Categories and Context in the Perception of Isolated Steady-State Vowels

bruno H. Repp, Alice F. Healy+, and Robert G. Crowder+

ABSTRACT

The noncategorical perception of isolated vowels has been attributed to the availability of auditory memory in discrimination. Using vowels from an /i/-/I/-/ɛ/ continuum in an AX (same-different) task and comparing the results with predictions derived from a separate identification test, we demonstrate that vowels are perceived more nearly categorically if auditory memory is degraded by extending the interstimulus interval (ISI) and/or filling it with irrelevant vowel sounds. In a second experiment, we use a similar paradigm but, in addition to presenting a separate identification test, elicit labeling responses to the AX pairs used in the discrimination task. We find that AX labeling responses predict discrimination performance quite well, regardless of whether auditory memory is available or not, whereas the predictions from the separate identification test are more poorly matched by the obtained data. The AX labeling responses show large contrast effects (both proactive and retroactive) that are greatly reduced when there is interference with auditory memory. We conclude from the presence of these context effects that vowels are not perceived categorically (that is, absolutely). However, it seems that by properly taking these context effects into account, discrimination performance can be quite accurately predicted from labeling data, suggesting that vowel discrimination, just like consonant discrimination, may be mediated by phonetic labels.

INTRODUCTION

One of the best-known findings of speech perception research is the phenomenon of categorical perception. Its experimental demonstration requires a continuum of synthetic speech sounds spanning at least two different

+Also Yale University.

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRSG Grant RR05596 to the Haskins Laboratories. We are greatly indebted to Virginia Walters for conducting the experiments and tabulating the results. Helpful comments on earlier versions of this manuscript were obtained from William Estes, Alvin Liberman, Virginia Mann, Richard Pastore and David Pisoni.

[HASKINS LABORATORIES: Status Report on Speech Research SR-54 (1978)]
phonemic categories. If listeners are asked to identify and discriminate the stimuli from such a continuum, two typical findings emerge: the perceptual boundary between the two categories is relatively abrupt, and discrimination of stimuli drawn from within the same category is near chance, while it is good across the phoneme boundary. Perception is said to be categorical if the discrimination results can be predicted from the identification results, under the assumption that discrimination is based exclusively on the phonetic category labels. This pattern of results has been typical for a number of speech sounds, particularly the stop consonants in initial position (Liberman, Harris, Hoffman and Griffith, 1957; Studdert-Kennedy, Liberman, Harris and Cooper, 1970; Pisoni, 1971).

Of all speech sounds, isolated vowels are least likely to be perceived categorically. Not only are the category boundaries less distinct on a vowel continuum, but, more significantly, discrimination of acoustically different stimuli from the same category is usually much better than chance and exceeds the predictions derived from labeling data (Fry, Abramson, Elmas and Liberman, 1962; Stevens, Liberman, Studdert-Kennedy and Uhman, 1969; Pisoni, 1971). Thus, the perception of isolated vowels has been called "continuous," in contrast to the categorical perception of stop consonants. However, even vowels often show better discrimination across category boundaries than within categories (Pisoni, 1971).

The distinction between categorical and continuous perception has been attributed to the differential availability of auditory memory traces for different kinds of stimuli (Darwin and Baddeley, 1974; Pisoni, 1971, 1973, 1975; Fujisaki and Kawashima, 1969, 1970). The assumption is that an accessible auditory memory representation facilitates continuous perception by providing an alternate basis for discrimination beyond phonemic categories. We will refer to this view as the dual-coding model. It assumes that speech sounds are discriminated by comparing both auditory and phonetic memory codes. The distinctive cues for stop consonants are of very brief duration and are followed (or preceded, if in final position) by a vowel that might interfere with the already fragile auditory memory representation of the relevant cues. Isolated steady-state vowels, on the other hand, are of much longer duration and contain distinctive information from onset to offset. Consequently, their auditory memory representation will be much more robust and can be utilized more easily in a discrimination task.

This explanation has found support in several experiments by Crowder (1971, 1973a). He showed that three standard phenomena of verbal short-term memory—the recency effect, the suffix effect, and the modality effect—are all obtained for lists of syllables differing only in their vowels, but are absent in lists of syllables differing only in their initial stop consonants. Since all three effects mentioned are assumed to reflect the existence of a relatively unanalyzed auditory memory (pre-categorical acoustic store), the conclusion was that stop consonants do not leave any significant auditory trace but vowels do. The relative strength of the auditory memory trace for a stimulus seems to be a direct function of its acoustic similarity to other stimuli to be remembered (Darwin and Baddeley, 1974).
Investigations of the role of auditory memory in categorical perception have taken two approaches. One line of research tries to make the perception of stop consonants less categorical by inducing listeners to make better use of their weak auditory memory traces. The other approach attempts to make the perception of vowels more categorical by interfering with their auditory memory representations, so that the listeners have to rely increasingly on category labels in discriminating the stimuli. The first type of experiment involves listener training and the use of sensitive discrimination paradigms; it has yielded some positive results suggesting that, under favorable conditions, listeners can make effective use of their auditory memory representations of stop consonants (Carney, Widin, and Viemeister, 1977; Ganong, 1977; Pisoni and Lazarus, 1974; Sachs and Grant, 1976; Samuel, 1977). The other approach—that of making the perception of vowels more categorical—is primarily due to Pisoni (1971, 1973, 1975), whose work provides the background for ours.

Prior to Pisoni's studies, there was already some evidence that vowels are perceived more categorically when they occur in phonetic (word) context (Stevens, 1968; Sachs, 1969). Thus, one way to decrease the strength of auditory memory is to change the structure of the stimuli themselves. Pisoni (1971, 1973), Fujisaki and Kawashima (1969, 1970), and Sachs (1969) took a related approach by decreasing the duration of isolated vowel stimuli. This made perception more categorical, but not completely so. Corresponding reductions in the stimulus suffix effect for shortened vowels were reported by Crowder (1973b) and Hall and Blumstein (1977). However, even very short vowels apparently permit distinguishable auditory traces to be established; discrimination is usually better than would be expected if only phonetic labels were used to discriminate the stimuli.

An alternative procedure is to leave the stimuli unchanged and to attempt to tamper directly with auditory memory. There are two methods that have been used to degrade auditory memory, decay and interference. The first technique was used by Pisoni (1971, 1973) and, more recently, by Cutting, Hosner, and Foard (1976). These authors systematically increased the interval between the vowel stimuli in an AX (same-different) discrimination task from 0 to 2 sec. The result was a decrease in performance that was taken to mean that auditory memory decayed over time. Whether this decay was complete after 2 sec is not clear from their data.

The interference technique was employed by Pisoni (1975). He used an ABX discrimination paradigm in which the "X" vowel was immediately preceded or followed by one of four irrelevant signals: a noise burst, a pure tone, a dissimilar vowel, or a similar vowel. Performance decreased in all conditions, but more so when the interfering stimulus followed the "X" vowel than when it preceded it. An acoustically similar vowel seemed to produce the most interference. These interference effects are similar to the phonetic context effects of Stevens (1968) and Sachs (1969), except for the fact the coarticulation with natural phonetic context modifies the acoustic properties of vowels, while unrelated interfering stimuli do not. We decided to employ both decay and interference in our experiments.
EXPERIMENT I

Pisoni's results are consistent with a role for auditory memory in vowel perception, but since Pisoni did not attempt to predict discrimination from identification performance, we do not know whether procedures designed to eliminate auditory memory would produce completely categorical perception for vowels. This is the hypothesis that we wished to test in our first experiment. To manipulate the availability of auditory memory, we varied the amount of time elapsing between members of AX (same-different) discrimination pairs and, orthogonally, whether there was an interpolated speech sound or not during this interval. To measure the degree of categorical perception, we relied on a comparison of identification and discrimination performance: if accuracy of AX discrimination can be predicted from phonetic labeling (identification), provided that both are better than chance, we may conclude that perception is categorical. The interesting possibility is that, although discrimination shows a surplus over identification when auditory memory is present, vowel perception will be categorical when auditory memory has been removed.

Method

Subjects. Sixteen college-age adults volunteered to participate as paid volunteers. All were native speakers of English and had little previous experience with synthetic speech.

Stimuli. The stimuli were modeled after Pisoni's (1971) vowel continuum. The formant frequencies given in Pisoni (1971, Table 2, p. 12) were realized as closely as possible (within a few Hz) on the O slime synthesizer at Haskins Laboratories. The complete set included 13 stimuli whose first formant increased, and whose second and third formants decreased in frequency in approximately equal logarithmic steps from stimulus 1 to stimulus 13. These frequencies are shown in Table 1. The fourth and fifth formants were hardware-fixed. All stimuli were 240 msec in duration and had a fundamental frequency that fell linearly from 125 to 80 Hz.

From these 13 stimuli, three pairs of vowels were selected which, according to Pisoni's data, were identified predominantly as /i/, /I/, and /ɛ/, respectively. In the notation of Table 1, they were stimuli 1 and 3 (/i/), 6 and 8 (/I/), and 11 and 13 (/ɛ/). Note that the acoustic distance between the vowels was greater between categories (three steps) than within (two steps); this was a deliberate attempt to avoid the fairly broad category boundary regions evident in Pisoni's data and to maximize between-category discriminability. Within-category discriminability of the stimuli selected had been about 80 percent correct in the AX test used by Pisoni (1971). An additional vowel-like sound was constructed by combining the first formant of stimulus 1 with the higher formants of stimulus 13 (See Table 1). This stimulus sounded approximately like the vowel /y/ (as in Swedish fyra) and was used for interference only.

Six experimental tapes were recorded using the Haskins Laboratories Pulse Code Modulation (PCM) System. Two were identification tapes; the remaining four were AX discrimination tapes. One identification tape contained 60 vowels (the six stimuli repeated ten times) in random order, with ISIs of 4
sec. The second identification tape contained the same 60 vowels in the same random sequence, but each vowel was preceded by the irrelevant /y/ sound. The interval between the /y/ and the following vowel was 120 msec; that between the vowel and the next /y/ was 4 sec. The /y/ precursor was included as a control to see whether it affected in any way the labeling of the following vowel.

The four discrimination tapes all contained the same random sequence of 80 vowel pairs consisting of five replications of sixteen different combinations of the six basic stimuli. The sixteen combinations included six identical pairs (1-1, 3-3, 6-6, 8-8, 11-11, 13-13), six within-category pairs (1-3, 3-1, 6-8, 8-6, 11-13, 13-11) and four between-category pairs (3-6, 6-3, 8-11, 11-8). The four tapes differed in the nature of the interval between the two vowels in a pair. In the "short unfilled" condition, it was 480 msec of silence. In the "long unfilled" condition, it was 1,920 msec of silence. In the "short filled" condition, the /y/ sound (240 msec in duration), preceded and followed by 120 msec of silence, intervened between the two vowels. In the "long filled" condition, five repetitions of the /y/ sound intervened; they were preceded, separated, and followed by 120 msec of silence. Thus, the temporal separation between the vowels in a pair was the same in corresponding filled and unfilled conditions. The interval between successive pairs was 4 sec throughout.

Procedure. The 16 subjects were divided into two equal groups. One group received the two identification tests prior to the discrimination tests; the other group was assigned the reverse order. All subjects listened to the regular identification series before the one with /y/ preceding each vowel. The sequence of the four discrimination conditions was counterbalanced across subjects in four Latin squares.

In the identification task, the answer sheets listed the words "beet," "bit," and "bet" for each trial. The subjects were instructed to circle the word whose vowel resembled most the stimulus presented. The /y/-sound was to be ignored, if present. In the discrimination tasks, the response sheet contained the letters "s" (same) and "d" (different) for each trial, and the subjects were instructed to circle the appropriate letter for each vowel pair. It was emphasized to respond "same" only when the two vowels were exactly the same. The different conditions were explained and announced in advance. Any occurrences of the /y/ sound were to be ignored.

The subjects were run in small groups in a single session of about one hour. The tapes were played back on a Sony TC-630 tape recorder with its own loudspeakers. Intensity was set at a comfortable level.

Results

Identification. The identification results, averaged across subjects and the two identification tests, are summarized in Table 2. The results of the two identification tests were combined, since an analysis of variance showed that the irrelevant /y/-precursor did not significantly affect identification performance, F(1,14) < 1. Table 2 shows that stimuli in the /i/ and /e/ categories were identified fairly consistently (89 percent correct or better), but many confusions occurred with stimuli in the /i/ category, especially
stimulus 8. This is in agreement with Pisoni's (1971) data: the /I/ category is the least stable of the three, probably because the relatively long stimulus durations employed here were least appropriate for this category which, in natural speech, is associated with shorter durations than /I/ and /ɛ/ (Peterson and Leniute, 1960). The statistical analysis indicated that confusions were somewhat more frequent when the identification tests were presented at the end of a session, $F(1,14) = 7.3$, $p = .017$; this may have been a result of fatigue.

**Discrimination.** The results of the discrimination tests are summarized in Figure 1. For each of the four experimental conditions, percentages of correct responses are shown as a function of stimulus pair. Each data point is plotted halfway between the locations of the two stimuli to be discriminated and represents the average of four percentages: those of "different" responses to the two stimulus orders of the given pair, and those of "same" responses to each member of the pair when paired with itself.

It is evident from Figure 1 that both manipulations of the ISI (delay and filling) affected discrimination performance. The subjects made more errors when the interval was long than when it was short, $F(1,14) = 56.4$, $p < .0001$, and when the interval was filled with irrelevant vowel sounds than when it was unfilled, $F(1,14) = 40.0$, $p < .0001$. The interaction of these two factors was not significant, $F(1,14) < 1$, nor was there any significant interaction of these two effects with vowel pairs, as is confirmed by the parallel functions in Figure 1. Newman-Keuls tests between individual conditions confirmed that both delay by itself and the presence of an interpolated stimulus by itself significantly reduced discrimination performance.

As expected, discrimination performance was poorest in the long filled condition. In order to find out whether the scores in this condition approached those expected under the categorical perception model, we predicted the percentages of correct responses in the discrimination test from the identification responses, under the assumption that discrimination is based solely on phonetic labels (Pollack and Pisoni, 1971). (A second assumption, often not stated explicitly, is that the labeling probabilities are the same in the identification and discrimination tasks; we will have reason to question this assumption later in this paper.) These predictions, averaged across subjects, are indicated in Figure 1 by the dashed function at the bottom. They are quite close to the scores obtained in the long filled condition, particularly for the first three stimulus pairs; the discrimination performance for the last two vowel pairs is somewhat better than predicted. Separate Chi-square tests were performed on the data from each subject, summing observed and expected frequencies of correct responses and errors across stimulus pairs, thereby enabling tests with only one degree of freedom. When comparing expected scores to those obtained in the long filled condition, 11 of the 16 subjects exceeded the expectations, but the difference was significant at the .05 level in only two cases. In addition, another subject performed significantly poorer than expected. Thus, overall performance in the long filled condition was not significantly different from that predicted by the categorical-perception model.
A separate analysis of "hits" ("different" responses to pairs of nonidentical stimuli) revealed an unexpected stimulus order effect, $F(1,14) = 15.3$, $p < .01$, which is shown in Table 3, averaged across conditions and subjects. This effect interacted with position on the continuum, $F(4,56) = 10.4$, $p < .0001$. In four stimulus pairs, the subjects gave substantially more "different" responses when the stimulus with the higher position on the continuum preceded the stimulus with the lower position, but the effect was reversed for the last stimulus pair: there were more "different" responses to the order 11-13 than to 13-11. The stimulus order effect was somewhat more pronounced when the ISI was unfilled than when it was filled, leading to a significant interaction between stimulus order and filling, $F(1,14) = 7.8$, $p = .014$, as well as an interaction between stimulus pairs, stimulus order, and filling, $F(4,56) = 4.1$, $p = .006$. We shall consider this effect in greater detail below.

Discussion

The results of Experiment I support the hypothesis that isolated steady-state vowels will be perceived categorically when there is interference with auditory memory. Discrimination performance in the long filled condition was close to that predicted under strict categorical perception assumptions. It seems likely that the combined effects of decay and interference in this condition impaired the auditory trace of the first stimulus in a pair to a degree that made an auditory comparison with the second stimulus rather difficult. Consequently, the listeners probably relied on phonetic memory codes in the most difficult condition, whereas, in the easier conditions, phonetic memory was supplemented by varying amounts of auditory memory.

Our results show that auditory memory is vulnerable to both decay and interference. The fact that performance in the long unfilled condition was better than in either of the filled conditions suggests that the auditory memory for the first stimulus in an AX pair took longer than 2 sec to decay, given that no interfering sounds followed.

The parallelism of the discrimination functions in Figure 1 is somewhat surprising. One might have expected that between-category comparisons, which -- according to the dual-coding model -- need not rely much on auditory memory, would be less affected by interference than within-category comparisons, which presumably rely more on auditory memory. (Of course, the distinction between between- and within-category comparisons is not an absolute one, considering the high frequency of confusions -- see Table 2.) In other words, the twin peaks in the discrimination function might have been expected to be more pronounced in the long filled condition but flatter in the short unfilled condition, where performance approached the ceiling. It would be implausible to assume that phonetic memory was affected by our manipulations of the ISI. However, the results can be explained by the unequal spacing of the stimuli. It will be recalled that there was a larger acoustic distance between categories than within. Thus, the contribution of auditory memory to between-category comparisons was increased, and this may explain why the peaks in the discrimination function remained pronounced even in the short unfilled condition. According to this interpretation, the peaks in the long filled condition reflect the higher phonetic discriminability of between-
category comparisons, whereas the similar peaks in the short unfilled condition reflect the higher auditory discriminability of these same comparisons.

This confounding of auditory distance with phonetic distance was an unintended consequence of our attempt to choose stimuli that were maximally representative of their respective categories. Now it could be argued that our listeners did not use phonetic categories at all but made discriminations exclusively on the basis of auditory stimulus codes -- a hypothesis that would be congenial to several recent discussions of categorical perception (Carney et al., 1977; Macmillan, Kaplan and Creelman, 1977). The peaks in the discrimination functions, it might be argued, represent simply the superior auditory discriminability of the between-category comparisons (stimulus pairs 3-6 and 8-11); the remaining peaks in the long filled condition reflect residual auditory memory for these larger stimulus differences, and their agreement with the predictions (Figure 1) is purely coincidental. Most likely, this view could be rejected by showing that, if even more severe interference with auditory memory is introduced, performance does not deteriorate further but remains at the level of the present long filled condition, and thus in accordance with the predictions based on phonetic labeling. Alternatively, an experiment with equally spaced stimuli might be conducted, in order to unconfound auditory and phonetic distinctiveness. We chose the latter course in our second experiment, which was designed to provide considerably more detailed data on the relationship between labeling and discrimination.

**EXPERIMENT II**

Experiment II employed a 13-member vowel continuum (Table 1), in which the stimuli were separated by nearly equal logarithmic steps. There was only one interference condition, corresponding to the long filled condition of Experiment I, and a short unfilled condition. As in the previous experiment, an identification test was included in order to predict discrimination performance and thus to test whether perception was categorical in the long filled condition.

However, the present experiment included an important new feature. In addition to obtaining discrimination responses to the AX pairs in the short unfilled and long filled conditions, we also asked the subjects, in two separate conditions, to give phonetic labels to the stimuli in exactly the same AX pairs. This provided us with information about the subjects' choice of labels as a function of the surrounding stimulus context, and with a new, and probably more appropriate, set of predictions to be compared to actual discrimination performance. The reason we expected these new "in-context" predictions to be more appropriate than those derived from a single-item identification test is the well-known fact that vowel identification is affected by the surrounding context, usually in the form of contrast (Pry et al., 1962; Eimas, 1963; Lindner, 1966; Thompson and Hollien, 1970; Ainsworth, 1974; Kanaamori, Kasuya, Arai and Kido, 1971). By taking such contrast effects into account, we expected to obtain a more accurate estimate of the probabilities of the various labels that the subjects may have covertly applied in the discrimination task, and thus a more accurate estimate of the degree to which
discrimination responses might have been based on such labels. Apart from this comparison, we were interested in the contrast effects themselves as an object of study: how large they would be; whether they would occur in both directions in an AX pair (proactive and retroactive contrast); and whether they would be affected by the interfering sounds in the long filled condition.

We should emphasize at this point that we were interested in separating two questions that often have been treated in the past as the single issue of categorical perception: whether perception is absolute, and whether discrimination is based on phonetic labels. The original definition of categorical perception, as put forth by Liberman et al. (1957) and Studdert-Kennedy et al. (1970), implies absolute, that is, context-independent perception. If the labeling of a stimulus depends on the preceding or following stimuli, as frequently seems to be the case with vowels, perception is by definition noncategorical. However, the listeners may nevertheless use these context-dependent categories in a discrimination task. Thus, it may be necessary to distinguish two kinds of noncategorical perception: one in which discrimination is based on a few discrete categories, and one in which perception is truly continuous -- that is, not mediated by categorization.

To illustrate these distinctions -- categorical, category-based noncategorical, and continuous perception -- consider the two methods of prediction employed in Experiment II. One set of predictions is derived from the labeling responses in a single-item identification test. These predictions provide an unambiguous test of categorical perception. In order for the discrimination results (in the long filled condition) to match these predictions, it is necessary not only that the subjects base their discrimination responses on the same labels as in the identification task, but also that the probabilities of applying these labels be the same in the two situations. Since the stimulus contexts are quite different in the single-item test and in the AX test, this equality of labeling probabilities necessarily requires that the process of categorization be context-independent. Experiment I suggested that this might be true in the long filled condition. Experiment II provided a much more stringent test of this hypothesis, since it included a direct comparison of labeling behavior in different contexts. It is essential to realize that the equality of the labeling probabilities in the single-item and AX tests is a necessary condition for the perception of the vowels to be called categorical (see Studdert-Kennedy et al., 1970).

Consider now the case in which the labeling probabilities are not the same in these two tests. This situation will hold almost certainly in the unfilled short condition, where contrast effects are expected; it may also turn out to apply to the long filled condition. It is here that the second set of predictions comes in -- predictions derived from the labeling probabilities in AX context. Context dependence implies that this set of predictions will be different from that derived from the single-item identification test. Moreover, there are separate sets of AX predictions for the long filled and short unfilled conditions, and they may also be different from each other, to the extent that the labeling probabilities vary between these two conditions. However, if these in-context predictions match the discrimination results obtained in corresponding conditions, it may be argued that discrimination is mediated by (context-dependent) category labels, despite the fact that percep-
tion is not categorical. Only if discrimination performance matches neither set of predictions, could we be convinced that perception is truly continuous.

Thus, we distinguish three different possibilities: (1) perception is truly categorical, and both sets of predictions (those derived from single-item identification and those derived from AX labeling) coincide with the discrimination results; (2) only the in-context predictions (derived from AX labeling) are matched by the discrimination results, suggesting that discrimination may be mediated by discrete categories; (3) the discrimination results match neither set of predictions, and perception is truly continuous.1

Method

Subjects. Sixteen new volunteers participated. They were Yale undergraduates who received course credit for their participation.

Stimuli. The vowel continuum included all 13 stimuli listed in Table 1, a subset of which had been used in Experiment I. Three experimental tapes were prepared. The single-item identification tape contained a random sequence of 130 stimuli (10 repetitions of each of the 13 stimuli) with ISIs of 3 sec. Each of the other two tapes contained five different random sequences of 35 vowel pairs consisting of each stimulus paired with itself (13 pairs) and with every other stimulus two steps removed on the continuum, in both stimulus orders (22 pairs). One of the discrimination tapes had a short (300 msec) unfilled interval between the stimuli in each AX pair; the other tape had a long (1,920 msec) interval, filled with five repetitions of the /y/ sound, exactly as in the corresponding condition of Experiment I. The interpair interval was 4 sec, and blocks of 35 pairs were separated by an extra 4 sec.

Procedure. All subjects first took the single-item identification test. As in Experiment I, they circled "beet," "bit," or "bet" on an answer sheet. This task was followed by the two AX tapes presented twice with different instructions. Under discrimination instructions, the subjects circled "s" (same) or "d" (different) on the answer sheet, as in Experiment I. Under labeling instructions, the subjects circled "beet," "bit," or "bet" for each of the two vowels in a pair. (On the answer sheet, the three response

1To the best of our knowledge, we are the first actually to compute in-context predictions in a categorical-perception task, although Lane (1965) suggested the idea long ago. Several of the earlier studies on vowel perception (Eimas, 1963; Fry et al., 1962; Fujisaki and Kawashima, 1969, 1970) obtained labeling responses to the precise stimulus sequences used in discrimination, but none of these studies made the predictions conditional on context. Instead, all labeling responses were lumped together, thus averaging out all context effects. Most likely, this accounts for the large discrepancies between predicted and obtained discrimination performance, particularly in the often-cited study by Fry et al. (see Lane, 1965). It is not clear why these authors used the discrimination sequences to collect labeling responses in the first place.
alternatives appeared twice side by side.) The subjects were instructed to listen to both vowels before responding. The sequence of the discrimination and labeling conditions was counterbalanced across subjects, and so was the sequence of the short unfilled and long filled conditions within each instruction condition.

Results

Simple Identification. The results of the single-item identification test are summarized in Figure 2. The percentages of responses in the three categories, /i/, /I/, and /ɛ/, are shown as a function of stimulus location along the continuum. As in Experiment I, and in agreement with Pisoni's (1971) results, the stimuli were less consistently assigned to the middle category, /I/, than to the other two categories.

As in Experiment I, we used these identification results to predict discrimination performance. The resulting predictions, averaged over subjects, are represented by the dashed function at the bottom of Figure 3. The function has two peaks, reflecting the prediction of higher discrimination performance in the category boundary regions. If vowels are categorically perceived in the absence of auditory memory, the discrimination results in the long filled condition should coincide with these predictions.

AX Discrimination. The results of the discrimination task are displayed in Figure 3 in terms of percent correct responses, derived and plotted in the same manner as in Experiment I (solid functions). Performance in the short unfilled condition was much better than in the long filled condition, as expected, $F(1, 14) = 78.8, p < .0001$. Discrimination performance also varied significantly with location on the stimulus continuum, $F(10, 140) = 7.8, p < .0001$; there was a pronounced peak in the region of stimulus 4. As in Experiment I, the short unfilled and long filled discrimination functions were strikingly parallel, as confirmed by the absence of a significant interaction between the factors of ISI and location on the continuum, $F(10, 140) = 1.4, p = .186$. This suggests that the parallelism of the discrimination functions in Experiment I was not an artifact of unequal stimulus spacing along the physical continuum.

In Experiment I, discrimination performance in the long filled condition resembled the predictions derived from single-item identification performance. However, the data of the present experiment do not support this earlier observation. Although predicted and obtained performance were close in the middle range of the continuum, the obtained scores were clearly better than predicted at the ends of the continuum, particularly at the right (/ɛ/) end. For all but three of the subjects obtained performance exceeded predicted performance in the long filled condition, and Chi-square tests on individual subjects revealed that 8 of the 16 subjects performed significantly better than predicted ($p < .05$). Even more importantly, the shape of the obtained discrimination function did not conform to the predictions. Specifically, the predicted peak in the /i/-/ɛ/ boundary region was absent, and the predicted peak in the /i/-/I/ boundary region was displaced to the left. These discrepancies could be explained by assuming that the labeling probabilities of the stimuli changed in the context of the AX pairs. Therefore, the
predictions derived from the single-item identification test were not appropriate, and, hence, perception was not truly categorical in the long filled condition. In Experiment I, the wider and unequal stimulus spacing forced similar zigzag shapes on both predicted and obtained functions, and thus prevented us from detecting any serious mismatch. We proceed now to a discussion of the AX labeling results that were expected to provide more accurate predictions of discrimination performance, since they were obtained in identical presentation contexts.

**AX Labeling**

The predictions derived from the AX labeling responses are shown in Figure 3 as the two dotted functions. We computed predicted percent correct discrimination scores, considering each pair of AX labeling responses placed in the same phonetic category as equivalent to a "same" response and each pair of responses placed in different phonetic categories as equivalent to a "different" response. If it had been true that each vowel was identified independently of its context, the predictions from the AX labeling task at both ISIs should have equalled the predictions from the single-item identification test. This was clearly not the case, not even in the long filled condition, thus providing indirect evidence for context effects in labeling.

The discrimination scores derived from the AX labeling task (the in-context predictions) were much closer to the results of the discrimination task than to the predictions from the single-item identification test. Like discrimination, AX labeling performance showed a strong effect of interference $F(1,12) = 52.6$, $p < .0001$, and of location on the continuum $F(10,120) = 13.3$, $p < .0001$. There was a small interaction between these two factors, $F(10,120) = 2.9$, $p = .003$; however, the functions for the short unfilled and long filled conditions were essentially parallel. They were also similar in shape to the functions obtained under discrimination instructions, showing only a single peak at stimulus 4.

The in-context predictions represent the discrimination performance to be expected when only the prescribed phonetic category labels are used. However, the scores actually obtained in the discrimination task, while similar in profile, significantly exceeded these expectations, $F(1,12) = 11.7$, $p = .005$. Figure 3 shows that this difference derived from the two ends of the vowel continuum, particularly the right (/ɛ/) end, while scores in the middle region were similar. This was reflected in a significant interaction of task and stimulus location, $F(10,120) = 3.7$, $p = .002$. Especially interesting is the fact that the advantage of discrimination over labeling responses was as large in the long filled condition as in the short unfilled condition, as confirmed by a nonsignificant interaction of task and ISI, $F(1,12) < 1$.

**Hits and False Alarms.** In order to understand the results in more detail, we examined separately the responses to pairs of identical and pairs of nonidentical stimuli. Figure 4 plots the percentages of "different" responses to each stimulus pair, separately for the discrimination and labeling tasks. (In the labeling task, a "different" response represents a pair of responses placed in two different phonetic categories.) The two pairs...
Figure 1: AX discrimination scores in the four conditions of Experiment I, together with scores predicted from identification responses.

Figure 2: Labeling functions for stimuli presented in the single-item identification test of Experiment II.
Figure 3: Two-step AX discrimination scores in the short unfilled and long filled conditions under discrimination and labeling instructions. The labeling results represent the "in-context" predictions of discrimination performance. Also shown are the predictions derived from the responses in the single-item identification test.

Figure 4: Percentages of hits and false alarms in AX discrimination (left-hand panel) and AX labeling (right-hand panel). Hit percentages are shown separately for the two different orders, I<J and I>J, of each pair (I,J), where I is the first and J the second stimulus.
Figure 5: Category boundaries between /i/ and /I/ (bottom) and between /I/ and /ɛ/ (top) as a function of the relative position of the nontarget stimulus, shown separately for the first and the second stimulus in a pair as the target (retroactive vs. proactive contrast). The boundaries for the stimuli in the single-item identification test are also shown ("in isolation").
TABLE 1: Formant frequencies of the vowel stimuli (in Hz).

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>$F_1$</th>
<th>$F_2$</th>
<th>$F_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>269</td>
<td>2296</td>
<td>3019</td>
</tr>
<tr>
<td>2</td>
<td>285</td>
<td>2263</td>
<td>2955</td>
</tr>
<tr>
<td>3a</td>
<td>297</td>
<td>2230</td>
<td>2912</td>
</tr>
<tr>
<td>4</td>
<td>315</td>
<td>2183</td>
<td>2829</td>
</tr>
<tr>
<td>5</td>
<td>336</td>
<td>2151</td>
<td>2769</td>
</tr>
<tr>
<td>6a</td>
<td>354</td>
<td>2105</td>
<td>2709</td>
</tr>
<tr>
<td>7</td>
<td>375</td>
<td>2075</td>
<td>2670</td>
</tr>
<tr>
<td>8a</td>
<td>397</td>
<td>2030</td>
<td>2632</td>
</tr>
<tr>
<td>9</td>
<td>420</td>
<td>2001</td>
<td>2567</td>
</tr>
<tr>
<td>10</td>
<td>442</td>
<td>1973</td>
<td>2557</td>
</tr>
<tr>
<td>11a</td>
<td>472</td>
<td>1930</td>
<td>2539</td>
</tr>
<tr>
<td>12</td>
<td>500</td>
<td>1902</td>
<td>2520</td>
</tr>
<tr>
<td>13a</td>
<td>530</td>
<td>1862</td>
<td>2484</td>
</tr>
<tr>
<td>/y/</td>
<td>269</td>
<td>1862</td>
<td>2484</td>
</tr>
</tbody>
</table>

*Stimuli used in Experiment I.

TABLE 2: Combined results of the two identification tests in Experiment I.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>/i/</th>
<th>/I/</th>
<th>/ε/</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>99</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>90</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>80</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>60</td>
<td>39</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>11</td>
<td>89</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>4</td>
<td>96</td>
</tr>
</tbody>
</table>
### TABLE 3: Stimulus order effect in Experiment I.

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Percentage of &quot;different&quot; responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>i &lt; j</td>
</tr>
<tr>
<td>1, 3</td>
<td>21</td>
</tr>
<tr>
<td>3, 6</td>
<td>93</td>
</tr>
<tr>
<td>6, 8</td>
<td>34</td>
</tr>
<tr>
<td>8, 11</td>
<td>63</td>
</tr>
<tr>
<td>11, 13</td>
<td>41</td>
</tr>
</tbody>
</table>

Note: i = first stimulus; j = second stimulus.

---

### Table 4: AX labeling task: Percentages of /i/ and /ɛ/ responses as a function of position of target stimulus in pair (first or second), relative location of nontarget stimulus (lower or higher), and ISI.

<table>
<thead>
<tr>
<th>Position of target stimulus</th>
<th>First</th>
<th>Second</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/i/</td>
<td>/ɛ/</td>
</tr>
<tr>
<td>Interval</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short unfilled</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>14</td>
<td>55</td>
</tr>
<tr>
<td>Higher</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>Long filled</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>18</td>
<td>60</td>
</tr>
<tr>
<td>Higher</td>
<td>21</td>
<td>53</td>
</tr>
</tbody>
</table>
of functions at the top of each panel represent "hits," that is, "different" responses to pairs of different stimuli, at the two ISIs. The difference between the two functions at each interstimulus interval may be ignored for the moment. The two functions at the bottom of each panel are "false alarms," that is, "different" responses to pairs of identical stimuli, at the two ISIs.

Although the percent-correct discrimination functions for the two ISIs had been quite parallel (Figure 3), this was not true for hits and false alarms considered separately. In both, the effect of interval interacted with stimulus location, in a complementary fashion: whereas hits showed strong effects of interference at the ends of the continuum only, false alarms showed large effects in the middle of the continuum only. This was true regardless of the task performed. The interaction of the interference effect with location on the continuum was significant for both hits, $F(10,120) = 5.6, p < .0001$, and false alarms $F(12,180) = 6.7, p < .0001$, and there were no interactions of these factors with task. Thus, at the long filled interval, discrimination errors in the middle of the continuum were largely due to false alarms, while errors at the ends of the continuum were largely misses. Another way of expressing this result is that, in the presence of interference, the subjects were more likely to respond "different" in the middle of the continuum than at the ends. This may have been a consequence of the general uncertainty about the middle category, /I/ (see Figure 2).

An analysis of variance showed that there were significantly more hits in the discrimination task than in the labeling task, $F(1,12) = 15.5, p = .002$; this difference derived primarily from the right end of the stimulus continuum, leading to a significant interaction of task with stimulus location for hits, $F(10,120) = 4.7, p < .0001$. On the other hand, the false-alarm rates did not differ between the two tasks (see Figure 4). Thus, the higher scores in the discrimination task were due to a higher hit rate, not a lower false-alarm rate.

**Stimulus Order Effect.** The strong stimulus order effect obtained in the discrimination task of Experiment I was replicated in the present study. Figure 4 displays two hit functions for each ISI in each task. These two functions are distinguished only by the order of the stimuli in a pair. The functions connecting squares represent the order I<J, where the first stimulus in a pair (I) had a lower position on the continuum than the second stimulus (J); the functions connecting triangles represent the reverse order, I>J. It can be seen that the majority of stimulus pairs received more "different" responses in the order I>J than in the order I<J, but this effect disappeared or was even reversed at the right end of the continuum. This pattern of results was reflected in a significant interaction of stimulus order and location, $F(10,120) = 2.9, p = .003$, together with a significant main effect of stimulus order, $F(1,12) = 8.8, p = .012$. The stimulus order effect was present in both tasks and, most interestingly, at both ISIs. In the middle of the continuum, the effect was actually increased by interference, which contributed to a significant three-way interaction involving stimulus order, location, and interference, $F(10,120) = 2.8, p = .004$. This contrasts with the results of Experiment I, where a small decrease in the stimulus order effect was observed as a function of interference. Taken together, however, the two findings justify the conclusion that the stimulus order effect was
little affected by interference.

Contrast Effects. The results of the AX labeling task offered an opportunity to investigate the degree to which the relationship between the stimuli in a pair influenced identification. Two effects were of special interest: whether the (expected) contrast effect would be stronger in one direction than in the other (proactive vs. retroactive contrast), and whether its magnitude would change as a function of interference with auditory memory.

In order to answer these questions, we first tabulated the labeling response frequencies in the three phonetic categories separately for stimuli occurring first and stimuli occurring second in pairs of different stimuli, and then examined these frequencies for one (target) stimulus contingent on the nature of the other (nontarget) stimulus in the pair. The nontarget stimulus could be either lower on the continuum (-2 steps), identical to the target, or higher on the continuum (+2 steps). For each of these three cases, perceptual boundaries between adjacent vowel categories were determined from the average data using Finney's (1971) probit algorithm. This procedure provides an estimate of the 50 percent crossover point of the labeling functions for adjacent categories, that is, of the category boundary. The two boundaries—that between /i/ and /ɪ/ and that between /ɪ/ and /ɛ/—were expressed in terms of their location on the stimulus continuum. Figure 5 shows these boundaries as a function of the relative position of the nontarget stimulus, separately for the first and second stimulus as target, and with separate panels for the short unfilled and long filled intervals. In addition, the boundaries obtained in the single-item identification test are shown (in isolation, see Figure 2). The functions on top represent /i/-/ɛ/ boundaries, while those at the bottom represent /i/-/ɪ/ boundaries. (Note that stimulus location, plotted on the abscissa in previous figures, is plotted on the ordinate in Figure 5.)

Since there were too few observations to compute reliable boundaries for individual subjects, an analysis of variance was conducted on response percentages pooled over all pairs in a given condition, with the following factors: vowel category (/i/ vs. /ɛ/; /ɪ/ responses were omitted), target stimulus (first vs. second), relative location of nontarget stimulus (higher vs. lower; identical pairs were not included in this analysis), and interference. Pairs including target stimuli 1, 2, 12, and 13 were omitted since these stimuli could not be paired with both higher and lower stimuli on the continuum. These response percentages, which formed the basis of the statistical tests, are shown in Table 4, averaged over subjects.

A contrast effect implies a shift in the category boundaries for target stimuli toward the category represented by the nontarget stimulus. In other words, if there is a contrast effect, the category boundaries for target stimuli will be shifted toward the lower end of the continuum when the nontarget stimulus has a lower position, and they will be shifted toward the higher end when the nontarget stimulus has a higher position on the continuum. This implies a positive slope for the "boundary functions" (the connected points) in Figure 5. Obviously, there were pronounced contrast effects in the short unfilled condition (left panel), but only negligible effects in the long filled condition (right panel). The overall contrast effect was significant,
F(1, 15) = 23.0, p = .0003, as well as its interaction with the interference factor, F(1, 15) = 19.8, p = .0005. In a separate analysis of the long filled condition, the contrast effect still reached significance, F(1, 15) = 5.1, p = .04, although it was obviously very small. A slope difference between the solid and dashed functions in Figure 5 would reflect a difference between proactive contrast (second stimulus as target) and retroactive contrast (first stimulus as target). It can be seen that, surprisingly, the retroactive effect was slightly stronger than the proactive effect at the unfilled short interval, although this difference turned out not to be significant.

A level difference between the solid and dashed functions in Figure 5 implies a boundary shift as a function of stimulus position in an AX pair. Such a difference can be observed in the long filled condition, F(1, 15) = 10.6, p = .006; it suggests that the perception of the vowels may have been influenced by the five interpolated /y/ sounds, although the control identification test in Experiment 1 had shown no effect of a single /y/ precursor on labeling. Five repetitions of /y/ may have been sufficient to produce selective adaptation (Morse, Kass, and Turkienicz, 1976) which, of course, is a kind of contrast effect. The fact that the category boundaries shifted toward the lower (/i/) end of the continuum is in agreement with this interpretation and with the intuitive observation that /y/ is perceptually more similar to /i/ (with which it shared the first formant) than to /ε/ (with which it shared the weaker, higher formants). Curiously, however, the data in Figure 5 (right panel) indicate that it was the vowel preceding the five /y/ stimuli (the first vowel in an AX pair) whose boundary shifted the most, not -- as one might expect in an adaptation situation -- the following vowel. Thus, these boundary shifts remain somewhat puzzling.

Apart from any differential effects, the boundaries in the AX labeling task were generally shifted towards the lower (/i/) end of the continuum, relative to the boundaries for the same stimuli in isolation. This is particularly evident when the latter are compared with the boundaries for identical AX pairs, in which the same stimulus was repeated once. It is not known what caused this shift in perception, but it obviously contributed to the discrepancy, shown in Figure 3, between the predictions derived from the AX labeling task and those derived from the single-item identification test. This finding demonstrates that stimulus context may affect the labeling probabilities even in the absence of contrast effects (that is, in pairs of identical stimuli).

Discussion

Are Vowels Perceived Categorically? The principal question of our research was whether isolated steady-state vowels would be perceived categorically when there is interference with auditory memory. Experiment 1 suggested an affirmative answer. However, the much more fine-grained analysis in Experiment II indicated that the answer depends on the exact form in which the question is asked. Recall that in the introduction to Experiment II, we distinguished among three modes of perception---categorical, category-based noncategorical and continuous. We have obtained a fairly close fit between the in-context predictions and obtained discrimination performance. (A small surplus of discrimination performance over the predictions is a common finding
even with stop consonants, and although it requires an explanation, it is not considered a major argument against categorical perception.) Such a reasonable fit between predicted and obtained discrimination functions -- without any qualifications about the nature of the identification test from which the predictions are derived -- has often been considered the sole criterion of categorical perception (see especially Macmillan et al., 1977). However, we have also found strong evidence for auditory contrast effects in vowel labeling, thereby indicating relative rather than absolute perception. These two results suggest that it is the second mode that is employed for vowels--noncategorical but category-based.

Thus, the answer to our original question depends entirely on how we choose to define categorical perception. By the predictability criterion (using the appropriate in-context labeling data) we succeeded in making vowel perception categorical when there was interference with auditory memory. In fact, even when auditory memory was intact, perception was categorical by this criterion. However, as we have pointed out earlier, categorical perception--as defined by Liberman et al. (1957)--is synonymous with absolute, context-independent perception. ("Absolute" is, incidentally, the primary dictionary definition of the word "categorical.") Therefore, any evidence indicating that labeling behavior depends on stimulus context argues against categorical perception. We obtained such evidence: in the case when the two test stimuli were close together, there were reciprocal contrast effects. When the two stimuli were separated by time and interference, reciprocal contrast effects in labeling were minimal. However, the labeling probabilities in this latter case nevertheless deviated considerably from those in the single-item identification test. This was probably due in part to contrast with the interpolated /y/ stimuli. The mere existence of phonetic category boundary shifts as a consequence of changes in stimulus arrangement indicates that the stimuli were not perceived absolutely. Also, the labeling probabilities depended on the absolute position of a stimulus in an AX pair -- the stimulus order effect. Thus, it appears that some stimuli, because of their particular acoustic structure, are perceived relative to the surrounding context, and cannot be made to be perceived absolutely by manipulating that context. Rather, any new context -- such as the /y/ vowels introduced to interfere with auditory memory -- will simply constitute a new frame of reference for the relative perception of the target stimuli.

**Vowels, consonants, and the operational definition of categorical perception.** Although we have made no direct comparisons between performance on vowels and on stop consonants, our results suggest some similarities and differences. By the predictability criterion, based on in-context identification data, vowels and stops are not likely to be very different in view of the high degree of predictability we observed here for vowels. Although we did obtain a significant discrepancy between predictions and discrimination performance for vowels, it is well known, as we observed earlier, that there is also, typically, a small discrepancy for stops. An interesting possibility, which we are presently testing, is that the discrepancy for stop consonants could be reduced considerably by basing the predictions on in-context identification rather than on single-item identification, as has been done previously. It is known that even stop consonants show small context effects (Emas, 1963), and the in-context prediction procedure would take such effects into
account. However, even though the fit between such predicted and obtained discrimination may turn out to be somewhat closer for consonants than for vowels, the fact remains that predictability is high for both kinds of stimuli. The fundamental difference between stop consonants and vowels seems to lie in their degree of susceptibility to context effects in identification. These effects seem to be larger for vowels than for consonants.

The traditional definition of categorical perception has included two aspects, absoluteness in phonetic labeling and predictability of discrimination performance from labeling performance. However, only the predictability requirement was ever directly operationalized. The other requirement—that labeling be context-independent—was generally satisfied by the wide stimulus spacing in the single-item identification test. Meeting the traditional operational definition (predictability of discrimination from single-item labeling data) does indeed indicate categorical perception, but failure to meet it is ambiguous. Lack of fit between identification and discrimination could be caused by a failure in either or both aspects of the definition. We prefer two separate tests, one for absoluteness and one for the predictability. Both tests make use of in-context labeling performance. Absoluteness is indexed by the effects on labeling of stimulus context, and the predictability measure is strictly analogous to the traditional one, except that in-context predictions are applied. Because interesting differences between vowels and stop consonants seem more likely to be found on the absoluteness test, that may turn out to be the more informative part of the new operational definition.2

The Roles of Auditory and Phonetic Processes in Vowel Discrimination. Having commented on the degree of categorical perception in vowels, we now wish to discuss the processing mechanisms that our subjects may have brought to bear on the AX discrimination task. According to the conventional logic, meeting the predictability requirement of categorical perception directly implies a process account of discrimination performance: the subject bases his response entirely on phonetic labels. However, even though our results demonstrated such predictability, there are several permissible process explanations.

The results to be explained. There are five findings that should be considered in any comprehensive process account:

2In our discussion, we have more or less ignored two other criteria for categorical perception commonly cited in the literature: the relative steepness of the labeling functions, and the presence of peaks and troughs in the discrimination function. We find these criteria less important because they are more difficult to quantify than the fit between predicted and obtained discrimination, and because they are more or less directly related to context effects in categorization. The relatively shallow slopes of the labeling functions (see Figure 2) and the irregularities in the discrimination functions in Experiment II generally support our conclusion that the perception of the vowels was not absolute.
(1) The AX discrimination performance was quite well predicted by in-context labeling performance, although there was a statistically significant difference due primarily to a discrepancy at the /ɛ/ end of the vowel continuum we used.

(2) As is usual with speech stimuli, we obtained discrimination peaks approximately at the category boundaries, although there was only a single peak in Experiment II.

(3) Discrimination was poorer after a long filled interval between AX stimuli than after a short unfilled one.

(4) There were large reciprocal contrast effects in the AX labeling task at the short unfilled interval; these were greatly reduced at the long filled interval.

(5) There were clear stimulus order effects in discrimination that were not consistently a function of interference or of delay.

The Fujisaki-Kawashima-Pisoni dual-coding model. Fujisaki and Kawashima (1969, 1970) and Pisoni (1971, 1973, 1975) have offered a process model for discrimination performance of speech stimuli. The main assumption of this model is that there are two codes that may be used to make comparisons of stimuli--phonetic and auditory memory. Whenever two stimuli cannot be distinguished by their phonetic codes, the listener is assumed to consult his auditory memory code. It follows that the differences between predicted and obtained discrimination performance, presumably even the small discrepancies obtained here with the in-context predictions, are due to the contribution of auditory memory. This model falsely predicts that in our short unfilled condition, where there should have been abundant auditory information about the first stimulus at the time of arrival of the second stimulus, the predicted-obtained discrepancy should have been considerably larger than in the long filled condition, where little auditory information should have survived. Instead, we found equally small discrepancies in the two conditions. Thus, our results strongly contradict the predictions of the dual coding model.

An all-auditory model for discrimination. One model that can deal successfully with our results is based on the assumption of a single auditory memory code for comparing the two stimuli. This model deals more or less successfully with each of the five results listed above. Deterioration of discrimination performance following a long filled interval is an obvious consequence of this model because auditory memory is assumed to deteriorate with time. The disappearance of contrast effects upon labeling with a long filled ISI is also consistent for the same reason. These contrast effects may have a sensory basis similar to that presumed to underlie brightness contrast in vision. Alternatively, contrast could be caused by the conscious strategy of giving different phonetic labels to two sounds whenever they sound different. In other words, phonetic categorization in the labeling task may be strongly influenced by the result of implicit auditory discrimination judgments.
The all-auditory model may account for discrimination peaks only on the basis of some acoustic discontinuity corresponding to category boundaries (Pastore et al., 1977). This model assumes no phonetic processing during AX discrimination, an assumption that makes it difficult to accommodate the close fit between discrimination performance and predictions based on phonetic labeling. It may be, however, that the relatively small range of the present stimulus continuum enabled the subjects to achieve relatively high resolution in labeling with only a small number of phonetic categories (see Pynn, Braida and Durlach, 1972; Ades, 1977). On the other hand, discrepancies between obtained discrimination and predictions are no problem for this model and, indeed, we obtained such a discrepancy, especially at the /ɛ/ end of our vowel continuum.

Two of our results provide some difficulty for the all-auditory model. First, performance in the long filled condition was well above chance. This means that there must have been some substantial auditory memory persisting over the long filled interval. The implication is that if we had used a more effective delay interval (or interference stimulus) the subjects either would have been left performing at chance or would have had to adopt a different processing strategy. More serious is the difficulty in accounting for the stimulus order effect on an auditory basis. This order effect was not consistently affected by time delay or interference, which do affect performance assumed to reflect auditory memory.

An all-phonetic model for discrimination. Another model that can deal successfully with our results is based on the assumption of a single phonetic code as the basis for comparisons. This model does not deny a role for auditory memory in discrimination: the discrimination responses are based entirely on phonetic distinctions, but those phonetic labels have themselves been subject to auditory influences. That is, phonetic coding occurs first on the basis of auditory information but it is only these codes that are then used for discrimination.

The close fit between discrimination and predictions based on phonetic labeling is the natural outcome of this model. According to this model there is no difference in the subject’s information processing in the two tasks, only that the labels are covert in one case and overt in the other. Any surplus discrimination over that predicted by labeling must be explained by the presence of additional covert phonetic categories used in discrimination but ineligible for the labeling task (see Chistovich and Kozhevnikov, 1970, for a similar argument); these additional categories should be equally available whatever the interference condition. Figure 3 shows that the obtained surplus occurred precisely in our stimulus continuum where there is reason to believe that an extra covert category did exist.3

3Informal evidence suggests that there may have been an additional phonetic category, /ɛ/, in the region of the /ɪ/-/ɛ/ boundary. Since /ɛ/ is not an English phoneme (the diphthong /eɪ/ occurs instead), it was not included among the response alternatives. Some subjects may have made covert use of this additional category in the discrimination task and thus may have widened the gap between predicted and obtained discrimination.

24
The occurrence of discrimination peaks located at category boundaries is another direct consequence of the all-phonetic model. The fact that only one peak was found in the discrimination task of Experiment II is no problem for the all-phonetic model since only one peak was found in the AX labeling task of Experiment II, and subjects are necessarily basing their responses on phonetic codes in that task.

In order for the all-phonetic model to account for the poor discrimination obtained at the long filled interval, hits and false alarms should be considered separately: the probability of hits (correctly saying "different" to physically nonidentical stimuli) is increased at the short unfilled interval by reciprocal contrast between the two stimuli, which is greatly reduced at the long filled interval. False alarms (incorrectly saying "different" to physically identical tokens), according to the all-phonetic model, result from inconsistency in the labeling of a given stimulus token. It is natural to expect that momentary fluctuations in the subject's state would produce increasing inconsistency as the two stimuli are more and more separated.

Another explanation of the performance level differences at the two intervals relies on the assumption that the subject waits to apply phonetic labels until both tokens have been received. Thus the first stimulus in the long filled condition will be in a state of degraded representation in auditory memory by the time the subject categorizes it. Therefore the phonetic label for this stimulus will have been based on poor information and both misses and false alarms will ensue. An internal test of this hypothesis is possible: in the long filled condition, the in-context labeling results should show steeper labeling functions for the second stimulus than for the first; there should be no such differences between the two stimuli in the short unfilled condition. The data partially bear out this hypothesis in that (a) there were essentially no differences in labeling for the two stimuli in the short unfilled condition, and (b) there were such differences in the long filled condition; however, (c) the latter differences existed only for the /I/ to /ɛ/ range of our continuum.

At present, the phonetic model provides no explanation of the stimulus order effect. However, the fact that the effect was equally large in both interference conditions suggests that a phonetic explanation may be appropriate.4

The all-phonetic model is designed to describe mechanisms resulting in the subject's "same" or "different" responses in discrimination; however, we may also expect from this hypothesis a statement on the contrast effects that were observed in in-context labeling. The form of this explanation is similar to one of the two all-auditory explanations of contrast effects: it is assumed that when two stimuli reside together in some stage of auditory processing, there are reciprocal interactions between the representations similar to those found in visual brightness contrast or other laterally

4See Appendix I.
inhibiting systems. At the short unfilled interval these processes are maximized in comparison to the long filled interval.

It might be objected that if there is a preliminary auditory stage in which contrastive effects occur, then why is not the AX discrimination response tapped directly off this contrastive process (rather than off the phonetic labels that, in turn, were influenced by the auditory contrast). Such direct use of the preliminary auditory stage would require consciousness of it, which, however, we consider to be unlikely. Conscious attention cannot be directed simultaneously to two different levels of analysis of the same message. When one level consists of a linguistically significant channel, it seems to be preferred. The Stroop color-word interference effect shows this rule operating in the visual realm; subjects cannot disregard the linguistically informative verbal channel while naming the colors of printed words. It is as if the linguistic level of analysis always dominates the nonlinguistic level; this is of obvious adaptive value for human communication. In our type of experiment, Bailey, Summerfield and Dorman (1977) have shown how difficult it is to leave the phonetic mode once subjects have begun to place phonetic interpretations on sounds at first perceived as nonspeech. The all-phonetic model for discrimination is in harmony with these considerations; phonetic mediation for purposes of discrimination is a natural and automatic consequence of inherent priority for the linguistic level of analysis when one exists in stimuli. Of course, the whole process of discrimination is fundamentally guided by auditory information processing. The information, after all, enters the system through the ears. The question at issue is whether or not there is a stage of phonetic mediation from which discrimination responses are drawn. By arguing in favor of this notion, we do not exclude the possibility that the subjects' attention may be directed to the auditory level by extended practice or special discrimination paradigms; this has been shown to be possible even with stop consonants (Ganong, 1977; Samuel, 1977). However, we find it plausible to assume that the relatively inexperienced listeners in our experiments followed a natural tendency to remain in the phonetic mode of processing.

**A mixed model for discrimination.** It is possible to combine the assumptions of the all-auditory and all-phonetic models for discrimination into a mixed model. It postulates that auditory processing dominates at the short unfilled interval and phonetic processing dominates at the long filled interval. The problems raised earlier for the all-auditory model apply with equal force to this hybrid model. Additionally, there is the unique problem associated with the hybrid model that the near identical goodness of fit between predicted and obtained discrimination levels is observed in the two conditions, which must be assumed to be purely coincidental. Similarly, the stimulus order effects were similar in the two conditions but would be explained by the hybrid model as resulting from two different mechanisms in the two conditions. Likewise, the obtained discrimination peak has to be explained in one way for the short unfilled condition and in another way for the long filled condition. Thus, despite the greater flexibility of the mixed model, it seems clearly unparsimonious compared to the others.
Conclusions

We remain in some doubt as to the detailed processing model that supports the AX discrimination of isolated vowels. However, this uncertainty should not detract from the positive conclusions permitted by our experiments:

Phonetic labeling is an excellent predictor of AX discrimination performance provided that the labels are obtained in the same context that is used in discrimination testing. This was true even under conditions presumed to be rich in auditory memory. Thus, the original version of the dual coding model for speech discrimination (the Fujisaki-Kawashima-Pisoni model) needs to be revised.

Reciprocal (proactive and retroactive) contrast effects are a major influence on phonetic labeling of vowels. It is on the basis of this evidence that we conclude that vowels are not perceived categorically. Nevertheless, it appears that vowel discrimination may be mediated by phonetic labels, possibly even to the same extent as is discrimination of stop consonants.

APPENDIX I

An interesting hypothesis was proposed by Smith* who apparently was the first to discover the stimulus order effect with vowels. She refers to the time-order error often found in studies of duration discrimination (see, for example, Jamieson and Petrusio, 1975) and links this finding with the fact that /I/ tends to be shorter than /i/ and /ɛ/ in natural speech (Peterson and Lehiste, 1960). Thus, stimulus duration provides an additional cue for distinguishing between these categories. If, as Smith assumes, the time-order error is negative, so that the first stimulus in a pair tends to be perceived as shorter and hence more /I/-like than the second, it would increase the discriminability of pairs of the type /I/-/I/ and /I/-/ɛ/ over the discriminability of the reverse order of these pairs. Since the effect would be mediated by the phonetic labels given to the stimuli, Smith's hypothesis fits well with the all-phonetic model of processing, and it predicts our results fairly well. There is a problem, however: two recent studies of duration discrimination using vowels comparable to our stimuli (Lehiste, 1976; Pisoni, 1976), have shown the time-order error to be positive, not negative. That is, the first stimulus in a pair tends to be perceived as longer than the second, probably due to the relatively short stimulus durations. In the light of this finding, Smith's hypothesis predicts just the opposite of her and our results. The hypothesis could be salvaged by assuming that, independently of the time-order error, there is a tendency to perceive /I/ as relatively longer than /i/ and /ɛ/, due to perceptual compensation when all vowels are of equal physical duration. If this is the case, a positive time-order error would tend to further increase the discrepancy in perceived duration when an /I/-like stimulus occurs first in a pair, thus enhancing discrimination. This is what

we found. This explanation assumes, however, that the subjects base their
discrimination responses on stimulus duration, not on phonetic labels.
Therefore, it is not compatible with an all-phonetic model of vowel discrimi-
nation, nor does it fit into an all-auditory framework because of the
mediating role of phonetic stimulus properties.

REFERENCES

Review 84, 524-530.
Ainsworth, W. A. (1974) The influence of prescriptive sequences on the percep-
tion of synthesized vowels. Language and Speech 17, 103-109.
Bailey, P. J., A. Q. Summerfield and M. F. Dorman. (1977) On the identifica-
tion of sine-wave analogues of certain speech sounds. Haskins
Carney, A. E., G. P. widin and N. F. Viemeister. (1977) Noncategorical per-
ception of stop consonants differing in VOT. Journal of the Acoustical
Society of America 62, 961-970.
Theory and Methods of Research on Perception of Speech, ed. by
L. A. Chistovich. (Washington, D.C.: Joint Publications Research Ser-
vice, JPRS-50423), 1-101.
Journal of Verbal Learning and Verbal Behavior 10, 587-596.
Crowder, R. G. (1973a) Representation of speech sounds in precategorical
Crowder, R. G. (1973b) Precategorical acoustic storage for vowels of short
and long duration. Perception and Psychophysics 13, 502-506.
Cutting, J. E., B. S. Rosner and C. P. Foard. (1976) Perceptual categories
for music-like sounds: Implications for theories of speech perception.
of speech. Cognitive Psychology 6, 41-60.
along speech and non-speech continua. Language and Speech 6, 206-217.
Press).
cation and discrimination of synthetic vowels. Language and Speech 5,
171-189.
perception. Annual Report of the Engineering Research Institute, Univer-
sity of Tokyo 28, 67-73.
and a model for the perceptual mechanism. Annual Report of the Engineer-
ing Research Institute, University of Tokyo 29, 207-214.
Unpublished Ph.D. Dissertation, Massachusetts Institute of Technology.
Hall, L. L and S. E. Blumstein. (1977) The effect of vowel similarity and
syllable length on acoustic memory. Perception and Psychophysics 22, 95-
99.
Jamieson, D. G. and W. M. Petrusic. (1975) Presentation order effects in
Kationsforschung 21, 102-106.