Memory & Cognition
1977, Vol. 5 (3), 331-334

Probing echoic memory with different voices

DAVID J. MADDEN and JARVIS BASTIAN
University of California, Davis, California 95616

Considerable evidence has indicated that some acoustical properties of spoken items are preserved in an "echoic" memory for approximately 2 sec. However, some of this evidence has also shown that changing the voice speaking the stimulus items has a disruptive effect on memory which persists longer than that of other acoustical variables. The present experiment examined the effect of voice changes on response bias as well as on accuracy in a recognition memory task. The task involved judging recognition probes as being present in or absent from sets of dichotically presented digits. Recognition of probes spoken in the same voice as that of the dichotic items was more accurate than recognition of different-voice probes at each of three retention intervals of up to 4 sec. Different-voice probes increased the likelihood of "absent" responses, but only up to a 1.4-sec delay. These shifts in response bias may represent a property of echoic memory which should be investigated further.

The concept of an "echoic" memory was first developed by Neisser (1967), who proposed that a listener possesses a fairly literal but rapidly fading representation of recent auditory events. Research motivated by this concept suggests that the exact time course of such auditory information depends upon the type of stimuli that are presented and decisions required. For example, the effect of backward masking on tone identification is limited to target-mask intertone intervals of 250 msec and less (Massaro, 1970); but in situations involving the recall of verbal events there is evidence that the effects of acoustical variables can endure for approximately 2 sec. Conrad (1967) found that the number of acoustically related errors in the recall of visually presented consonants was higher at a 2.4-sec than at a 7.2-sec retention interval. Lindley and Brown (1971) have shown that errors in the recall of acoustically similar paired associates were greater than errors in the recall of semantically related pairs at a 2.5-sec retention interval, but not after 12.5 sec. The ability of a stimulus suffix to displace recently presented words or syllables from "precategorical acoustic storage" (Crowder & Morton, 1969) is substantially reduced if the suffix is delayed for more than 2 sec after the final stimulus item (Crowder, 1971). As Crowder (Note 1) emphasizes, while these types of results all support the notion of an echoic memory, they need not imply that verbal identification is dependent upon an unprocessed echoic trace. Rather, echoic information is parallel and supplementary to verbal identity in remembering auditory events.

One set of findings that has been difficult to accommodate to the notion of echoic memory involves the relation between the voice in which the stimulus items are spoken and the speed of recognition judgments. Cole, Coltheart, and Allard (1974) presented listeners with pairs of spoken consonants or vowels in a timed same/different recognition test. The items of a pair could be spoken in either the same voice or different voices, a variable which was not relevant to the subjects' decisions. Correct reaction time (RT) to same-voice pairs was significantly faster than to different-voice pairs for both "same" and "different" judgments, which is consistent with the recognition advantage usually present for pairs of identical items as compared with those which contain nonidentical members of the same semantic category (Posner, 1969). Similar results had previously been obtained by Cole and Scott (1972) and Springer (1973). Yet Cole et al. (1974) found this advantage of same-voice pairs to be present up to an interitem interval of 8 sec, and this is considerably longer than the duration of other echoic effects. In fact, Craik and Kirsner (1974) have obtained an even more extensive facilitation for same-voice stimulus pairs. In their experiment subjects judged each word in a spoken list as "old" or "new" at presentation; both accuracy and RT were superior for same-voice repetitions over different-voice repetitions up to delays of 2 min and 31 intervening items. This result led Craik and Kirsner (1974) to doubt the validity of distinguishing echoic storage from long-term memory.

Since the long-lasting influence of the identity of the speaker's voice contrasts with the brevity of other acoustical effects in memory, there may be some component of this voice change effect which previous research has ignored. In particular, it is not clear from previous research what type of processing allows for such an extended recognition advantage for same-voice pairs. Alternating the speakers' voices could be affecting the subject's bias to respond "same" or "different" rather than making the stimuli more or less discriminable. Signal detection theory (e.g., McNicol, 1972) can be used to decide between these alternatives. Parks (1966), for example, has shown that performance in

We are grateful to Dean Malley and Dan Morrow for technical assistance, and to Ted Parks for his helpful comments on earlier versions of the manuscript.

recognition memory experiments can be described within the framework of signal detection theory, under the assumption that there is a psychological dimension of "familiarity" associated with the recognition items according to which decisions are made. Additionally, RT in auditory detection has been shown to be inversely related to the confidence subjects have in their decisions. The farther away an observation lies on either side of the value of familiarity adopted as a criterion by a subject, the greater is the confidence in the decision about this observation, and the shorter is the associated RT (Emmerich, Gray, Watson, & Tanis, 1972). Thus, the RT advantage obtained for same-voice stimulus pairs could be, at least in part, a result of subjects' adopting different response criteria under each of the voice conditions. If this were the case, then the duration of the voice effects obtained by Cole et al. (1974) and Craik and Kirsner (1974) would be more understandable, since these effects would be more closely related to subjects' biases to respond than to the retention of short-lived echoic information.

The effects of changing the voice speaking the stimulus items are consistent when RT is the dependent variable, so the present experiment instead used measures derived from signal detection theory to investigate the effects of these acoustical changes on response bias and memory sensitivity. On each trial, listeners were presented with an auditory ensemble of six digits and performed present/absent judgments of a single recognition probe digit after an interval of .5, 1.4, or 4.0 sec. It was hypothesized that changing the voice in which the probe appeared would affect bias but not sensitivity, and that this effect would not decrease over the retention interval. Secondarily, the relation of these acoustical changes to the judgments of presence and absence was examined.

METHOD

Subjects

Twenty-four undergraduates (15 males, 9 females) who reported that they possessed normal hearing were selected for the experiment. They received class credit in their introductory psychology courses as compensation.

Stimuli

The experimental tape employed was composed of 108 trials; each trial consisted of three dichotic pairs of digits followed by a binaural probe digit. The dichotic pairs were derived from a single master recording of the digits made by a male speaker. The members of each dichotic pair were first recorded so that the onsets coincided within 10 msec on each channel, and then the pairs were assembled into the trial configurations so that the onset-to-onset interval between each pair was 1 sec. Since natural speech was used, the offset-to-onset intervals varied, ranging from 325 to 619 msec, with an average duration of 487 msec. Each of the digits from 1 to 10 (excluding 7, which is disyllabic) appeared as a dichotic item 72 times over the course of the 108 trials. The distribution of the nine digits was random with regard to channel and serial position within a trial.

On 72 of the trials the binaural recognition probe actually matched one of the preceding six digits. Thirty-six of these "target" trials probed items present on the left channel, and 36 probed those in the right. Each of the three serial positions within a trial carried a target 12 times for each channel. On the remaining 36 "nontarget" trials, the recognition probe had not been present in the three pairs of that trial.

In order to investigate the effects of changing voices on memory for the dichotic items, half of the recognition probes were derived from a single recording of the digits spoken by a female speaker. Thus, on 18 of the target trials on each channel (six for each serial position) the probe was delivered by the female's voice, and on the other half of the target trials the probe was delivered by the male's voice which had produced the dichotic digits. For each of the target trials, each of the nine digits was a probe twice in each voice for items on each channel. The 36 nontarget trials were also divided equally between male- and female-voice probes, each digit being used twice in each voice.

Recognition judgments were made during a 5-sec interval following the probe. For target trials this interval was terminated by a feedback tone presented in the channel in which the target had been delivered, and for the nontarget trials a binaural buzz terminated the response interval. The next trial began 1 sec after this feedback had been presented.

Procedure

Instructions for the experimental task were tape recorded and played to each subject at the start of the session. Although no practice trials were given, the instructions described the possible voice changes of the probe and included examples of all trial types. Subjects were told that their task was to judge whether the probe digit, regardless of voice, had been presented in the preceding three dichotic pairs. Answer sheets were provided containing five blanks which formed a rating scale in which subjects indicated their confidence in their decision on each trial. The alternatives ranged from "certain presence" to "certain absence" of the probe digit, with the middle blank indicating uncertainty either way.

Each subject listened to three versions of the stimulus tape, which were identical except that each tape had one of three retention intervals between the offset of the third dichotic pair and the onset of the recognition probe: .5, 1.4, or 4.0 sec. Four subjects were tested at each of the six possible orders of three stimulus tapes. The stimuli were delivered over Sharpe HA-10A headphones to the subjects within a heavily sound-shielded chamber. Listening levels were adjusted to the subjects' comfort at the beginning of the experiment and then remained constant over the whole procedure. Headphone channels were reversed between subjects. The stimulus tapes were reproduced on an Ampex 300 tape deck located outside the listening chamber. The experimental session lasted approximately 2 h, including a short rest period between the second and third stimulus tapes.

RESULTS

Judgments of Presence and Absence

The percentage of correct decisions was obtained for each of the 12 combinations of trial type (target, nontarget), voice of the probe (same, different), and delay of the probe (.5, 1.4, 4.0 sec). For the target trials, this is the percentage of times the categories for certain or possible presence of the probe were used (blank 1 or 2); for the nontarget trials, this is the percentage of times the categories for possible or certain absence of the probe were used (blank 4 or 5). An analysis of variance performed on these percentages yielded a significant main effect of trial type [$F(1,23) = 8.09$, $p < .01$], perform-

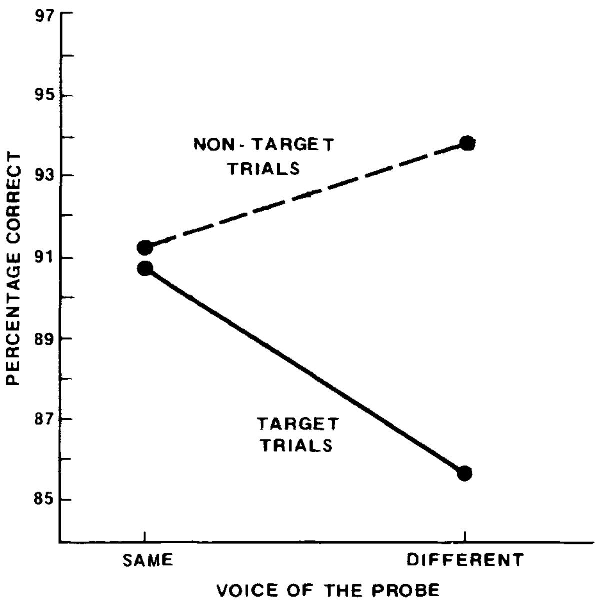

Figure 1. Trial Type by Voice of the Probe interaction.

ance on nontarget trials being superior to that on target trials. In addition, the Voice of the Probe by Trial Type interaction was significant [$F(1,23) = 16.78$, $p < .01$] and is shown in Figure 1. No other main effect or interaction was significant. Paired comparisons showed that subjects were more accurate on nontarget trials than on target trials when the probe was spoken in a different voice [$t(23) = 4.98$, $p < .01$], although the trial types were not different from each other when the probe occurred in the same voice. For the nontarget trials accuracy was significantly greater in the different-voice than in the same-voice condition [$t(23) = 2.77$, $p < .05$], while for the target trials the reverse was true [$t(23) = 3.86$, $p < .01$].

Memory Sensitivity

The present data were also analyzed within a signal detection framework. The measure P(A), a nonparametric estimate of the proportion of area under the ROC curve described by McNicol (1972, pp. 113-115) was obtained for each of the 18 combinations of voice of the probe (same, different), delay of the probe (.5, 1.4, 4.0 sec), and serial position (1, 2, 3). Larger P(A) values reflect greater recognition sensitivity. An analysis of variance performed on the arc-sine root transformations of the P(A) values yielded significant main effects of delay of the probe [$F(2,46) = 4.10$, $p < .05$] and serial position [$F(2,46) = 7.82$, $p < .01$]. Contrary to expectation, the voice of the probe main effect was also significant [$F(1,23) = 8.42$, $p < .01$], same-voice probes providing for greater accuracy than different-voice probes. These effects are shown in Figure 2a. Paired comparisons indicated that P(A) was lower at the 4.0-sec delay than at the 1.4-sec delay [$t(23) = -3.96$, $p < .01$]. However, performance at the .5-sec delay was not differ-

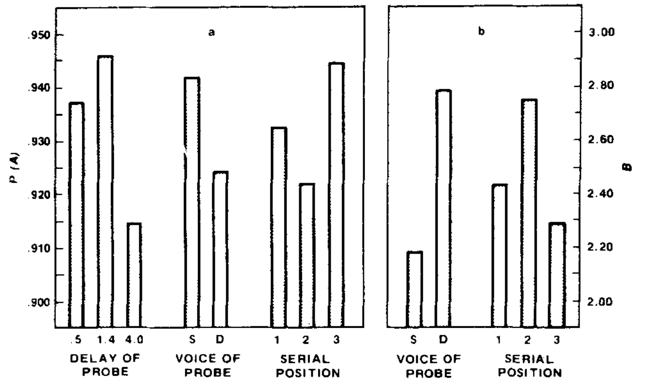

Figure 2. Significant main effects in the analyses of the measures P(A), an estimate of memory sensitivity, and B, an estimate of response bias. Increasing P(A) values represent greater sensitivity, and increasing B values represent stricter criteria for "present" judgments.

ent from that at either the 1.4-sec or 4.0-sec delay. Performance at the second serial position was worse than at both the first serial position [$t(23) = -2.15$, $p < .05$] and the third serial position [$t(23) = -3.84$, $p < .01$]. The first and third serial positions did not differ significantly. The two- and three-way interactions were not significant.

Response Bias

The measure B (McNicol, 1972, pp. 123-127), a nonparametric estimate of response bias, was also obtained for each subject under each condition. It is defined as the point on the rating scale where the subject is equally disposed toward "present" and "absent" responses; higher B values represent stricter criteria being set for "present" responses. An analysis of variance performed on this variable revealed the hypothesized main effect of voice of the probe [$F(1,23) = 21.16$, $p < .01$], different-voice probes showing higher B values than same-voice

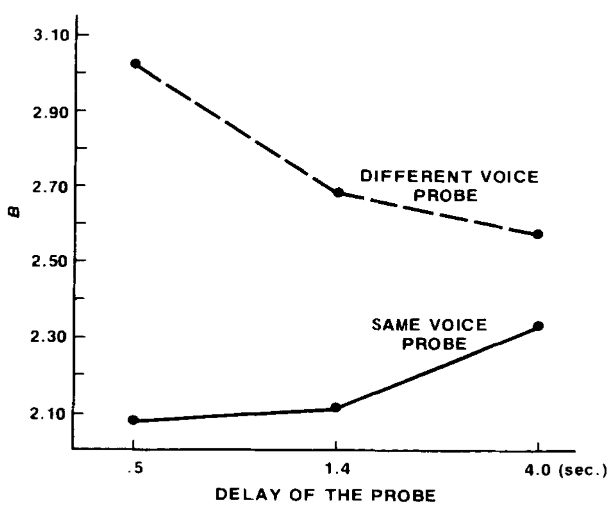

Figure 3. Delay by Voice of the Probe interaction for the measure B. Increasing B values represent stricter criteria for "present" judgments.

probes. The serial position main effect was also significant [$F(2,46) = 7.22$, $p < .01$]. These results are shown in Figure 2b. Paired comparisons of the serial position effect showed that B at the second serial position was higher than at both the first [$t(23) = 2.46$, $p < .05$] and third [$t(23) = 3.60$, $p < .01$] serial positions; the first and third serial positions did not differ significantly from each other. Surprisingly, the Delay by Voice of the Probe interaction, shown in Figure 3, was also significant [$F(2,46) = 3.42$, $p < .01$]. Different-voice probes resulted in higher B scores than did same-voice probes at the .5-sec delay [$t(23) = 5.08$, $p < .01$] and at the 1.4-sec delay [$t(23) = 3.95$, $p < .01$], but this effect of the probe voice did not extend to the 4.0-sec delay.

DISCUSSION

The superior accuracy on nontarget as opposed to target trials, evident in Figure 1, suggests that judgments of absence are easier than judgments of presence, particularly when the probe item occurs in a different voice. In addition, when the probe is actually present in the memory items, recognition is apparently assisted by acoustical identity between the memory and test items, while the recognition that the probe is absent is aided by acoustical dissimilarity. The signal detection analysis, however, shows that these changes in accuracy are the result of two different processes. When the voice of the recognition probe is shifted from same to different, there is first of all an impairment at all retention intervals in the ability to distinguish probe digits that were present from those that were absent, and this is reflected in the lower P(A) values for different-voice probes. A second effect is that, when the probe appears in a different voice, subjects are moved to respond "absent" more often, which is reflected in the increased B values. Yet, as Figure 3 shows, this response bias does not extend beyond the 1.4-sec retention interval, contradicting the hypothesis that this effect would mimic the long-lasting recognition differences obtained by Cole et al. (1974) and Craik and Kirsner (1974). Recognizing items spoken in different voices thus remains an anomaly, since voice identity can impair memory sensitivity over an interval nearly twice as long as that associated with other acoustical effects.

One way to bring the present results into line with other experiments on echoic memory may be to alter the current assumptions about what type of processing is associated with echoic information. In the present experiment the stricter criterion for "presence" required by the different-voice probes was restricted to the .5- and 1.4-sec retention intervals, which is consistent with traditional estimates of the duration of echoic memory. Thus, perhaps echoic memory does not just provide information about the auditory properties of linguistic items, but also briefly creates a bias in judgments regarding an item's reoccurrence. A first step in establishing this possibility would be to determine if such a bias were more closely related to the kinds of judgments required of listeners or to the nature of the particular acoustical changes imposed.

REFERENCE NOTE

1. Crowder, R. *Audition and speech coding in short-term memory.* Paper presented at the Seventh International Symposium on Attention and Performance, Senanque, France, 1976.

REFERENCES

Cole, R., Coltheart, M., & Allard, F. Memory of a speaker's voice: RT to same- or different-voiced letters. *Quarterly Journal of Experimental Psychlogy*, 1974, **26**, 1-7.

Cole, R., & Scott, B. Distinctive feature control of decision time: Same-different judgments of simultaneously heard phonemes. *Perception & Psychophysics*, 1972, **12**, 91-94.

Conrad, R. Interference or decay over short retention intervals? *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 49-54.

Craik, F., & Kirsner, K. The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, 1974, **26**, 274-284.

Crowder, R. Waiting for the stimulus suffix: Decay, delay, rhythm and readout in immediate memory. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 324-340.

Crowder, R., & Morton, J. Precategorical acoustic storage (PAS). *Perception & Psychophysics*, 1969, **5**, 365-373.

Emmerich, D., Gray, J., Watson, C., & Tanis, D. Response latency, confidence, and ROCs in auditory signal detection. *Perception & Psychophysics*, 1972, **11**, 65-72.

Lindley, R., & Brown, D. Acoustic and associative coding in short-term memory. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 14-21.

Massaro, D. Preperceptual auditory images. *Journal of Experimental Psychology*, 1970, **85**, 411-417.

McNicol, D. *A primer of signal detection theory.* London: Allen & Unwin, 1972.

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

Parks, T. Signal-detectability theory of recognition memory performance. *Psychological Review*, 1966, **73**, 44-58.

Posner, M. Abstraction and the process of recognition. In G. Bower & J. Spence (Eds.), *The psychology of learning and motivation* (Vol. 3). New York: Academic Press, 1969.

Springer, S. Memory for linguistic and nonlinguistic dimensions of the same acoustic stimulus. *Journal of Experimental Psychology*, 1973, **101**, 159-163.

(Received for publication March 15, 1977; accepted March 22, 1977.)