

The use of auditory and phonetic memory in vowel discrimination

Nelson Cowan

Department of Psychology, University of Missouri—Columbia, 210 McAlester Hall, Columbia, Missouri 65211

Philip A. Morse

New England Rehabilitation Hospital, Woburn, Massachusetts 01801 and Department of Psychology, Boston University, Boston, Massachusetts 02215

(Received 23 July 1985; accepted for publication 28 October 1985)

A commonly held assumption about memory for speech is that auditory memory is referred to only if phonetic memory does not contain the information needed for a particular trial. However, this assumption is in conflict with recent evidence [Crowder, *J. Exp. Psychol.: Learning, Memory, Cognition* 8, 153–162 (1982); Repp *et al.*, *J. Exp. Psychol.: Human Perception Performance* 5, 129–145 (1979)]. The present study provides additional data to help determine how auditory and phonetic memory are used in a vowel discrimination task, and what happens during memory decay. Experiment 1 was conducted to determine whether performance levels decline at similar rates on between- and within-category AX vowel comparison trials when certain methodological problems are removed. This was confirmed. Experiment 2 demonstrated that in the AX task there is a vowel order effect, as Repp *et al.* found, but that this effect increased across interstimulus delay intervals, in contrast to their findings. The results can be accommodated with a model in which the memory for a vowel is represented as a small, bounded area within the vowel space, and in which memory decay is represented by the expansion of that bounded area over time.

PACS numbers: 43.71.Es, 43.71.An

INTRODUCTION

A key question in the processing of language is how speech information is retained until the processing of that information can be completed. One procedure that has been proven useful in examining memory for speech is the two-stimulus, AX comparison procedure (Crowder, 1982; Pisoni, 1973; Repp *et al.*, 1979). In this paradigm, the subject's ability to determine whether the two sounds being compared are the same or different depends on the memory for the first sound across the interstimulus interval. Pisoni (1973) presented evidence for the role of phonetic memory, as well as auditory memory, in the processing of speech sounds within this paradigm. For example, when subjects compare two different vowels in this procedure, they rely on both phonetic memory (which can be useful in comparing vowels from different phonetic categories) and auditory memory (which can be useful in determining differences both within and between phonetic categories). However, the sensitivity of auditory memory decreases dramatically as the interstimulus delay is increased, reaching an asymptotically weak level by about 3 s (Crowder, 1982). In contrast, it has been assumed that phonetic memory does not decay as quickly as auditory memory. Crowder (1982) has summarized this widely held model of performance in the AX paradigm:

Subjects try first to make a same-different decision based on phonetic labels and, only after that has failed, go on to consult auditory memory. The rule is, *if the two sounds have different names, say "different," otherwise compare the sounds themselves* (p. 162).

Furthermore, as Crowder has noted, this model predicts that as the interstimulus interval is increased, differential rates of performance decrement should be obtained for pairs

of vowels drawn from different phonetic categories and for pairs drawn from within a single phonetic category. Specifically, because phonetic memory is assumed to decay less quickly than auditory memory, and phonetic memory can be useful only in between-category comparisons, performance should be more stable in between- than within-category comparisons.

However, in contrast to this prediction, evidence of an interaction between interstimulus interval and type of vowel comparison (between- versus within-category) has failed to emerge in the previous studies in this literature (Crowder, 1982; Pisoni, 1973; Repp *et al.*, 1979). Performance levels have been found to decline across interstimulus intervals at about the same rate in both the between- and within-category contrasts even though the absolute level of performance has been consistently found to be higher in between-category vowel contrasts.

The consistent failure to observe a differential decrement over time in between- versus within-category vowel discrimination certainly casts serious doubt upon this simple model of phonetic and auditory memory. However, there is one procedural aspect that characterizes all of the previous studies in this area that may account for the failure to observe this predicted interaction. In these studies subjects were presented with between- and within-category vowel comparison randomized together over trials. Since phonetic memory would have been of little use on within-category trials, this randomization procedure may have discouraged subjects on many or most between-category trials from utilizing their phonetic memory. As a result, they may have relied more on auditory memory (which decays rapidly over the interstimulus interval), yielding parallel decrements in both within- and between-category discriminations with in-

creasingly longer interstimulus delays.

If subjects could be encouraged to utilize their phonetic memory abilities in between-category comparisons, then this might result in more stable performance over increasingly longer interstimulus intervals, thereby providing support for the model of phonetic and auditory memory outlined above. Experiment 1 of the present study was designed to encourage subjects to employ their phonetic memory abilities in between-category comparisons in two different ways. First, between- and within-category vowel contrasts were presented in separate trial blocks. Second, subjects were familiarized with instructions that either provided them with the phonetic labels of the vowels used in the experiment or were nonspecific with regard to phonetic information. It was anticipated that a blocked presentation and phonetically explicit instructions would increase the subject's reliance on phonetic memory in the between-category comparisons, resulting in a significant interaction of delay interval and comparison type (between- versus within-category).

1. EXPERIMENT 1

A. Method

1. Subjects

The subjects were native speakers of English between the ages of 17 and 40, with no known hearing problems or experience with synthetic speech, and who were paid for their efforts. In the main experiment (1a), 32 subjects received AX vowel sequences with between- and within-category pairs in separate trial blocks. A smaller experiment (1b) was conducted with eight additional subjects, to replicate previous results with between- and within-category vowel pairs randomized together.

2. Design

Every subject received 240 between- and 240 within-category vowel comparison pairs, and also 480 control pairs in which the two vowels were identical. Trials of each type were divided evenly among five interstimulus delay intervals: 0, 250, 500, 1000, and 2000 ms. The experiment began with one of two slightly different sets of familiarization instructions, with half the subjects receiving each set in both experiments 1a and 1b.

3. Stimuli

The synthetic speech stimuli were 50-ms vowels identical to /i/₁, /i/₃, /i/₅, and /i/₇ from Pisoni's (1973) 13-point, /i-ɪ-ε/ continuum. The first three formant values for each of the stimuli (in Hz) were as follows: /i/₁: F1 = 270, F2 = 2300, F3 = 3019; /i/₃: F1 = 298, F2 = 2226, F3 = 2902; /i/₅: F1 = 336, F2 = 2144, F3 = 2776; /i/₇: F1 = 374, F2 = 2070, F3 = 2666. All vowels had fourth and fifth formants fixed at 3500 and 4500 Hz, respectively. The stimuli were constructed with a software series-resonance synthesizer (Klatt, 1980) at the Waisman Center of the University of Wisconsin, Madison, and were digitally stored and recorded using the VOCAL system of stimulus patterning (Gillman *et al.*, 1975).

A familiarization tape was constructed with eight series

of stimuli. Each series consisted of the four test vowels in order, with a 4-s interstimulus interval between vowels and 10 s between series. Odd-numbered series progressed from /i/₁ to /i/₇, whereas even-numbered series progressed from /i/₇ to /i/₁.

On each AX tape, stimuli were recorded in pairs with 0, 250, 500, 1000, or 2000 ms between vowels in a pair. A 100-ms, 600-Hz tone presented 700 ms before the first vowel in every pair served as a ready signal. AX trials were separated by 4-s silent intervals. There were four discrimination tapes for subjects in the blocked procedure (experiment 1a), each containing 120 trials. Half of the trials on each tape were control trials (/i/₃-/i/₃ and /i/₅-/i/₅). Two of these tapes contained between-category comparisons (/i/₃-/i/₅ and /i/₅-/i/₃), and two tapes contained within-category comparisons (/i/₃-/i/₁ and /i/₅-/i/₇). Every subject listened to each tape twice (once per day) for a total of 960 AX trials per subject. For the randomized procedure (experiment 1b), four different AX tapes were constructed. Each tape had 30 between-category, 30 within-category, and 60 control trials, and, across the experiment, these subjects heard the same 960 AX trials as subjects in the blocked procedure.

There was also an isolated-vowel identification tape containing 20 tokens of each of the four test vowels (or 80 tokens in all), with 4 s between vowels and a 1-min break halfway through the tape.

4. Procedure

Tapes were played binaurally over TDH-49 headphones in a sound-attenuated chamber, using a SONY-756 tape deck. The sound-pressure levels were set at 78 ± 1 dB (1a) with a General Radio sound level meter (1551-C) and ear-phone coupler (1560-P83).

Each subject participated on two consecutive days. On both days of the experiment, every subject began with the familiarization tape and then received four AX tapes. Before listening to the familiarization tape, "explicitly instructed" subjects were told that they would hear four slightly different vowels, shifting from /i/ as in "beet" to /ɪ/ as in "bit" and back again. In contrast, subjects who received nonspecific instructions were told only that they would hear four slightly different vowels, shifting gradually from one to the next and back. In the AX discrimination task, all subjects were asked to decide if the vowels in each pair were "the same" or "different," and were informed that the test vowels differed on roughly half of the trials. It was stressed that any acoustic difference between vowels should be considered. Responses were recorded on computer scorable test sheets. In the blocked presentation group, between- and within-category tapes were played in an alternating order, with half of the subjects receiving a between-category tape first and half receiving a within-category tape first. Out of the eight possible alternating orders of the tapes, four were employed on the first day of the experiment, and the reverse orders were employed on the second day. In the randomized presentation group, each subject heard the four randomized tapes in an order based on one row of a Latin square, with the order reversed on the second day. Finally, at the end of the second day of the experiment, all subjects were told that they

had heard /i/ and /ɪ/ vowels, and were administered the 80-item identification tape.

B. Results and discussion

1. Vowel identification

Consistent with Pisoni (1973), the vowels presented in isolation were identified with a sharp boundary between /i/₃ and /ɪ/₅. For the blocked presentation group, the proportion of tokens labeled /ɪ/ as in "bit" rather than /i/ as in "beet" were: #1, 0.0; #3, 0.11; #5, 0.94; and #7, 0.98. For the randomized presentation group, the proportions were: #1, 0.0; #3, 0.14; #5, 0.93; and #7, 0.98.

2. Blocked procedure

In order to assess the sensitivity of memory in the AX task, it is necessary to obtain a measure free of criterion bias, such as d' . However, an ordinary d' measure cannot be used when some cell means equal 0.0 or 1.0, as in the present experiment. In order to circumvent this problem a method devised by Crowder (1982) was used, in which subjects were grouped into "supersubjects" ($n = 4$) and d' scores were calculated for each of eight supersubjects in each condition using a table of d' for two-stimulus discrimination (Kaplan *et al.*, 1978). Criteria for the valid use of supersubject d' discussed by MacMillan and Kaplan (in press) were satisfied by the present data. These d' scores were entered into a delay (5) × contrast type (2) × supersubject (8) analysis. The instructions factor was omitted from this analysis, inasmuch as analyses of percent correct indicated that it did not alter the pattern of results.

In the d' analysis, there was a strong advantage for between- over within-category comparisons, $F(1,7) = 235.4$, $p < 0.001$, as well as a large delay effect, $F(4,28) = 122.4$, $p < 0.001$. There was also an interaction of delay × contrast type, $F(4,28) = 19.91$, $p < 0.001$. However, as can be seen in Fig. 1(a), this interaction is largely a result of the reduced advantage for the between-category condition at the 0-ms interval when compared to longer intervals. When only the last four delay intervals were entered into the analysis, the contrast type and delay effects became even larger, but the interaction of delay × contrast type did not approach significance. It is clear that between- and within-category performance levels declined at roughly equivalent rates despite the use of a blocked procedure.

Performance was also examined separately for vowel contrasts in which the second vowel was closer to /i/₁ than the first vowel, versus pairs in which the second vowel was closer to /ɪ/₇. The group d' scores for vowels presented in each order are shown in the top panel of Fig. 2. This figure suggests that the decrease in performance across delays may be greater for vowel pairs in which the second vowel is closer to /ɪ/₇ than the first vowel, although the experiment included the necessary balanced order comparison only for between-category trials. It was not possible to obtain d' scores for each vowel order and delay separately for subjects or supersubjects, so statistical analyses of between-category blocks that included the vowel order effect were performed using percent correct scores rather than d' . In these analyses

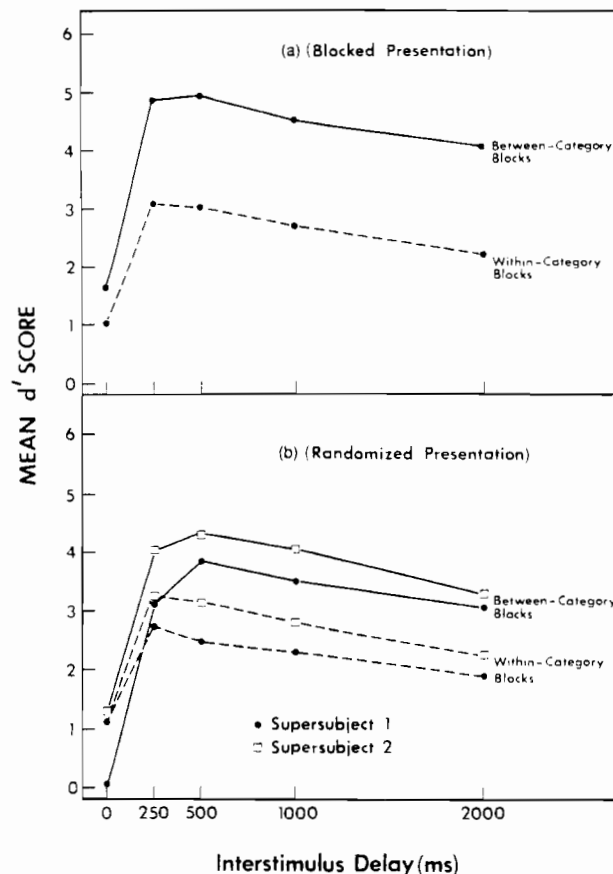


FIG. 1. Mean d' score in experiment 1 as a function of condition (between- versus within-category) and delay. Part (a): mean functions for eight supersubjects who received between- and within-category trials in separate trial blocks. Part (b): individual functions for two supersubjects who received a totally randomized presentation. Each supersubject is based upon four subjects.

there was a main effect of delay when all five delays were included in the analysis, $F(4,124) = 42.05$, $p < 0.001$, and when only the last four delays were included, $F(3,93) = 11.46$, $p < 0.001$. Moreover, performance levels were higher for pairs shifting toward /i/₁ than for pairs shifting toward /ɪ/₇ in an analysis with all five delays, $\bar{x} = 0.89$ vs 0.85, $F(1,31) = 12.2$, $p < 0.002$, and with only the last four delays, $\bar{x} = 0.95$ vs 0.91, $F(1,31) = 20.17$, $p < 0.001$. Most importantly, though, the delay × vowel order interaction for the between-category trials was significant when all five delays were included in the analysis, $F(4,124) = 4.06$, $p < 0.005$, and when only the last four delays were included, $F(3,93) = 14.09$, $p < 0.001$. The direction of the vowel order effect in the present study is consistent with the results of Repp *et al.* (1979), but Repp *et al.* obtained no interaction of delay × vowel order.

3. Totally randomized procedure

The eight subjects who received a totally randomized test procedure were divided into two supersubjects, and the results are plotted in Fig. 1(b). As in previous studies and the present blocked procedure, between- and within-category performance levels apparently declined at equal rates. However, the blocked presentation procedure does seem to

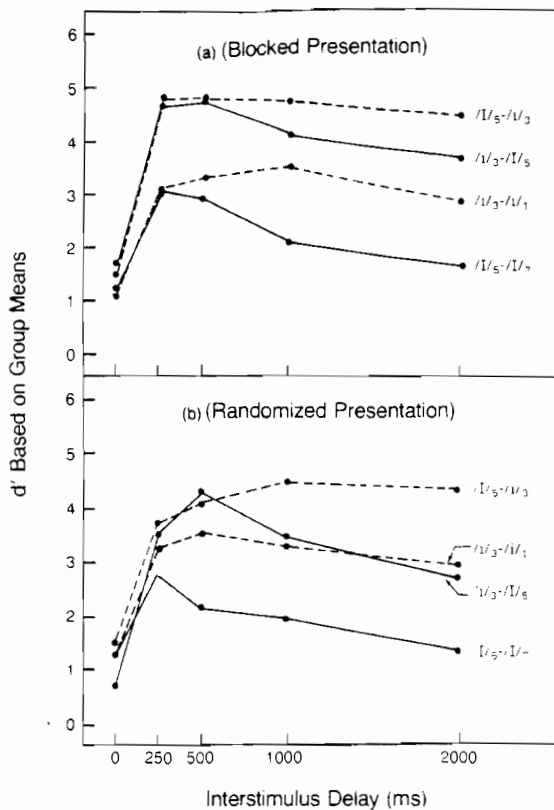


FIG. 2. The d' scores in experiment 1 based on group mean percentages, separately for each vowel contrast. Solid lines are pairs shifting toward /i/7 and broken lines, toward /i/1. (a) Blocked presentation; (b) randomized presentation.

have increased the advantage of between- over within-category discrimination in comparison to the randomized procedure [cf. Fig. 1(a) versus each of the two supersubjects in Fig. 1(b)]. This observation is supported by the pattern of d' means for the blocked and randomized groups averaged over supersubjects across the last four delays (so as to disregard backward masking at the 0-ms delay). In the blocked presentation group, the mean advantage of between- over within-category d' was 1.73, whereas in the randomized presentation group the between-category advantage was only 1.05. Moreover, the difference between the blocked and randomized presentations occurred primarily because the blocked presentation permitted superior performance on the between-category trials. The mean d' across the last four delays was 4.46 in the blocked presentation and only 3.63 in the randomized condition, a difference that was significant in an unequal n (general linear model), group \times delay ANOVA for d' scores, $F(1,8) = 6.46$, $p < 0.04$. Mann-Whitney U tests (1-tailed) indicated that this advantage for the blocked presentation was significant at each of these four delays ($p < 0.01$, $p < 0.05$, $p < 0.05$, and $p < 0.01$). In contrast, similar analyses for the within-category trials did not approach significance, $F(1,8) = 0.32$. The mean within-category d' across the last four delays was 2.73 in the blocked presentation group and 2.58 in the randomized group.

Thus, although blocking of between- versus within-category trials increases the subject's use of phonetic category

information, it does not alter the memory decay function observed in either condition; when subjects could tell in advance that phonetic information would be relevant (i.e., in between-category blocks), memory sensitivity declined across delay intervals at an unchanged rate.

Finally, in analyses of the percent correct scores with the vowel orders examined separately, there was a main effect of delay when all five delays were included in the analysis, $F(4,28) = 12.49$, $p < 0.001$, but not when only the last four delays were examined, $F < 1.0$. However, there was an advantage for vowel pairs shifting toward /i/1 rather than /i/7 with all five delays included, $\bar{x} = 0.85$ vs 0.77, $F(1,7) = 9.62$, $p < 0.02$, and also with only the last four delays included, $\bar{x} = 0.93$ vs 0.85, $F(1,7) = 11.34$, $p < 0.02$. Finally, the order-balanced contrast /i/3 vs /i/5 again produced a delay \times order interaction in the percent correct scores, both with all five delays included, $F(4,28) = 4.60$, $p < 0.006$, and with just the last four delays included, $F(3,21) = 7.07$, $p < 0.002$. As in the blocked presentation procedure, performance was more stable across delays in the /i/5-/i/3 order. The results of the randomized condition are shown separately for each vowel order in the bottom panel of Fig. 2.

Clarification of the vowel order effect may be quite important for an understanding of memory for vowel quality, inasmuch as the vowel order (rather than category knowledge) seems to have the largest effect on the rate of memory decay. Repp *et al.* (1979) were unable to determine the basis of the order effect. One appealing hypothesis considered by Repp *et al.* was that as the memory of a vowel decays, the representation shifts in the vowel space. For example, specific vowel memories might shift toward the neutral vowel /ə/ (schwa) (see Ladefoged, 1982, p. 198 for a description of the vowel space). This convergence in vowel representation could account for the loss of distinctiveness of vowel memories over time. In some ways, this hypothesis fits the Repp *et al.* (1979) data well. In their /i/-/ɪ/-/ε/ continuum, there was better performance on vowel pairs shifting toward /i/, but the order effect was reversed for the extreme /ε/ end of the continuum. This is exactly what one would expect on the basis of a memory shift toward /ə/, because the vowel closest to /ə/ on the continuum is between the prototypical /i/ and /ε/ values (the location of these vowels within the vowel space can be observed in Fig. 4).

Repp *et al.* (1979) rejected the vowel shift hypothesis because they did not obtain the delay \times vowel order interaction predicted by that hypothesis. However, there are several reasons to suspect that the vowel shift hypothesis was rejected prematurely by Repp *et al.* Only two delays were used in their experiments, and the dependent measure was the percentage of "hits" on experimental (change) trials rather than a measure that also takes into account "false alarms" on control (no-change) trials, such as d' . The results of the present experiment 1 suggest that, at least for between-category trials, there is an interaction of delay \times vowel order. However, in that experiment the vowel order effect was not examined with order-balanced stimuli for within-category contrasts. The purpose of experiment 2 was to obtain these within-category data in order to further assess the viability of the vowel shift hypothesis.

II. EXPERIMENT 2

A. Method

1. Subjects

The subjects were seven English-speaking adults with no known hearing problem or experience with synthetic speech, and who did not participate in the first experiment.

2. Stimuli

The same four vowels were used as in experiment 1, but the sets of vowel pairs on the AX tape were different. Each subject listened to 480 vowel pairs, divided equally among four experimental contrasts ($/i/_{1-}/i/_{3-}/$, $/i/_{3-}/i/_{1-}/$, $/I/_{5-}/I/_{7-}/$, and $/I/_{7-}/I/_{5-}/$) and four corresponding control contrasts ($/i/_{1-}/i/_{1-}/$, $/i/_{3-}/i/_{3-}/$, $/I/_{5-}/I/_{5-}/$, and $/I/_{7-}/I/_{7-}/$). The silent interstimulus delays used were 125, 250, 500, 1000, and 2000 ms. The purpose of including 125 ms rather than 0 ms [as in Pisoni (1973) and experiment 1 of the present study] was to map out more completely the discrimination function. From previous research, it is not clear if peak discrimination would occur before or after 250 ms.

3. Procedure

Subjects first received the familiarization tape used in experiment 1, with the "nonspecific" type of instructions, followed by the 480-item AX tape. Finally, each received the 80-item, isolated-vowel identification test.

B. Results and discussion

1. Vowel identification

In accord with experiment 1, the vowels were identified as in Pisoni (1973). The proportions of $/I/$ identification were: #1, 0.0; #3, 0.11; #5, 0.94; and #7, 0.99.

2. AX task

The AX discrimination results were examined in several ways. First, as in experiment 1, d' scores (based on Kaplan *et al.*, 1978) were calculated from group mean percentages (for a validation, see Macmillan and Kaplan, in press). The results shown in Fig. 3 suggest that the vowel order effect did increase across interstimulus delay intervals. As Fig. 3 illustrates, peak performance was slightly higher in vowel pairs shifting toward $/I/_{7-}$ rather than $/i/_{1-}$, but performance after a 2-s delay was much better in pairs shifting toward $/i/_{1-}$. Second, it was possible to calculate d' scores (again from Kaplan *et al.*, 1978) separately for six of the seven subjects. The seventh subject was 100% correct in some conditions, and therefore had to be omitted. When the delay \times vowel order effect was examined in individual subjects (using either the peak d' minus d' at 2000 ms or d' at 250 ms minus d' at 2000 ms), all six subjects had more memory loss for vowel pairs in which the second vowel was closer to $/I/_{7-}$, $p < 0.05$ (sign test). Third, performance of all seven subjects was examined by assigning the 100% correct performance of the seventh subject a z score of 2.33 (= 99%) and subjecting the data to an analysis of variance. In this analysis, there was a significant effect of delay, $F(4,24) = 34.22$, $p < 0.001$, and of vowel category, $F(1,6) = 25.60$, $p < 0.001$, the latter based on

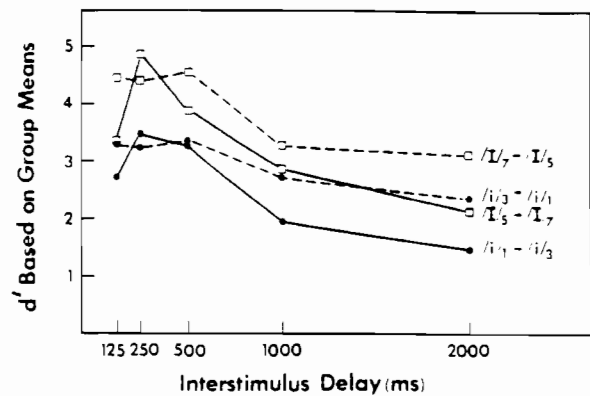


FIG. 3. The d' scores based on group percentages in experiment 2, separately for each vowel contrast. Solid lines are pairs shifting toward $/I/_{7-}$ and broken lines, toward $/i/_{1-}$.

the superiority of discrimination for the $/I/$ vowels (cf. Fig. 3). Most importantly, though, the delay \times vowel order interaction was significant, $F(4,24) = 4.12$, $p < 0.025$, because of the advantage for pairs shifting toward $/i/_{1-}$. Finally, this same result was obtained using the analysis employed by Repp *et al.* (1979) of p ("different" | different) responses or "hits," $F(4,24) = 3.56$, $p < 0.025$. In this analysis, the advantage of $/i/_{3-}/i/_{1-}$ over the opposite order increased from five percentage points at 250 ms to 31 points at 2000 ms, and the advantage for $/I/_{7-}/I/_{5-}$ over the opposite order increased from three points to ten points. The results of both experiments taken together indicate that the vowel order effect for both between- and within-category contrasts is in fact larger at longer delay intervals.

The pattern of vowel order effects was consistent across both experiments, but the performance levels for particular within-category contrasts differed markedly (cf. Figs. 2 and 3). Although we can offer no complete account of this aspect of the results, it is possible that performance levels were affected by the total set of contrasts that subjects had to detect. In experiment 1, for example, the most discriminable contrast was the between-category pair $/I/_{5-}/I/_{3-}$. The sensitivity to the within-category contrast beginning with the same vowel, $/I/_{5-}/I/_{7-}$, might be reduced as a consequence of subjects comparing vowel pairs across trial blocks. The $/i/_{3-}/I/_{5-}$ contrast is less discriminable, and probably would not affect the contrast $/i/_{3-}/i/_{1-}$ as strongly. Within experiment 2 there were no between-category trials, and differences between the levels of performance for $/i/$ vs $/I/$ may have reflected the differential sensitivity of these regions of the vowel space more accurately than in experiment 1.

A final important aspect of the results is the delay at which peak performance occurred. Delays of 125 and 250 ms were included in order to locate the peak more precisely than in previous experiments. As Fig. 3 illustrates, there was a peak at 250 ms rather than 125 ms. However, this peak occurred only for vowel pairs shifting toward $/I/_{7-}$; in pairs shifting toward $/i/_{1-}$, the levels of performance were approximately equal with 125- or 250-ms delays. One account of these results is based on two assumptions: (a) that the decrement in performance at delays shorter than 250 ms occurs because of backward recognition masking (Massaro, 1972);

and (b) that a masked vowel tends to sound closer to the neutral vowel /ə/, and therefore closer to /ɪ/, on the present vowel continuum (see Fig. 4). The latter assumption suggests that masking might only impair performance for vowel pairs shifting toward /ɪ/, a prediction that is consistent with the present results and with the model of vowel discrimination to be advanced below.

III. GENERAL DISCUSSION

The present study has examined vowel discrimination in an AX comparison task, and has focused on two aspects of performance as a function of the interstimulus delay: the advantage of between- over within-category performance, and the advantage for vowel pairs in which the second vowel is closer to the /i/ end of the vowel continuum than the first vowel. In experiment 1a, subjects listened to AX pairs with between- and within-category trials presented in separate trial blocks. It was expected that this blocking procedure would permit subjects to know whether or not phonetic labels would be helpful within a particular block, thereby allowing them to use stable phonetic labels in between-category blocks no matter what the delay interval. However, contrary to this expectation, the absence of a contrast type \times delay interaction at the last four delays suggests that subjects used a delay-dependent source of information on approximately

the same proportion of trials in between- and within-category trial blocks. The most likely delay-dependent source of information is a comparison of the auditory memory traces of the two vowels, or alternatively, of phonetic labels formed with reference to this auditory comparison (Repp *et al.*, 1979).

Although the blocking manipulation did not affect the performance functions across delays, it did appear to affect the overall magnitude of the between-category superiority. Further, the difference between the blocked and randomized procedures was primarily in the level of performance in between-category trials. Thus, the factors related to phonetic categories in these experiments affected the level of discrimination performance in a manner that did not interact with interstimulus delay, a similar effect of delay obtained for both between- and within-category vowel comparisons.

On the other hand, the order of the vowels in a pair did affect the performance function over delay intervals (Figs. 2 and 3). Discriminability decreased much more across delay intervals when the first vowel in the pair was closer to the /i/ end of the continuum than was the second vowel. This effect was observed both in the between-category contrasts (in experiment 1) and in within-category contrasts (in experiment 2).

The present results help to restrict the types of models that might be proposed to account for speech discrimination in an AX comparison task. The results clearly rule out a simple, two-component model described in the Introduction, in which discrimination performance on between-category trials depends on a stable phonetic memory, and performance on within-category trials depends on a transient auditory memory. That type of model predicts a delay \times contrast type interaction that was not obtained, despite methodological steps taken to encourage subjects' separate use of these two forms of speech information (experiment 1). Although it is necessary to adopt a model of performance that includes a stable form of phonetic memory in order to account for the main effect of contrast type (i.e., the between-category advantage), the obtained effect of delay in both between- and within-category performance indicates that this vowel category information does not take precedence over more transient forms of memory on between-category trials. Instead, the speech memory system seems to be characterized by some combination of stable phonetic and transient auditory information, regardless of the categories of the two stimuli.

This suggestion is consistent with a mathematical model of the auditory memory process proposed by Durlach and Braida (1969) and applied to speech perception by Ades (1977) and Macmillan (1985). Durlach and Braida (1969), basing their approach upon auditory intensity perception, proposed that there are two memory modes whose output can be combined: (a) a context-coding mode and (b) a trace mode. The context-coding mode results in categorical information that is relatively stable over time, whereas the trace mode results in detailed sensory information that decays over time. Both modes are active for any stimulus set, although the mode that predominates depends upon such factors as the perceptual range of the stimulus set and, in dis-

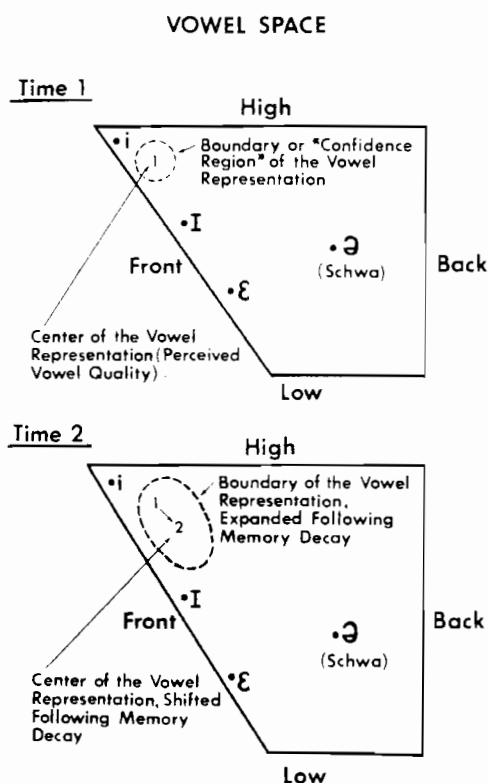


FIG. 4. A model of decay in the memory representation of a vowel, illustrated for the vowel /i/. At time 1 (e.g., 500 ms after the vowel is presented) the boundary of the confidence region for the representation is narrowly defined around the perceived vowel quality within the vowel space. By time 2 (e.g., 2000 ms) the confidence region has expanded or "smeared." Presumably because there is little room for expansion toward the edges of the vowel space, the perceived quality also shifts toward /ə/, thus remaining approximately centered in the confidence region.

crimination tasks, the time between stimuli. Ades (1977) and Macmillan (1985) proposed that the differences in the perceptual functions obtained for different sets of speech sounds are a consequence of their perceptual ranges (e.g., that consonants are perceived more categorically than vowels because they cover a smaller range that does not permit the trace mode to operate well). This approach does not identify the source of the between-category advantage, which could result either from a heightened sensitivity at the boundary or from practice in the context-coding mode. However, in either case, the categorical information from the context mode would be weighed or combined with information from the trace mode. As the interstimulus delay increased in the AX discrimination task for vowels, the information from the trace mode would decrease. Moreover, for vowels the context-coding mode would not be sufficiently predominant to prevent decline in between-category performance levels across delays.

Not surprisingly, the present data are insufficient to support a complete model of the speech discrimination process. For example, the between-category advantage could result from greater sensitivity at the category boundaries, or from learned categorization as Macmillan (1985) suggests. However, the data do bear on the plausibility of various classes of models. One result that may be especially important in the assessment of models of speech discrimination is the obtained interaction of delay \times vowel order. This interaction rules out delay-independent explanations of the vowel order effect, such as those discussed by Repp *et al.* (1979). Instead, it may support a model in which there is a shift in the vowel quality of the first vowel in an AX pair during the interstimulus delay.

The most restricted form of model incorporating a vowel quality shift would be one in which the memory representation for the first vowel shifts toward the /I/ end of the continuum during the delay interval. This would decrease the difficulty of contrasts in which the second vowel is closer to /i/₁ than the first, because the delay would broaden the distance or "gap" between the perceived vowel qualities of the two stimuli. Conversely, in pairs in which the second vowel was nearer to /i/₇ than the first, the shift in memory representation during the delay would bring the two vowels perceptually closer together and therefore would increase the difficulty of the discrimination.

A broader and somewhat more speculative form of this model is that the memory representation of the first vowel in a pair shifts toward the neutral vowel /ə/ at the center of the vowel space during the delay (see Ladefoged, 1982, for a discussion of the vowel space, and Repp *et al.*, 1979, and Sawusch *et al.*, 1980, for suggestions similar to the present one about the role of the neutral vowel /ə/ in relation to other vowels). This shift toward /ə/ implies a shift toward the /I/ end of the spectrum for the present stimuli, producing the delay \times vowel order interaction. For vowels at other locations within the vowel space, of course, the model predicts that similar delay \times vowel order interactions should be obtained. Although there are no available data permitting a direct examination of that prediction, Repp *et al.* (1979) did observe that the vowel order effect was reversed at the ex-

treme /ε/ end of their vowel continuum, just as it should have been if the memory representation shifts toward /ə/ in each case.

It is not sufficient, though, for a model of vowel memory decay to focus only on the delay \times vowel order interaction. It must also account for the general decrement in performance across delays that was obtained to some degree in both vowel orders and in both between- and within-category discrimination. One specific model that is capable of incorporating both the delay effect and its interaction with vowel order is illustrated in Fig. 4. According to this model, each vowel is represented initially as a small, bounded area (i.e., a particular vowel quality) within the two-dimensional vowel space. Memory decay is represented in the figure as an expansion of the bounded area corresponding to the memory. This representation assumes that the subject cannot recall the precise location of the vowel within the vowel space, but that he or she recalls its location within certain limits (the bounds of the area). When the second vowel in the pair is presented, the subject must determine whether or not the vowels originated in the same location within the vowel space. The longer the interstimulus delay, the greater would be the uncertainty as to the exact identity of the first vowel, and the more difficult the comparison would be. Presumably, the greater the overlap between the two vowel representations after the second vowel is presented, the less likely it is that the subject will detect a difference between the vowels. The amount of overlap increases as the representation of the first vowel expands across delays. This model also produces an interesting, additional prediction: Assuming that subjects respond "same" when they cannot detect a sufficiently large difference between vowels (i.e., when the memory representations of the two vowels overlap to a substantial degree within the vowel space), there should be more "same" responses at long delays. An examination of this aspect of the data from both experiments strongly confirmed this prediction.

The aspect of the model illustrated in Fig. 4 accounting for the vowel order effect is that the expansion of a vowel's bounded representation during the delay interval would not always be symmetrical. When a vowel is close to an edge of the vowel space little expansion is possible toward this edge, and the perceived quality presumably shifts so as to remain centered within the expansion region. Because the vowel /i/₁ occupies a cardinal location or "corner" of the vowel space, its perceived vowel quality (depicted in the figure) shifts away from this corner, toward both /I/ and schwa. Vowels closer to /i/₇ would shift in memory away from the front edge of the vowel space, but because these vowels are not near either the high or the low edge, the vowel space does not constrain them to shift either toward or away from /i/₁. The strong downward shift for vowels near /i/₁, in the absence of a strong shift of the other vowels toward either end of the /i/-/I/ continuum, would be the primary basis of the vowel order effect.

The model illustrated in Fig. 4 may provide a savings in the theoretical system needed to account for vowel discrimination, inasmuch as the delay effect and the vowel order effect can be explained with a single form of memory. How-

ever, there are several topics for future research that could make the model more complete. First, research is needed to specify the basis of the between-category advantage. It could be due to a heightened sensitivity of the perceptual mechanism in the region of the category boundary (see Ades, 1977) or it could be based on the subject's familiarity with the entire range of English vowels (see Macmillan, 1985). Second, there are interesting parallels to the vowel order asymmetry in research with other paradigms that could be explored. Sawusch *et al.* (1980) obtained anchor effects from both /i/ and /I/ vowels that altered the identification and discrimination functions along an /i/-/I/ continuum, but a signal detection analysis suggested that only the /i/ anchor altered the subjects' sensitivity, whereas the /I/ anchor caused shifts in criterion. Crowder and Repp (1984) found contrast effects from an /i/ vowel, but not an /I/ vowel, presented 350 ms before a target vowel. Finally, in a vowel adaptation study, Morse *et al.* (1976) obtained adaptation along an /i/-/I/-/ε/ continuum using /i/ or /ε/ as the adapting stimuli, but not /I/. Future research must determine the relationship between these various paradigms and must also examine vowel order effects in other regions of the vowel space.

- Ades, A. E. (1977). "Vowels, consonants, speech and nonspeech," *Psychol. Rev.* **84**, 524-530.
 Crowder, R. G. (1982). "Decay of auditory memory in vowel discrimina-

- tion," *J. Exp. Psychol.: Learning, Memory, Cognition* **8**, 153-162.
 Crowder, R. G., and Repp, B. H. (1984). "Single formant contrast in vowel identification," *Percept. Psychophys.* **35**, 372-378.
 Durlach, N. I., and Braida, L. D. (1969). "Intensity perception. I. Preliminary theory of intensity resolution," *J. Acoust. Soc. Am.* **46**, 372-383.
 Gillman, C. B., Wilson, D., Hirsch, J., and Morse, P. A. (1975). "A computer-based system for preparing auditory stimuli," University of Wisconsin Infant Development Laboratory Status Report **1**, 285-296.
 Kaplan, H. L., Macmillan, N. A., and Creelman, C. D. (1978). "Tables of *d'* for variable-standard discrimination paradigms," *Behav. Res. Methods Instrum.* **10**, 796-813.
 Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer," *J. Acoust. Soc. Am.* **67**, 971-995.
 Ladefoged, P. (1982). *A Course in Phonetics* (Harcourt Brace Jovanovich, New York).
 Macmillan, N. A. (1985). "On the existence, nature, and plasticity of perceptual categories," Paper presented at the 11th annual meeting of the Society for Philosophy and Psychology, Toronto, May 1985.
 Macmillan, N. A., and Kaplan, H. L. (in press). "Detection theory analysis of group data: Estimating sensitivity for average hit and false alarm rates," *Psychol. Bull.*
 Massaro, D. W. (1972). "Preperceptual images, processing time and perceptual units in auditory perception," *Psychol. Rev.* **79**, 124-145.
 Morse, P. A., Kass, J. E., and Turkienicz, R. (1976). "Selective adaptation of vowels," *Percept. Psychophys.* **27**, 421-434.
 Pisoni, D. B. (1973). "Auditory and phonetic memory codes in the discrimination of consonants and vowels," *Percept. Psychophys.* **13**, 253-260.
 Repp, B. H., Healy, A. F., and Crowder, R. G. (1979). "Categories and context in the perception of isolated steady-state vowels," *J. Exp. Psychol.: Human Perception Performance* **5**, 129-145.
 Sawusch, J. R., Nusbaum, H. C., and Schwab, E. C. (1980). "Contextual effects in vowel perception II: Evidence for two processing mechanisms," *Percept. Psychophys.* **27**, 421-434.