364

Journal of Experimental Psychology:
Human Perception and Performance
1982, Vol. 8, No. 1, 68–80

0096-1523/82/0801-0068$00.75

Context Independence and Phonetic Mediation in Categorical Perception

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

Bruno H. Repp
Haskins Laboratories, New Haven, Connecticut

Journal of Experimental Psychology:
Human Perception and Performance
1982, Vol. 8, No. 1, 68-80

0096-1523/82/0801-0068$00.75

Context Independence and Phonetic Mediation in Categorical Perception

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

Bruno H. Repp
Haskins Laboratories, New Haven, Connecticut

Categorical perception, as an ideal situation, is rarely if ever observed in the laboratory. Two separate requirements must be met for categorical perception: (a) Discrimination performance must be phonetically mediated (i.e., it should be predictable from labeling performance), and (b) labeling responses must be independent of the stimulus context. To determine the extent to which departures from the ideal are due to failures to meet one or both of these requirements, we used four stimulus continua in same-different (AX) discrimination and labeling tasks: stop-consonant-vowel (CV) syllables, isolated vowels, isolated fricative noises, and nonspeech sounds varying in timbre. We found that perception of CV syllables fell short of the ideal only because of contextual influences on labeling. Neither criterion of categorical perception was met by vowels or fricative noises, but subjects were more prone to phonetic mediation with vowels than with fricative noises, and they showed more context independence with fricative noises than with vowels. Surprisingly, the nonspeech timbre stimuli satisfied both requirements better than either vowels or fricative noises. This finding was attributed to the short duration of our timbre stimuli, which prevented them from sounding vowel-like but at the same time may have prevented stable auditory memory traces.

In categorical perception subjects encode stimuli in terms of a few discrete categories, rather than in terms of continuous attributes. If perception is perfectly categorical, stimuli drawn from a physical continuum are discriminated only to the extent that they can be assigned different category labels. In practice performance deviates from this ideal by varying amounts, so it becomes necessary to speak of degrees of categorical perception. The extent to which perception of a given stimulus set in a given task is categorical has typically been assessed by comparing results of a discrimination task with predictions derived from an independent identification task. Repp, Healy, and Crowder (1979) pointed out, however, that this method confounds two aspects of categorical perception: context independence (which they called *absoluteness*) and phonetic mediation (which they called *predictability*). Context independence refers to the degree to which the phonetic categorization of a given stimulus is independent of the context in which it occurs. Phonetic mediation refers to the degree to which discrimination appears to be based on category labels, rather than on continuous sensory stimulus attributes. Perfect categorical perception satisfies both of these criteria, but departures from this rarely observed ideal may be due to less context independence, less phonetic mediation, or both. In other words subjects' (covert) labeling responses in a discrimination task may be subject to contextual influ-

This research was supported by National Institute for Child Health and Human Development Grant HD01994 and Biomedical Research Support Grant RR05596 to Haskins Laboratories and by National Science Foundation Grants BNS77-00077 and BNS80-00283 to Yale University. Alice F. Healy was supported by a Senior Faculty Fellowship from Yale University during the final stages of preparing this manuscript.

We are indebted to Robert G. Crowder, William K. Estes, Alvin M. Liberman, and Michael Studdert-Kennedy for helpful discussions about this research and to Patti Jo Price and Mary LaRue for help with data analysis.

Requests for reprints should be sent either to Alice F. Healy, who is now at the Department of Psychology, Muenzinger Building, University of Colorado, Boulder, Colorado 80309, or to Bruno H. Repp, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510.

ences, but the subjects may nevertheless base their discrimination judgments on these labels; or it may be that discrimination is not based on category labels, whether or not they change as a function of context.

The proposal that categorical perception involves two separate aspects that are confounded in the standard predictability test was first made by Lane (1965), but it was subsequently rejected by Studdert-Kennedy, Liberman, Harris, and Cooper (1970) on the grounds that the standard test is sufficient to determine whether a stimulus continuum is categorically perceived. Such a test cannot reveal the reasons for any deviations from the ideal pattern, however, and since deviations are almost always observed, their explanation is a central issue.

In their recent study Repp et al. (1979) applied this logic to isolated vowels, a type of stimulus that has been shown by conventional methods to be perceived in a noncategorical fashion (e.g., Fry, Abramson, Eimas, & Liberman, 1962; Stevens, Liberman, Studdert-Kennedy, & Öhman, 1969). The stimuli used by Repp et al. formed an /i-I-ɛ/ continuum. Degree of context independence was assessed by examining whether the overt labeling of these vowels changed when they were paired with other vowels from the same continuum. Extent of phonetic mediation was determined by comparing the probabilities of assigning two vowels in a pair same or different phonetic labels to the probabilities of responding "same" or "different" to precisely the same vowel pairs in a discrimination test. In addition, a standard single-item identification test was run. This methodology revealed that the noncategorical perception of isolated vowels derived primarily from the context sensitivity of these stimuli (just as Lane, 1965, conjectured): Once context-induced (invariably contrastive) shifts in labeling probabilities were taken into account, discrimination performance could be predicted fairly closely, indicating that vowel discrimination may be mediated in large part by phonetic categories.

This result suggested to us that context sensitivity and phonetic mediation are independent aspects of perception. Repp et al. (1979) hypothesized (in their all-phonetic model) that contextual influences arise prior to categorization via a mechanism of auditory contrast similar to lateral inhibition, and the predictability of discrimination performance from labeling performance reflects the listeners' reliance on category labels and their reluctance or failure to refer to additional auditory stimulus information. According to that view the size of context effects is determined by auditory stimulus properties, whereas the extent to which discrimination can be predicted from labeling presumably depends both on the relative accessibility of auditory stimulus information (cf. Fujisaki & Kawashima, 1969) and on the familiarity of the categories used. If contextual influences are relatively independent of the use of category labels in discrimination, then it might be possible to find a stimulus set whose perception, unlike that of isolated vowels, is relatively context independent and yet not phonetically mediated. In addition, of course, there may be stimulus sets that are high or low on both of these dimensions (i.e., whose perception is highly categorical or highly noncategorical).

In the present study we compared four different stimulus sets with regard to the context independence and phonetic mediation criteria, using the methodology of Repp et al. (1979), which had been previously applied only to vowel stimuli. We expected these stimulus sets to exhibit different patterns of results, as explained in more detail below. Thus, the results of our experiment were expected to bear on the question of whether context independence and phonetic mediation are independent aspects of categorical perception.

Our first set of stimuli was a continuum of stop-consonant–vowel (CV) syllables ranging from /ba/ to /da/. It is well known that these stimuli are perceived highly categorically (e.g., Liberman, Harris, Hoffman, & Griffith, 1957). Therefore they were expected to be high on both the context independence and phonetic mediation criteria. Nevertheless there was more to be learned about their perception. We were interested in whether they show any reliable context effects at all and, if so (cf. Eimas, 1963; Rosen, 1979), how the magnitude of these effects compares to those found for other stimuli. It is a common finding in conventional studies of categorical perception that discrimination performance is somewhat higher than predicted, even for stimuli that are perceived highly categorically. We wondered whether this discrepancy could be accounted for by context effects in covert labeling; perhaps the difference would disappear when in-context predictions (derived from subjects' labeling responses to stimuli presented in the same format as in the discrimination task) were used.

Our second set of stimuli was a continuum of isolated vowels ranging from /i/ to /I/. This part of the experiment was expected to provide a partial replication of the Repp et al. results and to enable us to make a more direct comparison with the other stimulus sets. On the basis of the Repp et al. (1979) findings, we expected the vowels to exhibit large contrast effects in labeling but relatively high predictability of discrimination scores from in-context labeling results, indicating phonetic mediation. We wondered whether phonetic mediation would be as strong for vowels as for CV syllables, as suggested by Repp et al.

Our third set of stimuli was a continuum of isolated fricative noises ranging from /ʃ/ to /s/. Considerably less was known about the perception of these stimuli than about the preceding two sets. Mann and Repp (1980) recently used them in a labeling task, however, and found that subjects assigned them to phonetic categories reliably and without difficulty. Informal observations also suggested that these noises were not particularly sensitive to context and were easy to discriminate. Thus, this stimulus set was a candidate for being high on context independence but low on phonetic mediation—a result that (in conjunction with the vowel results) would indicate that the two dimensions can be dissociated. This part of the experiment also served as a partial replication of a previous study by Fujisaki and Kawashima (1969), who, to the best of our knowledge, were the only authors ever to use a continuum of isolated fricative noises in a categorical perception task. They, like Mann and Repp (1980), found very reliable identification of these noises, as well as better than chance discrimination within phonetic categories. They also found, however, a marked discrimination peak at the category boundary—a finding that was taken to indicate the involvement of phonetic categories in discrimination. We wondered whether this result could be replicated.

Our fourth set of stimuli was a continuum of nonspeech sounds varying in timbre. More precisely, they were isolated synthetic single-formant resonances varying in frequency but with a constant fundamental frequency. The categories that subjects used in classifying these stimuli were "low" and "high," referring to their relative pitch ("dull" and "sharp" or "dark" and "bright" might have been equally appropriate labels). Although this stimulus continuum had some aspects in common with a vowel continuum, it was expected to be perceived noncategorically, like other physical continua of simple nonspeech sounds. To prevent the stimuli from sounding vowel-like, a short duration (50 msec) was chosen. Because of the relative nature of the category labels, labeling responses were expected to be highly context dependent, and they were not expected to mediate discrimination.

Each of the four stimulus continua had the same number of stimuli (10) and categories (2). Since it is difficult to equate relative discriminability across continua without extensive pilot work, we instead chose to present stimulus comparisons one, two, and three steps apart on each continuum. Thus, one-step differences on a continuum of easily discriminable stimuli might give performance levels comparable to those of two-step or even three-step differences of other stimuli that were more difficult to tell apart, thus enabling us to make a fair comparison.

Aside form its primary purpose—the separation of the two aspects of categorical perception—our study served as a detailed investigation of perceptual contrast effects, that is, the tendency to give successive stimuli different labels. We were in a position not only to compare the magnitudes of contrast effects across different stimulus continua but also to compare forward and backward contrast effects within stimulus pairs, to separate phonetic contrast from auditory contrast, and to investigate the influence of varying step size (i.e., physical stimulus difference) on the size of contrast. We hoped that our results would bring us closer to an understanding of the stimulus characteristics that facilitate or inhibit contrast between successive stimuli.

Method

Subjects

The subjects were 12 paid volunteers, men and women recruited by posters on the Yale University campus. None of them were experienced in discrimination tasks, although several had listened to synthetic speech for other experimental tasks conducted in the laboratory.

Stimuli

Four different continua of synthetic sounds were used. Each continuum contained 10 stimuli spaced in approximately equal physical steps. The first three (speechlike) continua were generated on the OVE IIIc serial resonance synthesizer at Haskins Laboratories; the fourth (nonspeech) continuum was created on the Haskins Laboratories parallel resonance synthesizer.

CV syllables. The CV syllables (/ba/–/da/) differed in the onset frequencies of the second and third formants, which are listed in Table 1. The transitions from these onset frequencies to the formant steady states (at 1233 Hz and 2520 Hz, respectively) were stepwise linear and 40 msec in duration. All CV syllables had in common a 30-msec first-formant transition (from 200 Hz to 771 Hz), a fundamental frequency contour that was steady at 125 Hz over the first 50 msec and then fell linearly to 80 Hz, a flat amplitude contour with a final ramp, and a total duration of 250 msec.

Vowels. The vowels (/i/–/I/) differed in the frequencies of the first three formants, which are listed in Table 1. All vowels had constant formant frequencies, a linearly falling fundamental frequency contour (from 125 Hz to 80 Hz), a flat amplitude contour with initial and final ramps, and a total duration of 250 msec. Due to synthesizer characteristics stimulus amplitude increased slightly across the continuum.

Fricative noises. The fricative noises (/ʃ/–/s/) differed in the frequencies of two fricative formants (poles), which are listed in Table 1. All stimuli had constant formant frequencies, flat amplitudes with initial and final ramps, and a total duration of 250 msec. Due to certain adjustments in the amplitude specifications at the synthesis stage, the stimuli had increasingly lower amplitudes (a total decrease of about 4 dB), flatter amplitude ramps, and relatively more abrupt onsets towards the high (/s/) end of the continuum. These factors may have contributed to the discriminability of the noises, but this contribution was expected to be small because differences in noise spectra were salient to begin with.

Timbres. The timbres (low–high) were single (second) formant resonances varying in frequency (see Table 1). All timbres were steady state, with a fundamental frequency of 124 Hz, a flat amplitude contour, and a total duration of 50 msec. The short duration was chosen to reduce the speechlikeness of the stimuli (250-msec timbres sounded vowel-like) as well as their discriminability, which seemed too high initially. (Spacing on the continuum could not be reduced because of synthesizer limitations.)

For each of the four stimulus sets, two tapes were recorded using the Haskins Laboratories stimulus sequencing program. Except for the differences in stimuli, these tapes were identical for all four sets. The *simple identification tapes* contained 20 repetitions of each of the 10 stimuli on a given continuum, arranged in four random sequences of 50 (five repetitions of each stimulus) with 3-sec interstimulus intervals (ISIs). In addition, the two endpoint stimuli of the continuum were recorded five times in alternation at the beginning of the tape to provide examples of the two categories. The *same–different* (*AX*) *tapes* contained four random sequences of 68 stimulus pairs, with 300-msec ISIs within

Table 1
Stimulus Parameters (in Hz)

Stimulus no.	CV syllables		Vowels			Fricative noises		Timbres (F2)
	F2	F3	F1	F2	F3	P1	P2	
1	859	1795	269	2296	3019	1957	3803	2156
2	937	1929	281	2263	2976	2197	3915	2234
3	1022	2059	293	2247	2933	2466	4148	2307
4	1099	2197	304	2214	2912	2690	4269	2387
5	1181	2328	315	2198	2870	2933	4394	2462
6	1260	2466	327	2167	2829	3199	4655	2540
7	1345	2594	339	2151	2789	3389	4792	2615
8	1425	2729	351	2120	2749	3591	4932	2692
9	1510	2870	364	2105	2709	3917	5077	2762
10	1588	2998	375	2075	2670	4243	5322	2837

Note. CV = stop-consonant–vowel. F1 = Formant 1; F2 = Formant 2; F3 = Formant 3. P1 = Pole 1; P2 = Pole 2.

pairs and 4-sec ISIs between pairs. The 68 pairs in a block included the 10 identical, 9 one-step, 8 two-step, and 7 three-step pairs, in both possible stimulus orders, $2 \times (10 + 9 + 8 + 7) = 68$.

Procedure

Each subject participated in four sessions, one for each stimulus type. The sequence of stimulus types was counterbalanced across subjects according to a Latin-square design. There were three tasks in each session; the sequence of tasks was likewise counterbalanced across subjects but was fixed for a given subject across the four sessions.

Simple identification task. The subjects were first presented with the alternating endpoint stimuli to exemplify the response categories. Then they assigned a written label to each stimulus heard. The symbols used for the four stimulus types were b, d (CV syllables); i, I (vowels); sh, s (fricative noises); L, H (timbres).

AX labeling task. The subjects assigned labels to both stimuli in each pair. The same labels as in the simple identification task were used. If the AX labeling task was first in a session, it was preceded by examples of the endpoint stimuli (from the simple identification tape).

Ax discrimination task. In this task only the responses changed; they were now s (same) and d (different), and the subjects were carefully instructed to listen for any difference between the stimuli.

In all conditions the subjects were given a brief preview of the tapes: A randomly selected section was played for 1–2 minutes, and subjects listened without responding. The subjects listened to the stimulus tapes in a quiet room over TDH-39 earphones. The tapes were played back on an Ampex AG-500 tape recorder. Due to their different acoustic characteristics, the different stimulus types varied somewhat in overall amplitude, but all were within a comfortable listening range.

Results

Simple Identification

The results of the single-item identification test are summarized in Figure 1 in terms of percentages of b and d responses for CV syllables, i and I responses for vowels, sh and s responses for fricative noises, and L and H responses for timbres. The CV syllables differ from the other three stimulus sets in that the labeling functions are steeper and the category boundary (the 50% crossover point of the labeling function) is definitely off center (the b category being larger than the d category), whereas the other category boundaries fall close to the centers of the respective continua (between Stimuli 5 and 6).

We used these identification results to predict discrimination performance, following the classical low-threshold model (Pollack & Pisoni, 1971). The resulting predictions, averaged over subjects, are represented in the top row of Figure 2 in terms of percentage of responses of "different" as a function of stimulus number and step size.

Phonetic Mediation

The results of the AX discrimination task are displayed in the bottom row of Figure 2 in terms of percentage of responses of "different" as a function of stimulus number and step size. In the center row of Figure 2 are the corresponding scores derived from the AX labeling task by computing the percentages of trials on which the two stimuli in a pair were given different labels. The discrimination scores in the center row provide the in-context predictions of the scores in the bottom row; the accuracy of these predictions (together with other considerations discussed below) provides an indication of the extent to which AX discrimination might have been mediated by category labels.

Separate analyses of variance for each step size of each stimulus type were performed to compare the discrimination scores in the two tasks. These analyses revealed a significant discrepancy in favor of the discrimination task in all cases ($p < .05$ or less).[1] These significant differences between tasks, however, do not in themselves imply that discrimination scores were significantly higher in the discrimination task, since both hits (one- to three-step functions) and false alarms (zero-step functions) showed larger values than in the labeling task, indicating that subjects had a greater tendency to respond "different" in the discrimination task (particularly with CV syllables and timbres).

[1] It is interesting to note that the two tasks differ most markedly within, rather than between, phonetic categories. This observation is difficult to interpret, however, since in this experiment each continuum spanned only two phonetic categories, so all within-category comparisons occurred at the ends of a continuum and all between-category comparisons occurred in the middle of a continuum. Location on the continuum may have been the critical factor, rather than the categories of the two stimuli being compared.

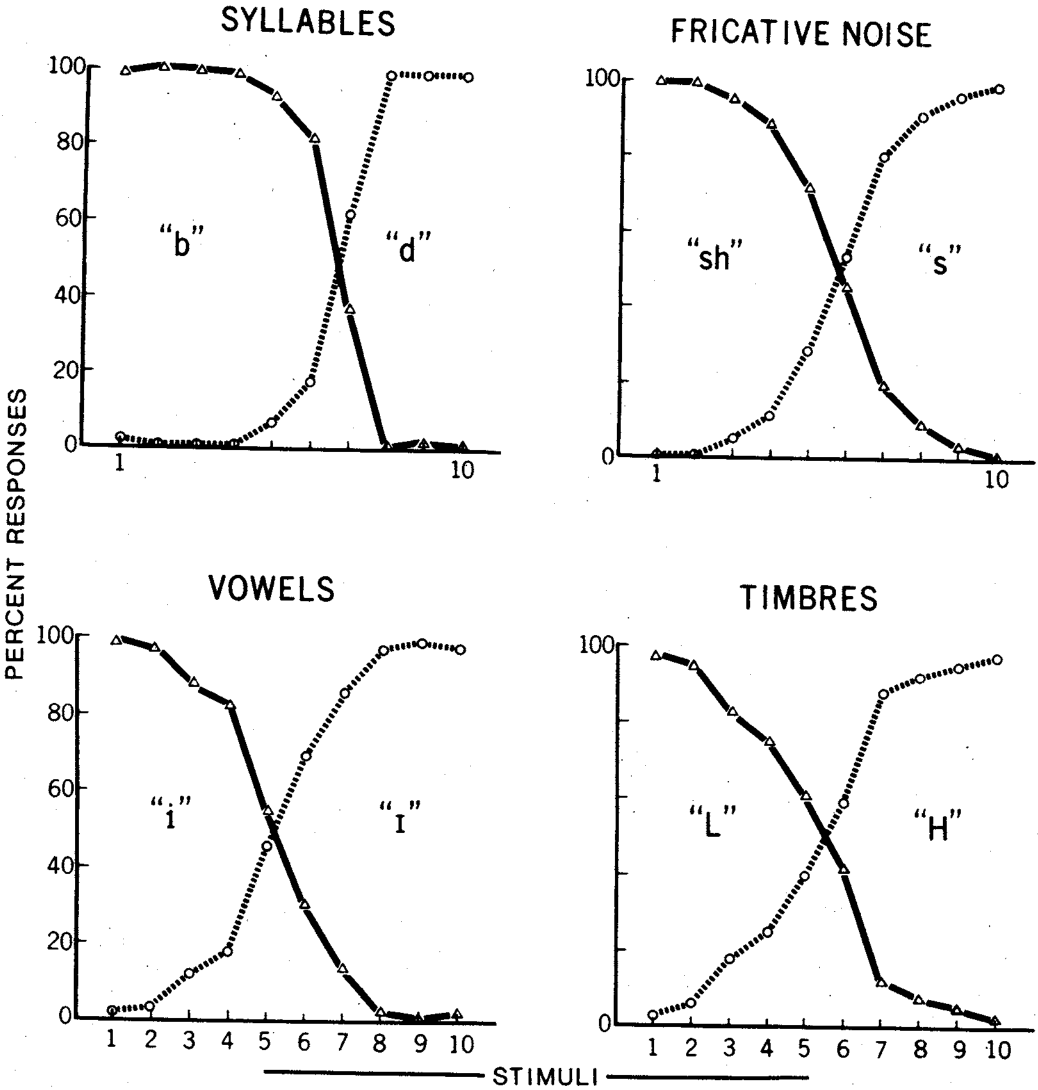

Figure 1. Labeling functions for the four stimulus continua in the simple identification task.

To control for this shift in response criteria, values of d' were obtained from the tables provided for the AX paradigm by Kaplan, Macmillan, and Creelman (1978). To obtain sufficiently stable estimates of d', it was necessary to average hit rates (separately for the three step sizes) and false alarm rates (based on pairs of identical stimuli) across stimulus pairs on each continuum before determining d' values for each subject and each stimulus type.[2] The values of d', averaged across subjects, are shown in Table 2.

An analysis of variance of these d' values included the following factors: step size, task (discrimination vs. labeling), and stimulus type. The overall difference between discrimination and labeling tasks was significant, $F(1, 11) = 60.8$, $p < .001$, as was the

[2] Unequal frequencies of individual stimuli were taken into account, and values of 0 and 1 were treated as .01 and .99, respectively, in the table look-up ($d'_{max} = 6.93$). Of course, the resulting d' values were rather crude indices, but since they were computed in the same way for both tasks, they provided a useful comparison.

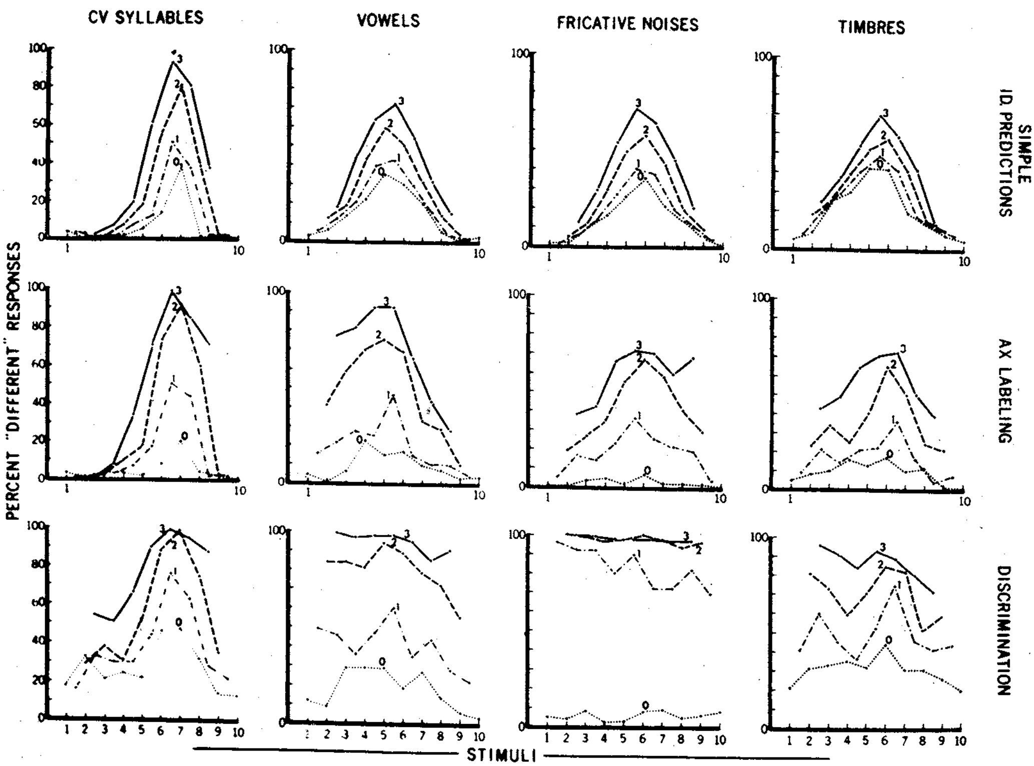

Figure 2. Percentage of "different" responses in the AX discrimination task (bottom row), in the AX labeling task (middle row), and as predicted from simple identification (top row). (CV = stop-consonant–vowel.)

Stimulus Type × Task interaction, $F(3, 33) = 48.0$, $p < .001$. The performance level in the discrimination task exceeded that in the labeling task for timbres, $F(1, 11) = 7.5$, $p = .019$, for vowels, $F(1, 11) = 21.4$, $p = .001$, and especially for fricative noises, $F(1, 11) = 131.8$, $p < .001$, whereas CV syllables showed a marginally significant difference in the opposite direction, $F(1, 11) = 4.5$, $p = .056$. This reversal for CV syllables suggests that listeners, in their (unsuccessful) attempt to make fine discriminations among CV syllables in the discrimination task, made less effective use of category labels than in the labeling task. It also suggests that the commonly observed advantage of CV syllable discrimination scores over predictions derived from single-item identification tests may be solely due to context effects in the discrimination paradigm (see below)—that is, that the advantage is an artifact of using inappropriate predictions. For vowels the significant performance advantage of the discrimination over the labeling task indicates that, contrary to the predictions of Repp et al. (1979), the discrimination of isolated vowels is not phonetically mediated to the same extent as is the discrimination of CV syllables.[3] Phonetic mediation seems to play little or no role in fricative noise discrimination, for which performance was exceedingly high even within categories.

Clearly, the magnitude of the overall difference between the scores from discrimination and labeling tasks cannot be taken as

[3] Note that our finding of a significant difference between performance in the discrimination and labeling tasks for vowels is consistent with the report of Repp, Healy, and Crowder (1979, Experiment 2), who also found a significant difference but, because the difference was small, predicted that it was not much larger than would be found under comparable conditions for stop-consonant–vowel syllables.

Table 2
Average Values of d' as a Function of Task and Step Size for Each Stimulus Type

Stimulus type and task	Step size 1	2	3
CV syllables			
Labeling (L)	1.20	2.14	2.90
Discrimination (D)	.93	1.75	2.90
D − L	−.27	−.39	.00
Vowels			
Labeling	1.24	2.41	3.15
Discrimination	1.57	3.32	4.38
D − L	.33	.91	1.23
Fricative noises			
Labeling	1.90	2.90	3.59
Discrimination	4.69	5.80	5.78
D − L	2.79	2.90	2.19
Timbres			
Labeling	.82	1.75	2.54
Discrimination	1.30	2.36	3.39
D − L	.48	.61	.85

Note. CV = stop-consonant–vowel.

the sole indicator of whether discrimination responses are mediated by category labels. Even if category labels play no role at all, performance differences between the two tasks shrink when discrimination is made sufficiently difficult. In other words one must beware of floor effects. To assess the possible role of mediation by category labels, the shapes of discrimination functions from the two tasks need to be compared as well. If category labels were used in the discrimination task, performance should be better in the category boundary region than within categories, as it is naturally in the labeling task. Thus, the discrimination functions should show peaks at the same points in both tasks.[4] (Compare the graphs in the bottom row with those in the middle row of Figure 2.)

In the discrimination task such peaks are clearly present for CV syllables. The vowels show small peaks in the boundary region, especially in the one-step function, indicating that category labels did play some role. Performance with fricative noises was too close to the ceiling, at least for two- and three-step functions, for any clear peaks to be exhibited. The timbre results are puzzling: The discrimination functions (especially one-step and two-step) do exhibit peaks in the category boundary region,[5] even though it might seem implausible that the subjects relied on the arbitrary category labels (high and low) in making their discriminations. There is no obvious psychoacoustic reason, however, why discriminability should have been higher in the center of the timbre continuum. We will return to this unexpected result with timbres in our discussion below. In summary, category labels seemed to play a role in the discrimination of CV syllables and—to some extent—in the discrimination of vowels and of timbres.

For three of the stimulus types—vowels, fricative noises, timbres—the listeners must have made (additional) use of auditory information in the discrimination task. Auditory information should become more available as the physical stimulus differences increase. As can be seen in Table 2, *d'* scores

[4] We take such a coincidence of peaks to indicate that category labels were involved in discrimination performance. We are aware of the possibility that discrimination peaks could be due, instead, either to inequalities in the spacing of the stimuli along the continuum or to a psychoacoustic boundary, which in turn may underlie phonetic labeling (cf. Macmillan, Kaplan, & Creelman, 1977; Pastore, 1981; Pisoni, 1977). There were some minor irregularities in the spacing of our stimuli, but none of these seem able to account for features of our discrimination functions. Furthermore, of the stimulus continua we used, only the stop-consonant–vowel syllables, because of their complex acoustic structure, might conceivably straddle a psychoacoustic boundary; however, the existence of such a boundary has not been well established or shown to coincide with the phonetic boundary. The assumption that irregularities in a discrimination function can be directly mapped into psychophysical distances between stimuli (Macmillan et al., 1977) seems misguided to us because it denies the possible contribution of category labels and thereby may lead to circular reasoning.

[5] Analyses of variance performed on the discrimination data yielded significant effects of stimulus location ($p < .01$) for each step size of the timbres, indicating that the discrimination functions were not flat. We also examined our data separately for those subjects who received the discrimination task first and those who received the labeling or single-item identification tasks first. They showed similar discrimination functions, indicating that the discrimination peaks did not result from specific practice with the labels provided by us. Presumably, then, subjects without previous labeling experience spontaneously used a dichotomous code to partition the stimulus range.

increase with step size in both tasks. To reflect a true increase in auditory information, however, the increase should be larger in the discrimination task than in the labeling task—that is, the difference between tasks should increase as a function of step size. Such a Task × Step Size interaction can indeed be observed for vowels, $F(2, 22) = 9.5$, $p = .001$, and to a much smaller extent, for timbres, $F(2, 22) = 2.6$, $p = .097$. For fricative noises the results were distorted by a ceiling effect; otherwise they presumably would have shown a similar pattern. For CV syllables the interaction was not significant. This pattern of results further establishes that auditory information beyond the category labels is available to naive listeners for vowels and timbres, but not for CV syllables.

Context Independence

To assess the effects of stimulus context on identification in the AX labeling task, we tabulated the labeling response frequencies separately for stimuli occurring first and second in the stimulus pairs, and we then examined these frequencies for one (target) stimulus contingent on the nature of the other (nontarget) stimulus in the pair. Only Target Stimuli 4–7 (but all relevant context stimuli) were considered, since the other target stimuli could not be paired with both higher and lower context stimuli one, two, and three steps apart on a given continuum. Figure 3 shows the percentage of responses in the lower response category (the category associated with Stimulus 1) for each of these target stimuli as a function of the identity of the context (nontarget) stimulus. Separate panels are provided for targets in first and second position. A contrast effect appears as a positive slope of the lines in each graph, whereas a flat function would imply no contrast.

It can be seen that all four stimulus types exhibit contrast effects: The percentage of responses in the lower category was greater when the context stimulus was above than when it was below the target on the continuum, $F(1, 11) = 46.4$, $p < .001$.[6] The magnitude of the effect varies, however, with stimulus type—the Stimulus Type × Position of Context Stimulus relative to target (lower versus higher on the continuum) interaction was significant, $F(3, 33) = 3.7$, $p = .022$. This interaction may in part be due to a ceiling effect for Stimuli 4 and 5 of the CV syllables; note that CV Stimulus 7 shows contrast effects comparable in magnitude to those obtained with vowels. There were also some clear differences, however, in the magnitude of context effects for different types of stimuli: Separate analyses revealed significant contrast effects for vowels, $F(1, 11) = 56.7$, $p < .001$, CV syllables, $F(1, 11) = 39.2$, $p < .001$, and fricative noises, $F(1, 11) = 10.2$, $p = .008$, but not for timbres, $F(1, 11) = 2.3$, $p = .153$. In accordance with the data of Repp et al. (1979) and a recent report by Shigeno and Fujisaki (1980), retroactive contrast (target first) was significantly larger than proactive contrast (target second) for vowels, $F(1, 11) = 8.5$, $p = .014$. None of the other stimulus types showed a significant difference in this direction; timbres actually showed a tendency in the opposite direction.

The percentage of responses in the lower category increased with context stimulus position on both sides of the target, $F(2, 22) = 82.9$, $p < .001$. In other words the extent of contrast increased with the physical separation between target and context stimuli. This increase was greater for some stimulus types than for others, as revealed in a significant Context Stimulus Position × Stimulus Type interaction, $F(6, 66) = 4.7$, $p = .001$. This interaction may also in part be due to a ceiling effect for the CV syllables. Separate analyses conducted on each stimulus type revealed significant effects of context stimulus position for each: vowels, $F(2, 22) = 53.8$, $p < .001$; CV syllables, $F(2, 22) = 6.9$, $p = .005$; fricative noises, $F(2, 22) = 28.8$, $p < .001$; timbres, $F(2, 22) = 4.9$, $p = .017$.

According to our results timbres are highest in context independence, whereas fricative noises, CV syllables, and especially vowels show large contrast effects. Retroactive contrast effects are larger than proactive effects for vowels, but retroactive and proactive contrast effects are essentially equal for

[6] For the purpose of this analysis, responses to pairs of identical stimuli (indicated by squares in Figure 3) were not included.

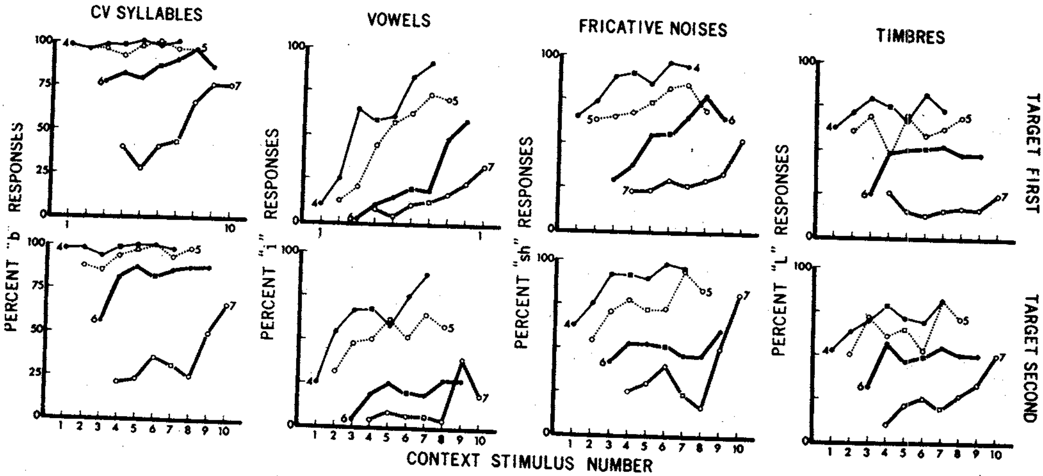

Figure 3. Context effects in the AX labeling task: Percentage of responses in the category associated with Stimulus 1, plotted as a function of target stimulus position (first or second), target stimulus number, and context stimulus number. (Pairs of identical stimuli are represented by squares. Only Target Stimuli 4–7 are included. CV = stop-consonant–vowel.)

other types of stimuli. Thus, the effects of stimulus context depend on the nature of the stimulus, and a simple explanation of these effects will not hold across different stimulus types. (In fact it is possible that the same explanation may not hold for all stimuli of a particular type, as suggested by the work of Sawusch, Nusbaum, & Schwab, 1980, who concluded that the vowels [i] and [I] produce contrast effects through different perceptual mechanisms.) Presumably, the pattern of context effects is conditioned by auditory stimulus properties, but at present it is not clear what the critical properties are, since there were a number of dimensions on which our stimuli differed.[7]

Discussion

Categorical perception is often understood to refer to the use of categories in discrimination (e.g., Macmillan, Kaplan, & Creelman, 1977); however, examination of the source literature (Liberman et al., 1957; Studdert-Kennedy et al., 1970) reveals that *categorical* was originally intended to mean *absolute*. Thus the original definition of categorical perception includes as criteria both context independence and the use of categories. One of the aims of the present study was to separate these two aspects by examining to what extent different sets of stimuli satisfy one or the other. Our results show that the two aspects are at least partially independent: Stimuli may exhibit large contrast effects even though discrimination is partially based on category labels (as in the case of vowels), or they may be less sensitive to context even though category labels play little role in discrimination (as in the case

[7] Another effect that also varied considerably across stimulus types was that of stimulus order. Although vowels did not show any consistent overall effect of stimulus order, the interactions of stimulus order and position were highly significant ($p = .002$ or less) at all three step sizes: At the left (/i/) end of the vowel continuum, more responses of different were obtained in both discrimination and labeling tasks when the first stimulus in a pair had a higher position on the continuum than did the second, but this effect was reversed at the right (/I/) end of the continuum. This stimulus order effect is similar to one found in the study by Repp, Healy, and Crowder (1979), although the reversal occurs at an earlier point on the vowel continuum in the present study.

Stop-consonant–vowel syllables showed stimulus order effects, but their direction was inconsistent across different step sizes. For fricative noises the high performance level may have prevented strong order effects. Timbres, when arranged from high to low frequency (in analogy to the second formant of the vowel continuum, which was in the same frequency range), showed weak trends in the same direction as vowels. These differences in the nature and size of the stimulus order effects as a function of stimulus type imply that these effects are not artifacts of the experimental design, but rather reflect properties of the stimuli employed.

of our fricative noises). Both vowels and fricative noises are noncategorically perceived, but apparently for different reasons—vowels primarily due to their high context sensitivity, fricative noises primarily due to absence of phonetic mediation.[8]

Using the methodology proposed by Repp et al. (1979), we demonstrated that discrimination performance for CV syllables does not exceed the predictions based on labeling when context effects are taken into account (so-called *in-context* predictions). Thus the small discrepancy between predicted and obtained discrimination performance in past studies was most likely due to context effects in covert labeling during the discrimination task. Our results strongly support the hypothesis that listeners, at least naive ones, discriminate CV syllables by relying exclusively on phonetic category information. In fact the task requirement of detecting within-category distinctions seems to lead to a somewhat less efficient use of category labels, but not to the recovery of auditory information. It has been shown, however, that auditory properties of stop consonants differing in place of articulation do become available after discrimination training (Edman, 1979).

A comparison of the results of vowels and fricative noises is revealing with regard to the possible determinants of context independence and phonetic mediation. In both stimulus types the distinctive spectral properties were constant throughout the stimulus duration, which was the same for vowels and fricative noises, and the labeling functions for the two stimulus continua were quite similar. Discrimination performance was much higher, however, for fricative noises than for vowels. Discrimination performance for two-step vowel pairs was similar to that for one-step fricative noise pairs (cf. Figure 2), so a fair comparison can be made between those portions of the results. Even when the obtained performance levels are thus equated, however, it is still true that vowel discrimination depends more on category labels, whereas fricative noises are less context sensitive. How are these differences to be explained?

The difference in phonetic mediation could arise from either or both of two sources: a difference in auditory distinctiveness or a difference in the salience of category labels. The much higher discrimination scores for fricative noises may reflect the greater auditory distinctiveness of these stimuli; alternatively (or in addition), listeners may have been able to ignore category labels and thus to access auditory information more successfully with fricative noises than with vowels. In other words isolated fricative noises are perhaps less speechlike than isolated vowels and, thereby, facilitate an auditory mode of processing. (However, see Footnote 8.)

The difference in the contrast effects exhibited by vowels and fricative noises is harder to explain. Although this difference is small overall, it is considerable when discrimination performance is equated (one-step fricative noises vs. two-step vowels). Some investigators have argued that contrast effects arise only after categorization of the stimuli (Fujisaki & Shigeno, 1979; Shigeno & Fujisaki, 1980), but there is evidence that this argument is not correct (Sawusch et al., 1980). Repp et al. (1979) found that contrast effects were greatly diminished when an irrelevant sound was interpolated between the two sounds in an AX pair. Such a manipulation should affect auditory (or precategorical) memory but not phonetic (or categorical) memory. The fact that in the present experiment the magnitude of contrast increased with the physical distance between target and context stimuli suggests strongly that the contrast effects arose at an auditory level. A more detailed analysis of our data revealed no consistent effects of the label given to a context stimulus on the magnitude of the contrast effect induced by that stimulus. Moreover, if response contrast had been operating, we should have observed frequent assignments of different labels to stimuli that were physically identical. In fact, however, Figure 2

[8] The possibility remains that fricative categories come into play when the step size on the continuum is small enough to make auditory discrimination very difficult, as in the study by Fujisaki and Kawashima (1969). Since we did not employ such a condition, our results are not necessarily incompatible with the Fujisaki and Kawashima data, which showed a discrimination peak at the category boundary.

reveals that these responses were less frequent in the AX labeling task than was predicted from simple identification. Therefore, we see no evidence for phonetic contrast. We must look at the auditory properties of the stimuli to understand the basis for the contrast phenomenon. The primary difference in auditory terms between vowels and fricative noises seems to be the periodic versus aperiodic nature of the waveform. Perhaps it is with periodic stimuli such as vowels that especially large contrast effects are found. (See May, 1979, for a similar hypothesis.) Clearly this hypothesis requires further testing (e.g., by using whispered vowels).

The pattern of results for the nonspeech stimuli, the timbres, was unexpected in several respects. We expected timbres to be the least categorically perceived of the stimuli we studied, since the category labels attached to the stimuli were completely relative. For that reason it seemed unlikely that subjects would base their responses on category labels (especially when the discrimination task preceded the labeling and single-item identification tasks—see Footnote 5) or that the category labels would be stable across changes in stimulus context. On the contrary, we found relatively small contrast effects and some indications of category mediation for timbres. Therefore, timbres satisfied to some extent both of the criteria for categorical perception, despite their status as nonspeech sounds and despite the arbitrary character of their category labels.

In attempting to explain these unexpected results, we are inevitably led to consider the fact that the timbre stimuli were very short in duration. Whereas all the other stimuli used were 250 msec long, the timbres were only 50 msec. This short duration was necessary to insure that our timbres would not be mistaken for vowels. Fujisaki and Kawashima (1969) and Pisoni (1973) have reported that short vowels are perceived more categorically than long vowels, presumably because they have a less stable representation in auditory memory, which increases listeners' reliance on category labels. Likewise our subjects may have been forced to rely on category labels, albeit arbitrary ones, in discriminating the short-duration timbres because they were unable to hold these sounds in auditory memory. This argument is consistent with the fact that the critical portion of the highly categorical CV syllables was short in duration, although the entire stimulus was 250 msec long.

An explanation must still be found for the fact that timbres were relatively context independent. The short duration of the stimuli may have been critical in this regard as well, since stable auditory memory traces may be required for contrast effects to be exhibited. Duration per se, however, may not provide a sufficient account for the context effects obtained in this experiment. The fricative noises were as long in duration as the steady-state vowels, but they exhibited a smaller contrast effect. In addition, Fujisaki and Shigeno (1979) have reported relatively small contrast effects with timbres that were 100 msec in duration, whereas they found larger contrast effects for vowels of the same duration; and Shigeno and Fujisaki (1980) found virtually no contrast with 200-msec vowels differing along a nonphonetic dimension (fundamental frequency), but large effects with 200-msec vowels from an /a/–/u/ continuum.

The relatively high auditory similarity of the timbre stimuli (as evidenced by their poor discriminability) may be another factor that contributed to the weakness of the contrast effect. Indeed Fujisaki and Shigeno (1979) have demonstrated that magnitude of the contrast effects is decreased when the stimuli being compared are highly similar. (See also Crowder, in press, for a relevant discussion.) Our own data corroborate these findings, since we also found smaller contrast effects for pairs of stimuli that were adjacent to each other on the continuum. (Note the tendency for the functions in Figure 3 to be flatter in the vicinity of the squares representing the identical pairs.) This line of reasoning, however, would lead one to expect the largest contast effects with fricative noises, since they were discriminated most easily. Instead, the fricative noises showed contrast effects that were smaller than those for vowels. Hence, auditory similarity alone cannot account for the magnitude of the contrast effects obtained with a given set of stimuli.

One factor that deserves serious attention is the relative speechlikeness of a stimulus. The hypothesis that only speechlike stimuli show large contrast effects appears to be consistent with our data and with those of Fujisaki and Shigeno (1979).

In conclusion, stimulus continua rarely, if ever, perfectly satisfy the standard predictability test of categorical perception, in which discrimination performance is predicted from performance on a single-item identification test. We have focused on two important causes for these departures from the ideal: Either the subjects may not rely wholly on category labels in discrimination, or the labels they use may be subject to contextual influences. Our data suggest that these two factors may vary independently. In particular we have shown that the departure from the ideal for CV syllables is entirely due to contextual influences on labeling. We have also shown that fricative noises and vowels are perceived noncategorically for both reasons, but with context effects playing a larger role for vowels and reliance on auditory information playing a larger role for fricative noises. The nonspeech continuum of timbres that we studied proved surprisingly to be more categorically perceived than either fricative noises or vowels, due to both smaller context effects and greater apparent reliance on category labels, albeit arbitrary ones. We tentatively ascribe this finding to the short duration of these stimuli, which may have prohibited the development of stable auditory memory traces.

References

Crowder, R. G. The role of auditory memory in speech perception and discrimination. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech*. Amsterdam, The Netherlands: North-Holland, in press.

Edman, T. R. Discrimination of intraphonemic differences along two place of articulation continua. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers presented at the 97th meeting of the Acoustical Society of America*. New York: Acoustical Society of America, 1979.

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.

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.

Fujisaki, H., & Shigeno, S. Context effects in the categorization of speech and nonspeech stimuli. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers presented at the 97th meeting of the Acoustical Society of America*. New York: Acoustical Society of America, 1979.

Kaplan, H. L., Macmillan, N. A., & Creelman, C. D. Tables of *d'* for variable-standard discrimination paradigms. *Behavior Research Methods and Instrumentation*, 1978, *10*, 796–813.

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

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.

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

Mann, V. A., & Repp, B. H. Influence of vocalic context on perception of the [ʃ]–[s] distinction. *Perception & Psychophysics*, 1980, *28*, 213–228.

May, J. G. *The perception of Egyptian Arabic fricatives*. Doctoral dissertation, University of Connecticut, 1979.

Pastore, R. E. Possible psychoacoustic factors in speech perception. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech*. Hillsdale, N.J.: Erlbaum, 1981.

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

Pisoni, D. B. Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, 1977, *61*, 1352–1361.

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

Repp, B. H., Healy, A. F., & Crowder, R. G. Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 129–145.

Rosen, S. M. Range and frequency effects in consonant categorization. *Journal of Phonetics*, 1979, *7*, 393–402.

Sawusch, J. R., Nusbaum, H. C., & Schwab, E. C. Contextual effects in vowel perception II: Evidence for two processing mechanisms. *Perception & Psychophysics*, 1980, *27*, 421–434.

Shigeno, S., & Fujisaki, H. Context effects in phonetic and non-phonetic vowel judgments. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* (University of Tokyo), 1980, *14*, 217–224.

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.

Received September 5, 1980
Revision received January 30, 1981 ■