, which, as Table 1 shows, is negligibly small.

Is it possible that the S, by alternating attention rapidly between ears, can sample each message and recover sufficient acoustic information to recognize words and their repetition? Miller and Licklider (1950) have shown how level of word recognition in a single message varies with rate of sampling of speech and the fractional size of the sample. One can examine, therefore, whether or not the signal detection rates obtained here could be achieved, for some rate of external switching, if the stimulus input were to alternate from one message (and ear) to the other. It follows from Miller and Licklider's results that at very high switching rates the results would be replicated and the S would not even notice that switching had occurred, but a hypothesis that attention can switch at such a rate would be trivial. Preliminary results show that for switching rates between 1 and 30 cps, detection rate does not reach even 50%.

We can also reject the hypothesis that all message inputs are recognized, but that only some can be held in memory or translated into response (Deutsch & Deutsch, 1963). Such a hypothesis predicts the relatively poor detection performance at short intersignal intervals, but it also predicts that detection probability should approach 0.95 at longer intersignal intervals. It may be argued, contrary to our expectation, that giving the S two response buttons in Monitor 2 conditions impairs performance by introducing a response choice. This may be a factor to take into account at short intersignal intervals but not elsewhere.

In order to obtain a critical test of the hypotheses, we have had to ignore possible theoretical niceties. If one postulates different levels of analysis between the acoustic input and full word recognition then it is possible to introduce selection or filtering at any of these intermediate levels so that the distinctions between hypotheses disappear. It is thus necessary for the theorist to justify such niceties.

None of the other hypotheses of attention currently available is tested by the present task because they make only the qualitative prediction that some recovery of the alternate message is possible.

Treisman (1964) modified the filter hypothesis by supposing that the alternate message was degraded rather than excluded. The results cannot reject this assumption, but we may question whether it is necessary to postulate any process that filters or degrades the input; a simpler assumption is that the process of categorizing the acoustic input into linguistic forms cannot exceed a certain rate and that this rate is usually exceeded with dichotic presentation. Any part of the acoustic input that is not categorized within a short time of its arrival is displaced by its sequel. In Jamesian language, it is heard but not listened to. Thus, in case of information overload, the selective process, which we call attention, allocates priority of recognition among the sources of input. A similar hypothesis was put forward by Neisser (1967) who, however, identified the rate limiting factor as an analysis-by-synthesis process of speech recognition (Halle & Stevens, 1964). Since there are no convincing reasons to prefer one theory of recognition to another, one may settle for the weaker, but sufficient, assumption of rate limitation.

It follows from this hypothesis that fewer signals should be detected at fast then at slow rates of speech, as shown here. We had also expected to show that if the S were required to recognize the input in units larger than the word, then this too should impair detection performance. Failure to find significant differences between U and Q conditions may have been due to inadequacy in the instruction procedure, and, since there were trends in the expected direction, this part of the experiment may be worth repeating. It may, however, be naive to suppose that word recognition is simpler than phrase or sentence recognition (cf. Miller & Isard, 1963).

What is not predictable from the hypothesis is that if two signals arrive close together, one in each message, then it is unlikely that both will be detected (see also Mowbray, 1964). It seems that the commitment to translate the signal into a response imposes a further limitation upon, or interferes with, speech recognition. The result is analogous to refractoriness in serial reaction (Shaffer, 1968) and the cause may be the same in each case. If interference produces a delay of recognition, then this would appear as a delayed second response in serial reaction and a missed signal in dichotic monitoring, since, with continuous messages, the alternate signal would be displaced by subsequent input prior to recognition.

REFERENCES

BROADBENT, D. E. *Perception and communication.* Oxford, Pergamon Press, 1958.

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

DEUTSCH, J. A., & DEUTSCH, D. Attention: Some theoretical considerations. Psychological Review, 1963, 70, 80-90.

HALLE, M., & STEVENS, K. N. Speech recognition: A model and a programme for research. In J. A. Fodor and J. J. Katz (Eds.), *The structure of language.* Englewood Cliffs, N.J.: Prentice-Hall, 1964.

MILLER, G. A., & ISARD, S. Some perceptual consequences of linguistic rules. Journal of Verbal Learning & Verbal Behavior, 1963, 2, 217-228.

MILLER, G. A., & LICKLIDER, J. C. R. The intelligibility of interrupted speech. Journal of the Acoustical Society of America, 1950, 22, 167-173.

MORAY, N. Attention in dichotic listening. Quarterly Journal of Experimental Psychology, 1959, 11, 56-60.
MOWBRAY, G. H. Perception and retention of verbal information presented during auditory shadowing. Journal of the Acoustical Society of America, 1964, 36, 1459-1464.
NEISSER, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.
SHAFFER, L. H. Refractoriness in information processing. Quarterly Journal of Experimental Psychology, 1968, 20, 38-50.
TREISMAN, A. Contextual cues in selective listening. Quarterly Journal of Experimental Psychology, 1960, 12, 242-248.
TREISMAN, A. Selective attention in man. British Medical Bulletin, 1964, 20, 12-16.
TREISMAN, A., & GEFFEN, G. Selective attention: Perception or response? Quarterly Journal of Experimental Psychology, 1967, 19, 1-17.

NOTES

1. Address: Department of Psychology, University of Exeter, Gandy Street, Exeter, England.

2. We gratefully acknowledge the financial support of the Science Research Council for this research, which is part of a study on serial skills.

(Accepted for publication April 23, 1969.)