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

Bruno H. Repp
Haskins Laboratories, New Haven, Connecticut

Alice F. Healy and Robert G. Crowder
Yale University and Haskins Laboratories
New Haven, Connecticut

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

One of the best-known findings of speech perception research is the phenomenon of categorical perception. The primary empirical criterion for categorical perception is the predictability of discrimination performance from performance in an identification test, under the assumption that discrimination is based solely on the phonetic labels used in identification.[1] This criterion has been essentially met for a number of speech sounds, particularly the stop consonants in initial position (Liberman, Harris, Hoffman, & Griffith, 1957; Pisoni, 1971; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970).

[1] Two other criteria for categorical perception are often cited in the literature: relative steepness of the labeling functions and 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, which are discussed in connection with Experiment 2.

Previous research indicates that isolated vowels are least categorically perceived of all speech sounds. Discrimination performance for these stimuli greatly exceeds the predictions derived from labeling data (Fry, Abramson, Eimas, & Liberman, 1962; Pisoni, 1971; Stevens, Liberman, Studdert-Kennedy, & Öhman, 1969). If we conceive of a continuum between the ideals of categorical perception and continuous perception, we find that stop consonants are closer to the categorical ideal and vowels are closer to the continuous ideal.

The distinction between categorical perception and continuous perception has been attributed to the differential availability of auditory memory traces for different kinds of stimuli (Darwin & Baddeley, 1974; Pisoni, 1971, 1973, 1975; Fujisaki & Kawashima, Note 1, Note 2). 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 includes the assumption that speech sounds are discriminated by comparing both auditory and phonetic memory codes.[2] The distinctive cues for stop consonants are of very brief duration and often involve rapid changes in formant frequencies; their auditory memory representations seem to be extremely fragile (Crowder, 1971, 1973). 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 representations are likely to be much more robust, so that they can be utilized more easily in a discrimination task.

Investigations that have illustrated the role of auditory memory in categorical perception have taken two approaches. One has been to make the perception of stop consonants less categorical by inducing listeners to make better use of their weak auditory memory traces. This 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 ef-

fective use of their auditory memory representations of stop consonants (Carney, Widin, & Viemeister, 1977; Ganong, 1977; Pisoni & Lazarus, 1974; Sachs & Grant, 1976; Samuel, 1977). The other approach involves making 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. This approach 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, Note 3). Thus, one way to decrease the strength of auditory memory is to change the structure of the stimuli. Pisoni (1971, 1973), Fujisaki and Kawashima (Note 1), and Sachs (Note 3) took a related approach by decreasing the duration of isolated vowel stimuli. This made perception somewhat more categorical. 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 more directly with auditory memory. There are two methods that have been used to degrade auditory memory, time delay and interference. The first technique was used by Pisoni (1971, 1973) and, more recently, by Cutting, Rosner, and Foard (1976). These authors systematically increased the interval

---

[2] The dual-coding model, as originally formulated by Fujisaki and Kawashima (Note 1, Note 2), was designed to apply to the ABX discrimination paradigm only; also, it was a serial model. Pisoni (1971) implicitly extended the model to the AX paradigm. We have in mind a general dual-coding model that applies to any paradigm and may be either serial or parallel in nature. In other words, we assume only that both auditory and phonetic codes are available for comparison, but we make no assumptions about the order or contingency of auditory and phonetic comparisons in a given discrimination task.

between the vowel stimuli in a same–different (AX) discrimination task from 0 to 2 sec. The result was a decrease in performance, indicating 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. We decided to employ both time delay and interference in our experiments.

Experiment 1

Pisoni's results are consistent with a role for auditory memory in vowel discrimination. However, since Pisoni did not attempt to predict discrimination from identification performance, we did not know whether procedures designed to eliminate auditory memory would produce completely categorical perception for vowels. This was 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 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.

Table 1
Formant Frequencies of the Vowel Stimuli (in Hz)

Stimulus	F_1	F_2	F_3
1[a]	269	2,296	3,019
2	285	2,263	2,955
3[a]	297	2,230	2,912
4	315	2,183	2,829
5	336	2,151	2,769
6[a]	354	2,105	2,709
7	375	2,075	2,670
8[a]	397	2,030	2,632
9	420	2,001	2,567
10	442	1,973	2,557
11[a]	472	1,930	2,539
12	500	1,902	2,520
13[a]	530	1,862	2,484
/y/	269	1,862	2,484

[a] Stimuli used in Experiment 1.

Method

Subjects. Sixteen college-age adults participated 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 OVEIIIc synthesizer at Haskins Laboratories. The complete set included 13 stimuli in which the first formant increased in frequency and in which the 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 physical distance between the vowels was greater between categories (three steps) than within categories (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% correct in the ABX test used by Pisoni (1971). An additional vowellike sound was constructed by combining the first formant of stimulus 1 with the higher formants of stimulus 13 (cf. Table 1). This stimulus sounded approximately like the foreign vowel /y/ (as in Swedish *fyra*) and was used for interference only.

Six experimental tapes were recorded after digitizing the stimuli 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 6 stimuli repeated 10 times) in random order with interstimulus intervals 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; the interval 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 16 different combinations of the six basic stimuli. The 16 combinations included 6 identical pairs (1-1, 3-3, 6-6, 8-8, 11-11, 13-13), 6 within-category pairs (1-3, 3-1, 6-8, 8-6, 11-13, 13-11), and 4 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 tests in 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 tested in small groups in a single session of about 1 hr. The tapes were played

Table 2

Combined Results of the Two Identification Tests in Experiment 1

Stimulus	% responses		
	/i/	/I/	/ɛ/
1	99	1	0
3	90	8	2
6	9	80	11
8	1	60	39
11	0	11	89
13	0	4	96

back on a SONY TC-630 tape recorder with loudspeakers. Intensity was set at a comfortable level.

Results

Identification. The identification results, averaged across the 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 /ɛ/ categories were identified fairly consistently (89% 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 was the least stable of the three categories, probably because the relatively long stimulus durations employed were least appropriate for /I/ vowels, which are rather short in natural speech (Peterson & Lehiste, 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
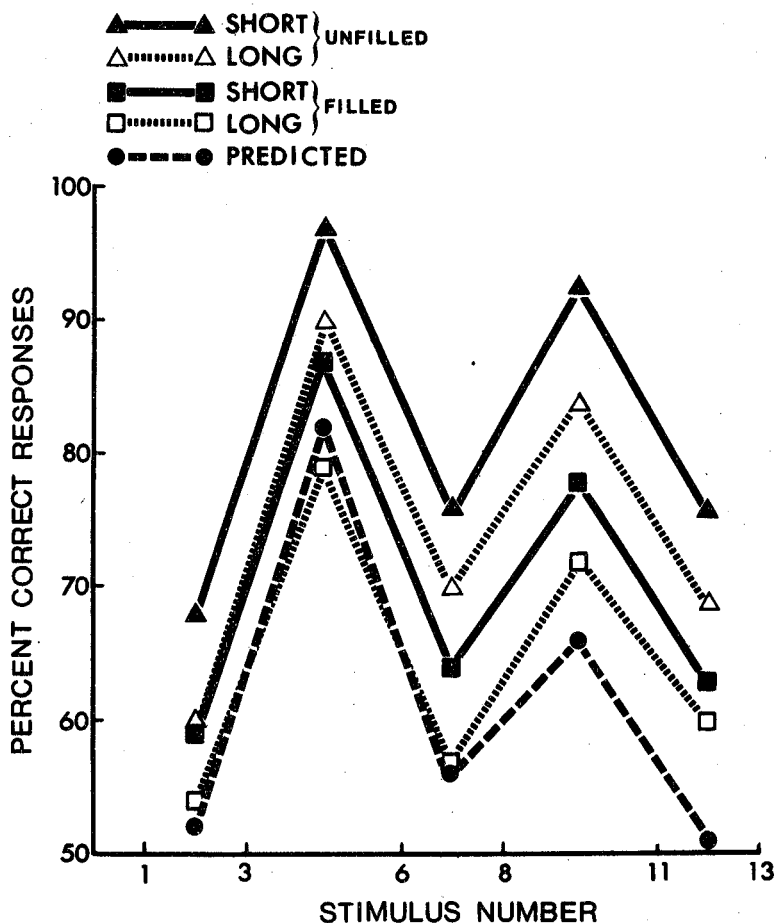
Figure 1. AX discrimination scores in the four conditions of Experiment 1, together with scores predicted from identification responses.

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 interstimulus interval (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 it 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.[3] 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.

---

[3] The parallelism of the functions in Figure 1 and the corresponding absence of an interaction between the effects of delay and filling characterize only the raw percent-correct scores and, therefore, should not be taken at face value. For example, it should not be inferred that the contribution of auditory memory was equally large in all stimulus pairs. Note also that the filling at the short and long interstimulus intervals was not strictly equivalent, as it involved different numbers of interpolated sounds.

Table 3
Stimulus Order Effect in Experiment 1

Stimuli	% "different" responses	
	$i < j$	$i > j$
1, 3	21	43
3, 6	93	97
6, 8	34	63
8, 11	63	79
11, 13	41	34

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

As expected, discrimination performance was poorest in the long-filled condition. To find out whether the scores in this condition approached those expected under the traditional categorical perception model (Liberman et al., 1957), we predicted the percentages of correct responses in the discrimination test from the identification responses, under the assumption that discrimination was based solely on phonetic labels (Pollack & Pisoni, 1971). These predictions, computed separately for each subject and then averaged, are indicated in Figure 1 by the heavy 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. At the individual level, only 11 of the 16 subjects exceeded the expectations. Thus, overall performance in the long-filled condition was only marginally superior to 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, averaged across conditions and subjects, in Table 3. 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. The stimulus-order effect was more pronounced

when the interstimulus interval 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 1 support the hypothesis that isolated steady-state vowels will be perceived categorically when auditory memory is interfered with. 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.

One problem with the present study is the wide, unequal stimulus spacing that we employed. Originally chosen to avoid the most ambiguous regions on the vowel continuum, the unequal spacing turned out to be disadvantageous when comparing predicted and obtained discrimination functions. Clearly, the stimulus spacing forced similar zigzag shapes on the predicted and obtained functions and thus prevented us from detecting any serious mismatches. More critically, 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, & Creelman, 1977). The peaks in the discrimination functions, it might be argued, represent simply the superior auditory discriminability of stimulus pairs 3–6 and 8–11. Even the peaks in the long-filled condition may reflect residual auditory memory for these larger stimulus differences, and their agreement with the predictions (Figure 1) may be 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. However, a more convincing test of categorical perception would include stimuli that are equally spaced along the continuum. Such a test is provided in Experiment 2, which was also designed to yield considerably more detailed data on the relationship between labeling and discrimination.

Experiment 2

Experiment 2 employed a 13-member vowel continuum (see 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 1, and a short-unfilled condition. As in Experiment 1, 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, Experiment 2 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 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 (Ainsworth, 1974; Eimas, 1963; Fry et al., 1962; Lindner, 1966; Thompson & Hollien, 1970; Kanamori, Kasuya, Arai, & Kido, Note 4). 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.[4] 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.

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 1. Three experimental

---

[4] To 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 & Kawashima, Note 1, Note 2) 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. (cf. Lane, 1965). Presumably, these authors used the discrimination sequences to collect labeling responses only to avoid the (in earlier years, considerable) effort involved in constructing a separate identification test.

tapes were prepared. The single-item identification tape contained a random sequence of 130 stimuli (10 repetitions of each of the 13 stimuli) with interstimulus intervals 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 1. 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 1, they circled beet, bit, or bet on an answer sheet. This task was followed by the two AX tapes presented twice each with different instructions. Under discrimination instructions, the subjects circled s (same) or d (different) on the answer sheet, as in Experiment 1. Under labeling instructions, the subjects circled beet, bit, or bet for each of the two vowels in a pair. The subjects were instructed to listen to both vowels before responding. The sequence of the discrimination and labeling conditions was counterbalanced across subjects, as was the sequence of the short-unfilled and long-filled conditions within each instruction condition.

Results

Single-item 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 1 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 1, we used these identification results to predict discrimination performance. The resulting predictions, averaged over subjects, are represented by the heavy 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.
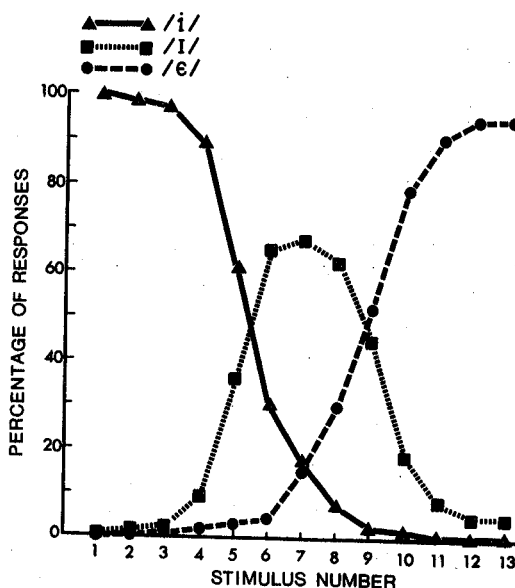

Figure 2. Labeling functions for stimuli presented in the single-item identification test of Experiment 2.

AX discrimination. The results of the discrimination task are displayed in Figure 3 in terms of percentage of correct responses, derived and plotted in the same manner as in Experiment 1 (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. There was no significant interaction between the factors of interstimulus interval and location on the continuum, $F(10, 140) = 1.4$, $p = .186$.

In Experiment 1, 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 13 of the 16 subjects, obtained performance
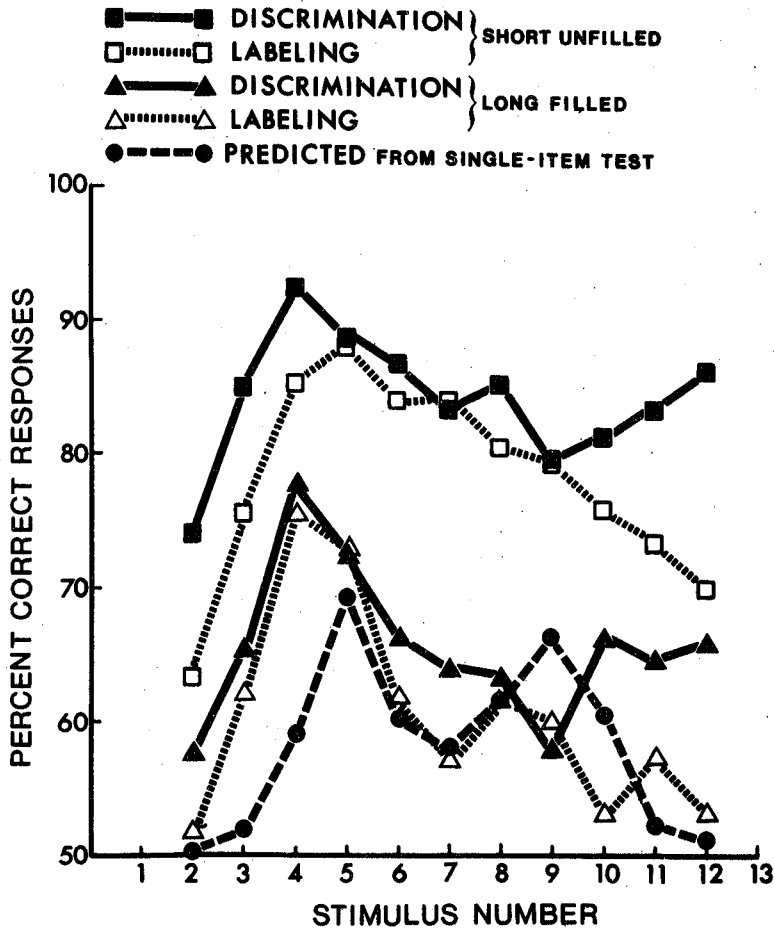
Figure 3. Two-step AX discrimination scores in the short-unfilled and long-filled conditions under discrimination and labeling instructions. (The labeling results represent the in-context predictions of discrimination performance. Also shown are the predictions derived from the responses in the single-item identification test.)

exceeded predicted performance in the long-filled condition. 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 suggest that the labeling probabilities of the stimuli changed in the context of the AX pairs. 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 interstimulus intervals should have equaled 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 performance, 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, although similar in profile, significantly exceeded these expectations, $F(1, 12) = 11.7$, $p = .005$. Figure 3 shows that this difference occurred at the two ends of the vowel continuum, particularly the right (/ε/) end, whereas scores in the middle region were similar. This pattern of results 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 interstimulus interval, $F(1, 12) < 1$.[5]

Stimulus order effect. The strong stimulus-order effect obtained for hits in the discrimination task of Experiment 1 was replicated in Experiment 2. The majority of the stimulus pairs received more "different" responses when the first stimulus in a pair had a lower position on the continuum (i.e., was more /i/-like) than the second stimulus, 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 interstimulus intervals. 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 1 in which 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.

To answer these questions, we 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

---

[5] The qualifications of Footnote 3 apply here, too. We cannot be sure that the differences between AX labeling and discrimination performance were equally large in the two conditions, but we would like to emphasize the small size of the difference in the short-unfilled condition. A breakdown of the responses into hits and false alarms may provide some additional information. A brief discussion of hit rates and false alarm rates may be found in an earlier version of the present paper (Repp, Healy, & Crowder, Note 5). They showed an interesting pattern, but we are not sure how to interpret it. We will be glad to supply appropriate summaries of our data to anyone interested in reanalyzing our results from a different viewpoint.

Table 4

AX Labeling Task: Percentages of /i/ and /ε/ Responses as a Function of Position of Target Stimulus in Pair (First or Second), Relative Location of Nontarget Stimulus (Lower or Higher), and Interstimulus Interval

| | Position of target stimulus | | | |
| | First | | Second | |
Interval	/i/	/ε/	/i/	/ε/
Short unfilled				
Lower	14	55	20	49
Higher	28	27	28	30
Long filled				
Lower	18	60	22	45
Higher	21	53	25	42

the other (nontarget) stimulus in the pair. The nontarget stimulus could be either two steps lower on the continuum, identical to the target, or two steps higher on the continuum. We then conducted an analysis of variance on response percentages pooled over stimulus pairs, with the following factors: vowel category (/i/ vs. /ε/; /I/ responses were omitted), position of 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 are shown in Table 4, averaged over subjects.[6]

There were pronounced contrast effects in the short-unfilled condition but only negligible effects in the long-filled condition. The overall contrast effect (the interaction of vowel category and relative location of nontarget stimulus) 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 very small. It also can be seen that, surprisingly, the retroactive effect (target stimulus first) was

slightly stronger than the proactive effect (target stimulus second) at the unfilled-short interval, although this difference turned out not to be significant.

In the long-filled condition, response frequencies also varied as a function of stimulus position in an AX pair, $F(1, 15) = 10.6$, $p = .006$. This 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. Moreover, the response frequencies in the AX labeling task were generally shifted toward the higher (/ε/) end of the continuum, relative to the labeling responses for the same stimuli in isolation. (This is not shown in Table 4.) It is not known what caused this shift in criteria, 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.

Discussion

Are Vowels Perceived Categorically?

Two aspects of categorical perception. The principal question of our research was whether isolated steady-state vowels would be perceived categorically when auditory memory is interfered with. Experiment 1 suggested an affirmative answer. However, the much more fine-grained analysis in Experiment 2 indicated that the answer depends on the exact form in which the question is asked.

By taking contrast effects into account, we have separated 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. (Absolute is, incidentally, the primary

---

[6] For a discussion of these results in terms of category boundary shifts, see Repp et al. (Note 5).

dictionary definition of the word categorical.) If the labeling of a stimulus depends on the preceding or following stimulus, perception is by definition noncategorical. However, listeners may nevertheless use these context-dependent categories in a discrimination task. Thus, rather than conceiving of a single continuum from continuous to categorical perception, it is necessary to define two aspects or dimensions of categorical perception: *Absoluteness* (the extent to which stimuli are perceived absolutely) and *predictability* (the extent to which discrimination performance can be predicted from labeling performance). For perception of a set of stimuli to be considered categorical, the stimuli must be high on both of these dimensions. However, stimuli that are not perceived categorically may take one of three forms: They may be low on both dimensions and hence be truly continuous; they may be low on absoluteness but high on predictability; or they may be high on absoluteness but low on predictability. Whether these two dimensions are truly independent and whether there exist stimuli of each of these forms are questions open to empirical investigation.

We have obtained a fairly close fit between the in-context predictions and obtained discrimination performance for our vowel stimuli. Such a reasonable agreement 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 Macmillan et al., 1977). However, we have also found strong evidence for stimulus-context effects in vowel labeling, thereby indicating relative rather than absolute perception. Specifically, in the case in which the two test stimuli in an AX pair 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. Also, the labeling probabilities depended on the absolute position of a stimulus in an AX pair—the stimulus order effect. This pattern of results suggests that vowel perception is not truly categorical. While the predictability criterion is satisfied, the absoluteness criterion is not.

Vowels, consonants, and the operational definition of categorical perception. Although we have made no direct comparisons between performance on vowels and performance on stop consonants, our results suggest some similarities and differences between the two. 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 that we observed for vowels. Although we did obtain a significant discrepancy between predictions and discrimination performance for vowels, it is well-known 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 (Eimas, 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 labeling—the absoluteness criterion. Context effects seem to be much larger for vowels than for consonants.

Although the traditional definition of categorical perception has implied two dimensions, or requirements, only the predictability requirement was ever directly operationalized. The other requirement—that labeling be absolute—was generally satisfied by the long interstimulus intervals 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 dimensions. We prefer two separate tests, one for each dimension. 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, this may turn out to be the more informative part of the new operational definition. (However, see Footnote 1.)

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: Subjects base their responses 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:

1. The AX discrimination performance was 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 that 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 2.

3. Discrimination was poorer when there was a long-filled interval between AX stimuli than when there was a short-unfilled interval.

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 dual-coding model. Fujisaki and Kawashima (Note 1, Note 2) 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. (However, see Footnote 2.) Whenever two stimuli cannot be distinguished by their phonetic codes, it is assumed that listeners consult their auditory memory codes. 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, in which 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, in which 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 interstimulus interval 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 can account for discrimination peaks only by assuming that there are psychoacoustic discontinuities (natural boundaries) coinciding with phonetic-category boundaries (cf. Pastore et al., 1977). It is not clear what might cause such natural boundaries for steady-state vowels. The all-auditory 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 (cf. Ades, 1977; Pynn, Braida, & Durlach, 1972). On the other hand, discrepancies between obtained discrimination and predictions are no problem for this model, and, indeed, we obtained such a discrepancy 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, meaning 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. The second and more serious difficulty lies 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; however, these 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 a natural outcome of the all-phonetic 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 (cf. Chistovich & Kozhevnikov, 1970, for a similar argument).[7]

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 2 is no problem for the all-phonetic model, since only one peak was found in the AX labeling task of Experiment 2, and subjects are necessarily basing their responses on phonetic codes in that task.

To account for the performance level difference between the short-unfilled and long-filled conditions, the all-phonetic model assumes that the stimuli in an AX pair are

---

[7] Informal evidence suggests that there may have been an additional phonetic category, /e/, toward the right end of the vowel continuum, in which the largest discrepancies between predicted and obtained scores occurred. Since /e/ is not a phoneme in American English (the diphthong /eⁱ/ 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.

subject to reciprocal auditory contrast before they are categorized. We assume that when two stimuli reside together in some stage of auditory processing, there are reciprocal interactions between their representations similar to those found in visual-brightness contrast or other laterally inhibiting systems. At the short-unfilled interval these processes are maximized in comparison to the long-filled interval. In addition, listeners may wait to apply phonetic labels until both stimuli in a pair have been received. Thus, the first stimulus in the long-filled condition may be in a state of degraded representation in auditory memory by the time the subject categorizes it.

At present, the all-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.[8]

The all-phonetic model assumes that 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.[9] This model does 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. The mixed model 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 in the two conditions

must be assumed to be purely coincidental. Similarly, the stimulus-order effects were similar in the two conditions but would be

---

[8] An interesting hypothesis was proposed by Smith (Note 6), 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, e.g., Jamieson & Petrusic, 1975) and links this finding with the fact that /I/ tends to be shorter than /i/ and /ɛ/ in natural speech (Peterson & 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 light of this finding, Smith's hypothesis predicts just the opposite of both her own 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 (Chuang & Wang, 1978). If this were the case, a positive time-order error would tend to increase further 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 discrimination, nor does it fit into an all-auditory framework because of the mediating role of phonetic stimulus properties.

[9] The Stroop color-word interference effect shows this rule operating in the visual realm. It is as if the linguistic level of analysis always dominates the nonlinguistic level; this is of obvious adaptive value for human communication. In an experiment more directly related to ours, Bailey, Summerfield, and Dorman (Note 7) 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.

explained by the hybrid model as resulting from two different mechanisms. 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 and its resemblance to the dual-coding model, it seems clearly unparsimonious compared to the single-process models.

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 is true even under conditions presumed to be rich in auditory memory. 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.

Reference Notes

1. Fujisaki, H., & Kawashima, T. *On the modes and mechanisms of speech perception.* Annual Report of the Engineering Research Institute, University of Tokyo, 1969, *28*, 67–73.

2. Fujisaki, H., & Kawashima, T. *Some experiments on speech perception and a model for the perceptual mechanism.* Annual Report of the Engineering Research Institute, University of Tokyo, 1970, *29*, 207–214.

3. Sachs, R. M. *Vowel identification and discrimination in isolation vs. word context.* (Quarterly Progress Report No. 93). Research Laboratory of Electronics, Massachusetts Institute of Technology, 1969. Pp. 220–229.

4. Kanamori, Y., Kasuya, H., Arai, S., & Kido, K. *Effect of context on vowel perception* (Paper 20 C4). Seventh International Congress on Acoustics, Budapest, 1971. Pp. 37–40.

5. Repp, B. H., Healy, A. F., & Crowder, R. G. *Categories and context in the perception of isolated steady-state vowels.* New Haven, Conn.: Haskins Laboratories (Status Report on Speech Research SR-54), 1978. Pp. 1–30.

6. Smith, M. R. *An investigation of changes in categorical perception of vowels.* Unpublished manuscript, Department of Linguistics, University of Connecticut, 1976.

7. Bailey, P. J., Summerfield, Q., & Dorman, M. F. *On the identification of sine-wave analogues of certain speech sounds.* New Haven, Conn.: Haskins Laboratories (Status Report on Speech Research, SR-51/52) 1977. Pp. 1–25.

References

Ades, A. E. Vowels, consonants, speech, and nonspeech. *Psychological Review*, 1977, *84*, 524–530.

Ainsworth, W. A. The influence of precursive sequences on the perception of synthesized vowels. *Language and Speech*, 1974, *17*, 103–109.

Carney, A. E., Widin, G. P., & Viemeister, N. F. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 1977, *62*, 961–970.

Chistovich, L. A., & Kozhevnikov, V. A. Perception of speech. In L. A. Chistovich et al. (Eds.), *Theory and methods of research on perception of speech.* Washington, D. C.: Joint Publications Research Service (JPRS-50423), 1970. Pp. 1–101.

Chuang, C.-K., & Wang, W. S.-Y. Psychophysical pitch biases related to vowel quality, intensity difference, and sequential order. *Journal of the Acoustical Society of America*, 1978, *64*, 1004–1014.

Crowder, R. G. The sound of vowels and consonants in immediate memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 587–596.

Crowder, R. G. Representation of speech sounds in precategorical acoustic storage. *Journal of Experimental Psychology*, 1973, *98*, 14–24.

Cutting, J. E., Rosner, B. S., & Foard, C. F. Perceptual categories for musiclike sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology*, 1976, *28*, 361–378.

Darwin, C. J., & Baddeley, A. D. Acoustic memory and the perception of speech. *Cognitive Psychology*, 1974, *6*, 41–60.

Eimas, P. D. The relation between identification and discrimination along speech and non-speech continua. *Language and Speech*, 1963, *6*, 206–217.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. The identification and discrimination of synthetic vowels. *Language and Speech*, 1962, *5*, 171–189.

Ganong, W. F., III. *Selective adaptation and speech perception.* Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1977.

Jamieson, D. G., & Petrusic, W. M. Presentation order effects in duration discrimination. *Perception & Psychophysics*, 1975, *17*, 197–202.

Lane, H. The motor theory of speech perception: A critical review. *Psychological Review*, 1965, *72*, 275–309.

Lehiste, I. Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics*, 1976, *4*, 113–117.

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.

Lindner, G. Veränderung der Beurteilung synthetischer Vokale unter dem Einfluss des Sukzessivkontrastes. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 1966, *19*, 287–307.

Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. The psychophysics of categorical perception. *Psychological Review*, 1977, *84*, 452–471.

Pastore, R. E., et al. Common-factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 686–696.

Peterson, G. E., & Lehiste, I. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 1960, *32*, 693–703.

Pisoni, D. B. *On the nature of the categorical perception of speech sounds.* Unpublished doctoral dissertation, University of Michigan, 1971.

Pisoni, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 1973, *13*, 253–260.

Pisoni, D. B. Auditory short-term memory and vowel perception. *Memory and Cognition*, 1975, *3*, 7–18.

Pisoni, D. B. Fundamental frequency and perceived vowel duration. *Journal of the Acoustical Society of America*, 1976, *59*, (Suppl. No. 1), S39.

Pisoni, D. B., & Lazarus, J. H. Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, 1974, *55*, 328–333.

Pollack, I., & Pisoni, D. B. On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science*, 1971, *24*, 299–300.

Pynn, C. T., Braida, L. D., & Durlach, N. I. Intensity perception. III. Resolution in small-range identification. *Journal of the Acoustical Society of America*, 1972, *51*, 559–566.

Sachs, R. M., & Grant, K. W. Stimulus correlates in the perception of voice onset time (VOT): II. Discrimination of speech with high and low stimulus uncertainty. *Journal of the Acoustical Society of America*, 1976, *60* (Suppl. No. 1), S91 (A).

Samuel, A. G. The effect of discrimination training on speech perception. *Perception & Psychophysics*, 1977, *22*, 321–330.

Stevens, K. N. On the relation between speech movements and speech perception. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 1968, *21*, 102–106.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., & Öhman, S. E. G. Cross-language study of vowel perception. *Language and Speech*, 1969, *12*, 1–23.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 1970, *77*, 234–249.

Thompson, C. L., & Hollien, H. Some contextual effects on the perception of synthetic vowels. *Language and Speech*, 1970, *13*, 1–13.