

Auditory and phonetic memory codes in the discrimination of consonants and vowels*

DAVID B. PISONI

Indiana University, Bloomington, Indiana 47401

Recognition memory for consonants and vowels selected from within and between phonetic categories was examined in a delayed comparison discrimination task. Accuracy of discrimination for synthetic vowels selected from both within and between categories was inversely related to the magnitude of the comparison interval. In contrast, discrimination of synthetic stop consonants remained relatively stable both within and between categories. The results indicate that differences in discrimination between consonants and vowels are primarily due to the differential availability of auditory short-term memory for the acoustic cues distinguishing these two classes of speech sounds. The findings provide evidence for distinct auditory and phonetic memory codes in speech perception.

Current theories of speech perception suggest that the perception of speech sounds may involve processes that are in some way basically different from the processes involved in the perception of other sounds (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). A large body of experimental work indicates that when listeners are exposed to certain classes of speech sounds, their ability to identify and discriminate between them on an auditory basis is limited to a large degree by their linguistic knowledge. Differences in the perception of certain classes of speech sounds have led investigators to propose a "special" speech perception mode to characterize the way these phonetic segments are heard (Liberman, 1970). Other results have suggested that a "special" perceptual mechanism may exist for the processing of speech sounds (Studdert-Kennedy & Shankweiler, 1970; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970).

One of the findings that has been cited as evidence for a special speech perception mode is the difference in perception between synthetic stop consonants and steady-state vowels. Stop consonants have been found to be perceived in a categorical mode, unlike other auditory stimuli. Discrimination is limited by absolute identification. Listeners are able to discriminate stimuli drawn from *different* phonetic categories but cannot discriminate stimuli drawn from the *same* phonetic

category, even though the acoustic difference between stimuli is comparable.

For example, Liberman, Harris, Hoffman, and Griffith (1957) found that synthetic speech stimuli that varied in acoustically equal steps through the range sufficient to produce the initial stop consonants /b/, /d/, and /g/ were perceived as members of discrete categories. When listeners were required to discriminate pairs of these stimuli, they were able to discriminate stimuli drawn from different phonetic categories but could not discriminate stimuli drawn from the same phonetic category. The obtained discrimination functions were not monotonic with changes in the physical scale, but showed marked discontinuities at points along the continuum that were correlated with changes in identification.

On the other hand, steady-state vowels have been found to be perceived continuously, much like nonspeech sounds. Listeners are able to discriminate many more differences than would be predicted on the basis of absolute identification. Fry, Abramson, Eimas, and Liberman (1962) reported that synthetic vowel stimuli varying in acoustically equal steps through the range of /i/, /ε/, and /æ/ were perceived in a continuous mode. The discrimination functions did not yield discontinuities along the continuum that were related to changes in identification but were relatively constant across the entire continuum. In addition, they reported that listeners could perceive many more intraphonemic differences for the vowel series than for the consonant series. Similar differences between stop consonants and steady-state vowels have been reported more recently by Stevens, Liberman, Studdert-Kennedy, and Ohman (1969) and by Pisoni (1971). The results of these studies have led investigators to propose two different modes for the perception of speech stimuli: a categorical or phonetic mode and a continuous or auditory mode¹ (Liberman et al. 1967; Studdert-Kennedy, 1973; Studdert-Kennedy, Shankweiler, & Pisoni, 1972).

*This paper is based on a portion of a thesis submitted to the University of Michigan in partial fulfillment of the requirements for the PhD degree. I am very grateful to Dr. Franklin S. Cooper and Professor Alvin M. Liberman for making the unique facilities of Haskins Laboratories available to me for preparation of the stimulus materials and for their interest in this work. I am also indebted to Professor Irwin Pollack and Professor Michael Studdert-Kennedy for their help and advice. This research was supported in part by a grant from NICHD to Haskins Laboratories, an NSF grant to Irwin Pollack, and a Rackham Prize Fellowship from the Graduate School of the University of Michigan. A shorter version of this paper was presented at the meetings of the Acoustical Society of America, Denver, Colorado, October 1971.

Although the distinction between categorical and continuous modes of perception has played an important role in theoretical discussions of speech perception, the differences between these two modes of perception are not well understood. Recently, Fujisaki and Kawashima (1969, 1970) proposed a model of the perceptual processes involved in speech discrimination that considers the separate contributions of phonetic and auditory short-term memory. They suggest that the categorical-continuous distinction in speech perception may be related to the degree to which separate auditory and phonetic memory components are employed in the decision process during discrimination.

According to Fujisaki and Kawashima's model, when a listener discriminates two *different* phonetic types, he bases this decision on the derived phonetic properties and features of the auditory stimulus as represented in some type of *phonetic short-term memory*. The listener determines whether the two stimuli have been identified as belonging to the same or different phonetic categories with a binary decision. For example, are the two stimuli the "same" phonetic segment or "different" phonetic segments? Following this strategy, a listener's performance in a discrimination task should be completely predictable from his performance on an identification task and should approach the ideal case of categorical perception. The listener can discriminate two stimuli only to the extent that he can identify the stimuli as being different phonetic segments (Liberman et al, 1957).

Categorical perception, which appears to be unique to certain kinds of speech sounds, may be contrasted with continuous perception where the listener is able to discriminate between two identical phonetic types, or allophones. In order to make a correct decision, the listener must rely on some stored auditory information about the acoustic parameters of the stimuli as represented in *auditory short-term memory*, because he has categorized both stimuli as the same phonetic segment. The listener now makes a comparative judgment rather than an absolute judgment, attending to the specific acoustic properties of the two stimuli. As a consequence, discrimination performance is independent of identification.

The purpose of the present experiment was to test the hypothesis that consonants and vowels differ in the degree to which distinct auditory and phonetic memory codes are employed in discrimination. If categorical perception is related to the differential use of auditory short-term memory in discrimination, it should be possible to demonstrate this by the use of several procedures which have not been employed previously in speech perception experiments. One such procedure is a delayed comparison recognition memory task. In this task, two stimuli are presented with a varying comparison interval between them. The S's task is to indicate whether the two stimuli were the same or different. We can produce a relative preponderance of

comparative and absolute judgments and, in turn, a differential reliance on auditory and phonetic memory codes by selecting pairs of stimuli to be discriminated from either "between" phonetic categories or "within" phonetic categories. By examining discrimination performance over a range of comparison intervals, the temporal course of recognition memory for vowel and consonant stimuli may be assessed.

METHOD

Experimental Design

The design of the present experiment involved the manipulation of three independent variables: stimulus conditions (four levels, two classes of vowels, and two classes of consonants); stimulus comparisons (two levels, within phonetic category comparisons and between phonetic category comparisons); and delay interval (0.0, .25, .50, 1.0, and 2.0 sec). Ss were assigned to one of four different stimulus conditions, with the stimulus variable distributed over all levels of the other two independent variables.

Subjects

Sixteen undergraduate students at the University of Michigan served as Ss in the present experiment. The Ss were obtained from the paid S pool at the Mental Health Research Institute. All Ss were right-handed native speakers of English and reported no history of a hearing disorder or speech impediment. Ss were paid for their services at the rate of \$2/h. None of the Ss had ever heard any synthetic speech stimuli before the experiment.

Description of Stimuli

The following four sets of synthetic speech stimuli were prepared, and they correspond to the different stimulus conditions in the experiment. All of the stimuli were digitized, and their wave forms were stored on the Pulse Code Modulation System at Haskins Laboratories (Cooper & Mattingly, 1969).

Voiced Stop Consonants (/bæ/ - /dæ/)

A set of voiced stop consonant-vowel stimuli were synthesized on the Haskins Laboratories parallel resonance synthesizer. The set consisted of seven three-formant syllables that were 300 msec in duration. The final 220 msec of each stimulus was a steady-state vowel appropriate for an American English /æ/, with the first three formants fixed at 743, 1,620, and 2,862 Hz, respectively. During the initial 40 msec, a period of closure voicing was simulated on the synthesizer by a low-amplitude F_1 at 150 Hz. This period of prevoicing appropriate for the voiced stops was followed immediately by a 40-msec transitional period, during which the first three formants moved toward the steady-state frequencies of the vowel. The experimental variable was the starting frequencies of the second and third formant transitions. Stimulus 1 had second and third formant frequencies, beginning at 1,232 and 2,180 Hz. For successive stimuli in the series, the F_2 and F_3 starting frequency increased in approximately equal steps from 1,232 to 1,695 Hz and from 2,180 to 3,195 Hz, respectively. The change in both F_2 and F_3 transitions from Stimulus 1 to Stimulus 7 has been shown to be the major acoustic cue for distinguishing place of production between the syllables /bæ/ and /dæ/ (Liberman et al, 1967). The fundamental frequency was set at 120 Hz for the entire duration of the syllable.

Bilabial Stop Consonants (/ba/-/pa/)

A set of seven three-formant bilabial stop consonants was also produced on the Haskins synthesizer. The stimuli varied in 10-msec steps along the voice onset time (VOT) continuum from 0 through +60 msec, which distinguishes /ba/ and /pa/. VOT has been defined as the interval between the release of the articulators and the onset of laryngeal pulsing or voicing. Synthesizer control parameter values for these stimuli were similar to those employed by Lisker and Abramson (1967). Each of the seven stimuli had a duration of 300 msec. The final 250 msec of the CV syllable was a steady-state vowel appropriate for an American English /a/. The frequencies of the first three formants were fixed at 769, 1,232, and 2,525 Hz, respectively. During the initial 50-msec transitional period, the first three formants moved upward toward the steady-state frequencies of the vowel. For each successive stimulus in the set, the amplitude of F_1 was "cutback" (amplitude reduced) and the excitation source was switched from buzz (periodic) to hiss (aperiodic) in 10-msec steps. Lisker and Abramson (1967) have showed that changes in amplitude in the lower frequency region and type of excitation characterize the voicing and aspiration differences between /b/ and /p/ in English.

Long Steady-State Vowels (/i/-/I/)

Seven 300-msec steady-state vowels were synthesized on the vocal tract analogue synthesizer at the Research Laboratory of Electronics, M.I.T. The stimuli were arranged so that the first three formants varied in approximately equal logarithmic steps through the English vowels /i/ and /I/. The fourth and fifth formants were fixed at 3,500 and 4,500 Hz, respectively. The formant frequency values employed here were identical to those provided by Stevens et al (1969) in their study of vowel perception.

Short Steady-State Vowels (/i/-/I/)

An additional set of seven vowel stimuli were also produced on the vocal tract analogue synthesizer at M.I.T. These seven stimuli were identical to the long vowels described above, except that their duration was reduced to 50 msec. The 50-msec vowel condition was included because both Fujisaki and Kawashima (1970) and Pisoni (1971) had reported that vowels of very brief duration could be perceived categorically.

Experimental Materials

The experimental materials were produced under computer control. Two types of tests were prepared for each stimulus condition: an identification test and a delayed comparison recognition memory test.

Absolute Identification Tests

Two different 70-item identification tests were prepared for each of the four stimulus conditions. Each identification test contained 10 different randomizations of an entire series of the seven stimuli in the set. The stimuli were recorded singly with a 4-sec interval between presentations and an 8-sec interval after every 10 presentations.

Delayed Comparison Recognition Memory Tests

Four different delayed comparison recognition memory tests were constructed for each of the four stimulus conditions. Each test tape contained two separate replications of 50 trials. A set of 50 trials consisted of 10 basic test pairs at each of five delay intervals. The 10 test pairs were constructed in the following manner. Stimuli 1, 3, 5, and 7 were selected from each of the four original sets of seven stimuli. The four stimuli were arranged

into three AB pairs (i.e., 1 with 3, 3 with 5, and 5 with 7). These three pairs appeared in two permutations (i.e., AB and BA), producing six "different" AB pairs at a given delay interval. Each of the four stimuli was also paired with itself once, resulting in four "same" AA pairs at each delay interval. Each set of 50 trials appeared in a different random arrangement in each test.

The stimuli were recorded in pairs, with a 5-sec interval between successive trials. The delay intervals (0, .25, .50, 1.0, 2.0 sec) were arranged automatically under computer control. A 100-msec 1,000-Hz tone was recorded 750 msec before the onset of the first stimulus in each pair as a ready signal.

Procedure

The experiment was conducted in an anechoic chamber located in the Phonetics Laboratory at the University of Michigan. The experimental tapes were reproduced on an Ampex 351-2 tape recorder and were presented binaurally through Telephonics (TDR-39) matched and calibrated headphones. The gain of the tape recorder was adjusted to give a voltage across the earphones equivalent to 75 dB SPL re 0.0002 dynes/cm² for a 1,000-Hz calibration tone. Measurements were made on a Ballantine VTVM (Model 300) before the presentation of each experimental tape.

At the beginning of the experiment, Ss were told that this was an experiment dealing with speech perception and that the sounds they would hear were made by a computer to approximate human speech. The instructions for the identification test were identical to those used in previous speech perception experiments. The Ss were told that the stimuli would be presented individually and that they were required to identify each stimulus as belonging into one of two categories, depending on the particular stimulus condition employed (e.g., /b/ or /d/, /b/ or /p/, /i/ or /I/).

For the delayed comparison recognition memory task, Ss were told that they would hear two stimuli separated by a varying interval on each trial and that their task was to decide whether the two stimuli were the "same" or "different." They were told that approximately half of all the pairs were the same and half of the pairs were different. Ss were encouraged to guess if they were not sure of a judgment. Judgments for both identification and discrimination tests were recorded in prepared booklets containing IBM test sheets for later analyses.

The Ss were tested for 1 h/day on 2 consecutive days. Each session began with a 70-item identification test, which was followed by four 100-trial delayed comparison recognition memory tests. The order of presentation for the delayed comparison tests was reversed on the second day.

RESULTS

Absolute Identification

The average identification function for each of the four stimulus conditions is shown in Fig. 1. Each point is based on 80 judgments summed over the four Ss in each condition.

Inspection of this figure reveals that Ss partitioned each of the stimulus continua into two relatively distinct phonetic segments. The average percent correct discrimination functions summed over all delay intervals are also plotted on the corresponding identification functions in Fig. 1. Two aspects of these data are of interest. First, discrimination performance is better between phonetic categories than within phonetic categories for every stimulus condition. Second,

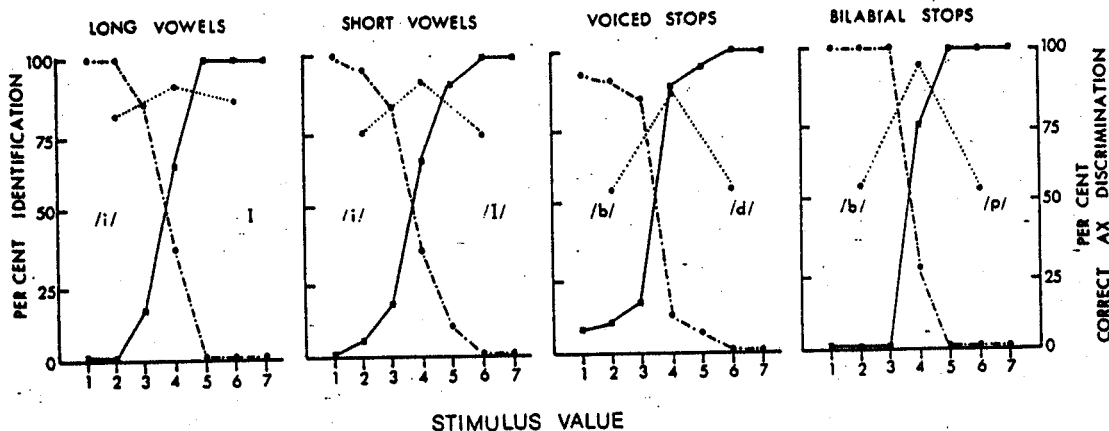


Fig. 1. Average identification functions for each of the four stimulus conditions, with discrimination functions averaged over all delay intervals superimposed on the corresponding identification functions.

within-category discrimination is close to chance for both consonant conditions but well above chance for both vowel conditions. Within-category comparisons for the long-vowel condition also appear to be more discriminable than they are for the short-vowel condition.

Delayed Comparison Recognition Memory

The major stimulus comparisons under consideration are also shown in Fig. 1. Within-category discrimination scores were obtained at each delay interval by averaging the judgments for Stimulus Comparisons 1-3 and 5-7 in each stimulus condition. Between-category discrimination scores were obtained at each delay interval for judgments of Stimulus Comparisons 3-5 in each stimulus condition. Two separate scoring procedures were used to assess recognition memory performance. The first procedure examined only responses to trials in which the two stimuli to be

discriminated were different, i.e., $P("D" | D)$. The second procedure employed a d' measure, which considered responses to both same and different trials. The d' measure was employed in order to account for possible response biases which might enter into the same-different task.

$P("D" | D)$ Discrimination Scores

Discrimination probabilities were obtained from conditions in which the pairs of stimuli were different. Figure 2 shows the average probability of a "different" response when the stimuli were different as a function of delay interval. Each point is based on 32 judgments per S at each delay interval. Filled circles represent between phonetic category comparisons, whereas open circles represent within phonetic category comparisons.

The $P("D" | D)$ scores were analyzed by means of a three-factor analysis of variance for mixed designs. F ratios were evaluated against the corresponding mean

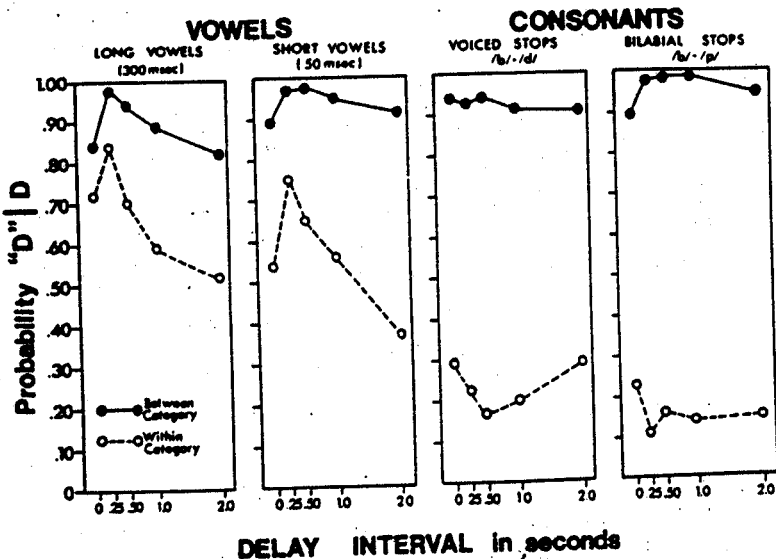


Fig. 2. Average probability of a "different" response when the stimuli within a pair were different [$P("D" | D)$] as a function of delay interval for each stimulus condition for within and between phonetic categories.

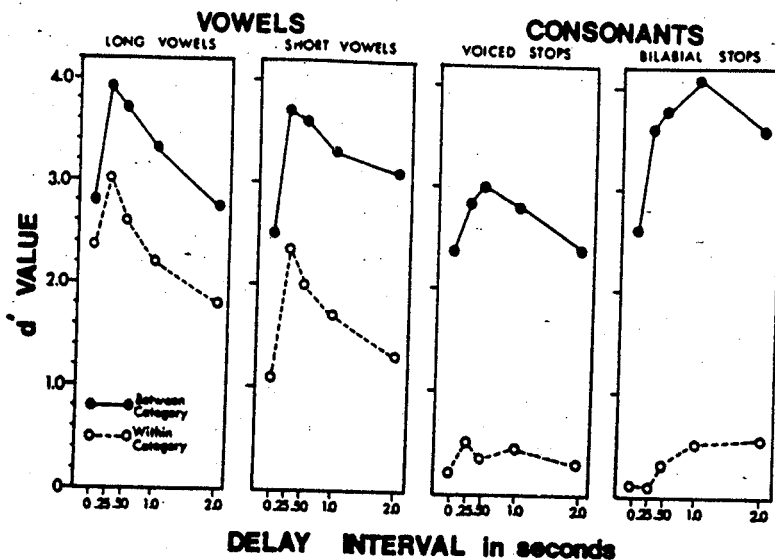


Fig. 3. Average d' scores for within- and between-category comparisons as a function of delay interval for each stimulus condition.

square term including Ss. The main effects of stimulus condition (i.e., vowels vs consonants) [$F(3,12) = 16.47$, $p < .001$], delay interval [$F(4,48) = 5.58$, $p < .005$], and stimulus comparison (i.e., between vs within) [$F(1,12) = 238.35$, $p < .001$] were all significant. Second-order interactions of Stimulus Condition by Delay Interval [$F(12,48) = 2.77$, $p < .01$] and Stimulus Condition by Stimulus Comparison [$F(3,12) = 16.37$, $p < .001$] were also significant.

The overall $P("D" | D)$ scores for between-category comparisons are quite high and relatively stable across the delay intervals for three of the four stimulus groups. One exception appears to be the long-vowel condition, where there is a slight decline in the between-category discrimination scores as the delay interval increases. However, Newman-Keuls tests on the differences among between-category means for each group indicated that they did not differ significantly from each other.

Inspection of the within-category scores, however, shows very marked differences in discrimination between consonants and vowels. Within-category consonant scores are quite low and do not appear to be related in any way to changes in the delay interval. Newman-Keuls tests showed that there were no significant differences for the consonant means across delay intervals. On the other hand, the within-category vowel scores are much higher and are systematically related to changes in the delay interval. For both vowel stimulus conditions, within-category discrimination is maximum at .25 sec and then decreases with increases in the delay interval. Newman-Keuls tests established that the .25-sec interval was significantly different from the other delay intervals for both vowel conditions.

d' Discrimination Scores

To separate the effects of recognition memory and possible response bias on the observed judgments, a d'

score was computed. False alarm rates were obtained from trials on which Ss responded "different" when the pairs of stimuli were the same, i.e., $P("D" | S)$. Since these false alarm rates were available for each stimulus at each delay interval, it was possible to obtain a d' score and use this as a reliable measure of recognition memory that was independent of response bias. The d' value has been used in this manner as a measure of trace strength in recognition memory (Wickelgren, 1966; Massaro, 1970). Figure 3 shows the average d' scores for within- and between-category comparisons as a function of delay interval for each stimulus condition. Better discrimination accuracy is shown by the higher d' levels.

The results presented in Fig. 3 indicate that discrimination accuracy decreases for both within- and between-category comparisons as the delay interval increases beyond .25 sec. The effect also appears to be greater for the vowel conditions than for the consonants. Another three-factor analysis of variance was applied to the d' scores. This analysis revealed essentially the same results as the analysis of variance performed on the $P("D" | D)$ scores. However, one important difference appeared between the two analyses. The Stimulus Condition by Delay Interval interaction, which was significant in the first analysis with the $P("D" | D)$ scores, did not reach significance with the d' scores.

It is clear from the results of both scoring procedures that increases in the delay interval affect vowel discrimination accuracy much more than consonant discrimination. Moreover, the differences that obtain appear to be most pronounced for the within-category vowel comparisons.

DISCUSSION

The overall results of the present experiment strongly support the claim that differences in discrimination between consonants and vowels are related in some way

to the differential use of auditory short-term memory for these two classes of speech sounds. Furthermore, the results provide some insight into the temporal course of recognition memory in speech discrimination.

The major findings of this experiment can be illustrated by considering again the two types of discrimination trials employed: within phonetic category comparisons and between phonetic category comparisons. It was argued at the outset that discrimination judgments for these comparisons may be considered to represent two types of memory components: short-term auditory memory and short-term phonetic memory, respectively. The results shown in Fig. 2, based on the $P("D" | D)$ scores, indicated that only within-category vowel discrimination was related to changes in the comparison delay interval. Within-category consonant discrimination was not only quite poor overall, but also showed no relationship to changes in delay interval. These differences indicate that while auditory short-term memory facilitates vowel discrimination within categories, it contributes little to within-category consonant discrimination. On the other hand, the $P("D" | D)$ scores for between-category discrimination were quite high and appeared to be unaffected by changes in the delay interval. Phonetic memory appears to be quite reliable for both vowels and consonants.

The d' analysis, which considered the response bias inherent in the same-different task, revealed several additional findings which were obscured by the $P("D" | D)$ scoring analysis. First, there appears to be some tendency even for the between-category d' scores to decrease with increases in the delay interval. This was most noticeable for the long-vowel condition. Secondly, the absence of a significant Stimulus Condition by Delay Interval interaction with the d' scores suggests that consonant discrimination is also affected to some extent by changes in the delay interval. The conclusion that the course of short-term memory is different for vowels than for consonants is, perhaps, a premature oversimplification. The differences that obtain are entirely due to within-category comparisons, which we have argued are based on auditory short-term memory. For these pairs, large and consistent differences do obtain for consonants and vowels in both the $P("D" | D)$ and d' analyses.

While it may be concluded from these results that auditory short-term memory for consonants is different from auditory short-term memory for vowels, an explanation for the differences is clearly warranted. It is apparent that the acoustic information needed to discriminate two physically different but phonetically identical consonants is somehow not available for use in discrimination, even at very short delay intervals. However, it is not clear why this acoustic information is unavailable for use in discrimination.

One possible explanation has been suggested by Fujisaki and Kawashima (1970). Their explanation may be called the cue-duration hypothesis. According to this

hypothesis, the major factor responsible for the inferior auditory short-term memory with consonants is the duration of the critical information in the signal. The acoustic cues that distinguish stop consonants (i.e., formant transitions) are relatively short in duration and presumably cannot be stored well in memory. On the other hand, the acoustic cues that distinguish vowels (i.e., formant frequencies) extend the entire duration of the stimulus. Although Fujisaki and Kawashima (1970) and Pisoni (1971) have reported that short vowels are perceived more categorically than long vowels, their findings must be considered in light of the present experiment. If the cue-duration hypothesis were correct, we would expect, in the present experiment, to find the short vowels to be more similar to the stop consonants than the long steady-state vowels. The recognition memory data of this experiment argue against this prediction. Short vowels of 50 msec duration behave almost identically to the long vowels. Although discrimination is somewhat lower overall in the short-vowel condition, the effect of the delay interval is still present, especially for within-category comparisons.

Since categorical perception can be defined only by examining the relationship between identification and discrimination (see Studdert-Kennedy et al, 1970, for further discussion), short vowels may show a tendency towards categoricity under certain experimental conditions. However, the type of categorical perception previously found with short vowels may, in fact, be qualitatively different from that found with stop consonants. For example, both Fujisaki and Kawashima (1970) and Pisoni (1971) used an ABX discrimination procedure, which may have prevented a direct comparison between successive stimuli and forced their listeners to use an encoded categorization in discrimination. Thus, the categorical perception observed with short vowels by Fujisaki and Kawashima and by Pisoni may be attributed to the use of the ABX discrimination procedure rather than being inherent in the perceptual processes underlying vowel perception.

Several other findings deserve some comment. First, there was a noticeable decrease in discrimination performance at the 0.0-sec delay interval for both the $P("D" | D)$ and d' scores. It is possible that this decrease represents an interruption of processing at an early stage of perceptual analysis. Using a backward masking paradigm, Massaro (1970) has found that perceptual processing of a brief tone is terminated if a masking tone follows the test tone at very short delay intervals. Massaro suggests that the masking tone interrupts a readout of information from a preperceptual sensory store, which holds the image of a stimulus until the features needed for identification can be extracted. It is interesting to note that the interruption at the 0.0-sec delay interval is more apparent for the within-category vowel comparisons than for the within-category consonant comparisons. Moreover, the interruption is less marked overall for the between-category comparisons than for the within-category comparisons.

The direction of the interruption is that anticipated if the subsequent signal terminated the processing of auditory features, assuming that phonetic processing was more nearly completed.

Secondly, discrimination accuracy for within-category vowel comparisons reaches a maximum at the .25-sec delay interval. Successive increases in the delay interval beyond .25 sec produce a steady decline in discrimination performance. This value may represent the processing time necessary for auditory recognition (see Massaro, 1972). If perceptual processing is completed within .25 sec, then the acoustic information may still be relatively salient for use in subsequent discrimination.

The most important difference between categorically and continuously perceived stimuli appears to rest on the level of within-category discrimination performance. Although the discrimination functions for vowels show a peak at the boundary between phonetic segments, the level of within-category discrimination is still well above chance. These findings suggest that categorization for vowels is not absolute and that purely auditory information is still available for use in discrimination judgments. Since prosodic information is carried almost entirely by vowels, auditory information could be available for use in vowel discrimination.

The situation with respect to the consonants is somewhat peculiar. Although the discrimination functions for the consonants show comparable peaks at phonetic boundaries, the level of within-category discrimination is very close to chance, indicating that categorization is absolute and binding. The extraction of relevant features from the acoustic signal for consonant recognition may preclude the further use of auditory information for nonphonetic judgments. It seems reasonable to conclude from these results that consonant recognition may be mediated by some specialized decoder which is tuned to specific phonetic features (Liberman et al, 1967). However, it is not possible to determine from the outcome of the present experiment whether this mediation involves an overlap with some articulatory-motor component or whether it is due to some inherent limitation of the auditory system.

The differences obtained in this study between consonants and vowels are also related to the findings reported by Crowder (1971) on recency effects in immediate memory. Crowder found that for lists of auditorily presented synthetic stop-vowel syllables, a recency effect is observed in ordered recall if the syllables in the list contrast only on vowels. However, the recency effect is curiously absent if the syllables contrast only on stop consonants. His findings are exactly what we would expect if auditory information about consonants were unavailable for use in later recall as a consequence of phonetic classification. The vowel data indicate that some type of auditory information is still available in memory for later use in immediate recall.

In summary, the results of this experiment suggest

that the differences between consonant and vowel discrimination are primarily due to the differential availability of auditory short-term memory for the acoustic cues which distinguish these two classes of speech sounds. Vowel discrimination within categories was considerably better than consonant discrimination within categories. As a result, we can conclude that auditory short-term memory for the acoustic properties of vowels is better than auditory short-term memory for the acoustic properties of consonants.

REFERENCES

- Cooper, F. S., & Mattingly, I. G. Computer-controlled PCM system for investigation of dichotic speech perception. Status Report on Speech Research (SR-17/18), Haskins Laboratories, New York, 1969, 17-21.
- Crowder, R. G. The sound of vowels and consonants in immediate memory. *Journal of Verbal Learning & Verbal Behavior*, 1971, 10, 587-596.
- Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. The identification and discrimination of synthetic vowels. *Language & Speech*, 1962, 5, 171-189.
- Fujisaki, H., & Kawashima, T. On the modes and mechanisms of speech perception. Annual Report of the Engineering Research Institute, Vol. 28, Faculty of Engineering, University of Tokyo, Tokyo, 1969, 67-73.
- Fujisaki, H., & Kawashima, T. Some experiments on speech perception and a model for the perceptual mechanism. Annual Report of the Engineering Research Institute, Vol. 29, Faculty of Engineering, University of Tokyo, Tokyo, 1970, 207-214.
- Liberman, A. M. Some characteristics of perception in the speech mode. In D. A. Hamburg (Ed.), *Perception and its disorders; Proceedings of A.R.N.M.D.* Baltimore: Williams & Wilkins, 1970. Pp. 238-254.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, 74, 431-461.
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme-boundaries. *Journal of Experimental Psychology*, 1957, 54, 358-368.
- Lisker, L., & Abramson, A. S. Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences, Prague, September 1967.
- Massaro, D. W. Retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, 1970, 83, 32-39.
- Massaro, D. W. Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, 1972, 79, 124-145.
- Pisoni, D. B. On the nature of categorical perception of speech sounds. Doctoral thesis, University of Michigan, August 1971.
- Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., & Ohman, S. E. G. Cross-language study of vowel perception. *Language & Speech*, 1969, 12, 1-23.
- Studdert-Kennedy, M. The perception of speech. In T. A. Sebeok (Ed.), *Current trends in linguistics*. Vol. XII. The Hague: Mouton, 1973.
- Studdert-Kennedy, M., Liberman, A. M., Harris, K., & Cooper, F. S. The motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 1970, 77, 234-249.
- Studdert-Kennedy, M., & Shankweiler, D. P. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 1970, 48, 579-594.
- Studdert-Kennedy, M., Shankweiler, D. P., & Pisoni, D. B.

Auditory and phonetic processes in speech perception: Evidence from a dichotic study. *Cognitive Psychology*, 1972, 3, 455-466.

Wickelgren, W. A. Phonemic similarity and interference in short-term memory for single letters. *Journal of Experimental Psychology*, 1966, 71, 396-404.

modes of perception reflect processing of information at two distinct stages of perceptual analysis. The auditory stage refers to the analysis of the acoustic wave form into a set of time-varying psychological dimensions (pitch, loudness, timbre), whereas the phonetic stage refers to the transformation of auditory dimensions into abstract phonetic features.

NOTE

1. In this paper, it is assumed that auditory and phonetic

(Received for publication July 31, 1972;
revision received November 17, 1972.)