CONTEXT SENSITIVITY AND PHONETIC MEDIATION IN CATEGORICAL PERCEPTION: A COMPARISON OF FOUR STIMULUS CONTINUA

Alice F. Healy and Bruno H. Repp

Abstract. Categorical perception is an ideal rarely, if ever, observed in the laboratory. Two separate requirements must be met for categorical perception: (1) predictability of discrimination performance from labeling performance, and (2) independence of labeling responses from stimulus context. In order to determine the extent to which instances of noncategorical perception are due to failures to meet one or both of these requirements, we employed four stimulus continua in AX discrimination and labeling tasks: stop-consonant-vowel (CV) syllables, steady-state vowels, fricative noises, and complex tones varying in timbre. We found that CV syllables departed from the ideal only because of contextual influences on labeling. Neither requirement was met by vowels or fricative noises, but fricative noises were less predictable than vowels, and vowels were somewhat less context independent than fricative noises. Surprisingly, the timbre stimuli were more predictable and showed smaller context effects than vowels or fricative noises. This finding was attributed to the shorter duration of the timbre stimuli, which may have prevented stable auditory memory traces.

INTRODUCTION

Categorical perception is a mode of perception in which stimuli are encoded in terms of a few discrete categories rather than in terms of continuous attributes. It is said to obtain when stimuli drawn from a physical continuum are discriminated not much better than would be predicted from a knowledge of the way in which they were assigned category labels. The degree of categorical perception of a stimulus set has typically been assessed by comparing results of a discrimination task with predictions derived from an independent identification task. However, Repp, Healy, and Crowder (1979)

Acknowledgment: This research was supported by the National Institute for Child Health and Human Development Grant HD01994 and Biomedical Research Support Grant RR05596 to Haskins Laboratories, and 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.

[HASKINS LABORATORIES: Status Report on Speech Research SR-63/64 (1980)]
pointed out that this method confounds two aspects of categorical perception: "context independence" (which they called "absoluteness") and "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. Predictability is the degree to which discrimination appears to be based on category labels, rather than on continuous sensory stimulus attributes. While a set of stimuli that is categorically perceived must satisfy both of these criteria, a set that is perceived not so categorically may be less context independent, less predictable, or both. In other words, subjects may change their (covert) labeling responses in the context of the discrimination task but 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 acknowledgment that categorical perception involves two separate aspects that are confounded in the standard predictability test was originally made by Lane (1965) but 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. However, such a test cannot reveal the reasons for any deviations from the ideal pattern, 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, & Ohman, 1969). The stimuli used by Repp et al. formed an /i-I-ε/ continuum. Degree of context independence was assessed by examining whether the labeling of these vowels changed when they were paired with other vowels from the same continuum. Extent of predictability was determined by comparing the probabilities of assigning two vowels in a pair same or different phonetic labels to the probabilities of assigning "same" and "different" responses 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 presumed noncategorical perception of isolated vowels derived primarily from the context sensitivity of these stimuli: Once context-induced (invariably contrastive) shifts in labeling probabilities were taken into account, discrimination performance could be predicted fairly closely, thus leaving open the possibility that vowel discrimination is mediated in large part by phonetic categories.

This result suggested to us that context sensitivity and phonetic mediation (predictability) 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, while the predictability of discrimination 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 that, unlike isolated vowels, shows small context effects (i.e., context independence) but poor predictability. In addition, of course, there may be stimulus sets that are high or low on both of these dimensions.

EXPERIMENT

In the present study, we compared four different stimulus sets with regard to the context independence and predictability criteria, using the methodology of Repp et al. (1979). We expected these stimulus sets to exhibit quite 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 predictability are independent aspects of categorical perception.

Our first set of stimuli was a continuum of 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 predictability 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) are 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 a basis for a more direct comparison with the other stimulus sets. On the basis of the Repp et al. findings, we expected the vowels to exhibit large contrast effects in labeling but relatively high predictability of discrimination scores from in-context labeling results. Whether predictability would be as high for vowels as for CV syllables was of particular interest, because of the suggestion by Repp et al. (1979) that vowels may be as predictable as CVs.

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. However, Mann and Repp (1980) recently used them in several labeling tasks 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 easy to discriminate. Thus, this stimulus set was a candidate for being high on context independence but low on predictability—a result that 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 dis-
crimination within phonetic categories. However, they also found a marked
discrimination peak at the category boundary—a finding that was taken to
indicate the involvement of phonetic categories in discrimination. We won-
dered whether this result could be replicated.

Our fourth set of stimuli was a continuum of brief complex tones varying
in timbre. They were isolated synthetic single-formant resonances varying in
frequency, but with a constant fundamental frequency. The categories 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.
Classification into essentially arbitrary categories was expected to be highly
context-dependent, and predictability was expected to be poor, because of the
absence of mediation by category labels.

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 differ-
ences of other stimuli that were more difficult to tell apart.

Aside from its primary purpose—the separation of the two aspects of
categorical perception—our study served as a detailed investigation of
perceptual contrast effects, i.e., 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, and to investi-
gate 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 was experienced in
discrimination tasks, although several had listened to synthetic speech for
other experimental tasks conducted in our 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 IIci serial
resonance synthesizer at Haskins Laboratories; the fourth (nonspeech) contin-
uum was created on the Haskins Laboratories parallel resonance synthesizer.
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 and 2520 Hz, respectively) were stepwise-linear and 40 msec in duration. All CV syllables had in common a 30-msec transition in the first formant (from 200 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.

<table>
<thead>
<tr>
<th>Stim. No.</th>
<th>CV Syllables</th>
<th>Vowels</th>
<th>Fric. Noises</th>
<th>Timbres</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F2</td>
<td>F3</td>
<td>F1</td>
<td>F2</td>
</tr>
<tr>
<td>1</td>
<td>859</td>
<td>1795</td>
<td>269</td>
<td>2296</td>
</tr>
<tr>
<td>2</td>
<td>937</td>
<td>1929</td>
<td>281</td>
<td>2263</td>
</tr>
<tr>
<td>3</td>
<td>1022</td>
<td>2059</td>
<td>293</td>
<td>2247</td>
</tr>
<tr>
<td>4</td>
<td>1099</td>
<td>2197</td>
<td>304</td>
<td>2214</td>
</tr>
<tr>
<td>5</td>
<td>1181</td>
<td>2328</td>
<td>315</td>
<td>2198</td>
</tr>
<tr>
<td>6</td>
<td>1260</td>
<td>2466</td>
<td>327</td>
<td>2167</td>
</tr>
<tr>
<td>7</td>
<td>1345</td>
<td>2594</td>
<td>339</td>
<td>2151</td>
</tr>
<tr>
<td>8</td>
<td>1425</td>
<td>2729</td>
<td>351</td>
<td>2120</td>
</tr>
<tr>
<td>9</td>
<td>1510</td>
<td>2870</td>
<td>364</td>
<td>2105</td>
</tr>
<tr>
<td>10</td>
<td>1588</td>
<td>2998</td>
<td>375</td>
<td>2075</td>
</tr>
</tbody>
</table>

The vowels (/i/-/I/) differed in the frequencies of the first three formants, which are listed in Table 1. All vowels were completely steady-state, with a linearly falling fundamental frequency contour (from 125 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.

The fricative noises (/ʃ/-/s/) differed in the frequencies of two fricative formants (poles), which are listed in Table 1. All stimuli were steady-state, had 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 quite salient to begin with.
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 4 random sequences of 50 (5 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 AX tapes contained 4 random sequences of 68 stimulus pairs, with 300-msec ISIs within 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 x (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.

In the simple identification task, the subjects were first presented with the alternating endpoint stimuli to exemplify the response categories. Then, they assigned in writing a 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).

In the AX labeling task, the subjects assigned labels to both stimuli in each pair. The same labels were employed as in the simple identification task. If the AX labeling task was first in a session, it was preceded by examples of the endpoint stimuli (from the simple identification tape). In the AX discrimination 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 before responding began: 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 at a comfortable loudness. Due to their different acoustic characteristics, the different stimulus types varied somewhat in overall amplitude, but all were within a comfortable listening range.
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 and 2520 Hz, respectively) were stepwise-linear and 40 msec in duration. All CV syllables had in common a 30-msec transition in the first formant (from 200 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.

<table>
<thead>
<tr>
<th>Stim. No.</th>
<th>CV Syllables</th>
<th>Vowels</th>
<th>Fric. Noises</th>
<th>Timbres</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F2</td>
<td>F3</td>
<td>F1</td>
<td>F2</td>
</tr>
<tr>
<td>1</td>
<td>859</td>
<td>1795</td>
<td>269</td>
<td>2296</td>
</tr>
<tr>
<td>2</td>
<td>937</td>
<td>1929</td>
<td>281</td>
<td>2263</td>
</tr>
<tr>
<td>3</td>
<td>1022</td>
<td>2059</td>
<td>293</td>
<td>2247</td>
</tr>
<tr>
<td>4</td>
<td>1099</td>
<td>2197</td>
<td>304</td>
<td>2214</td>
</tr>
<tr>
<td>5</td>
<td>1181</td>
<td>2328</td>
<td>315</td>
<td>2198</td>
</tr>
<tr>
<td>6</td>
<td>1260</td>
<td>2466</td>
<td>327</td>
<td>2167</td>
</tr>
<tr>
<td>7</td>
<td>1345</td>
<td>2594</td>
<td>339</td>
<td>2151</td>
</tr>
<tr>
<td>8</td>
<td>1425</td>
<td>2729</td>
<td>351</td>
<td>2120</td>
</tr>
<tr>
<td>9</td>
<td>1510</td>
<td>2870</td>
<td>364</td>
<td>2105</td>
</tr>
<tr>
<td>10</td>
<td>1588</td>
<td>2998</td>
<td>375</td>
<td>2075</td>
</tr>
</tbody>
</table>

The vowels (/i/-/I/) differed in the frequencies of the first three formants, which are listed in Table 1. All vowels were completely steady-state, with a linearly falling fundamental frequency contour (from 125 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.

The fricative noises (/ʃ/-/s/) differed in the frequencies of two fricative formants (poles), which are listed in Table 1. All stimuli were steady-state, had 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 quite salient to begin with.
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 speechlike-ness 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 4 random sequences of 50 (5 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 AX tapes contained 4 random sequences of 68 stimulus pairs, with 300-msec ISIs within 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.

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

In the AX labeling task, the subjects assigned labels to both stimuli in each pair. The same labels were employed as in the simple identification task. If the AX labeling task was first in a session, it was preceded by examples of the endpoint stimuli (from the simple identification tape). In the AX discrimination 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 before responding began: 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 at a comfortable loudness. Due to their different acoustic characteristics, the different stimulus types varied somewhat in overall amplitude, but all were within a comfortable listening range.
Results and Discussion

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-percent cross-over 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). This pattern of results, which was also found at the individual level, indicates a certain amount of context independence of CV syllables. The arbitrary category boundary for timbres was naturally expected to fall right in the center, as it did; the central locations of the vowel and fricative boundaries may have been simply a consequence of our selection of stimulus ranges.

We also 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 percent "different" responses as a function of stimulus number and step size.

Predictability. The results of the AX discrimination task are displayed in the bottom row of Figure 2 in terms of percent "different" responses as a function of stimulus number and step size. In the center row of Figure 2 are the corresponding scores ("in-context" predictions) derived from the AX labeling task by computing the percentages of trials on which the two stimuli in a pair were given different labels.

Separate analyses of variance for each step size of each stimulus type were performed to compare the discrimination functions to the analogous functions based on AX labeling. These analyses revealed a significant discrepancy in favor of the discrimination task for each stimulus type at each step size (p < .05 or less in each case). However, these significant differences between tasks do not in themselves imply that performance was significantly better than in the discrimination task, since both hits (1- to 3-step functions) and false alarms (0-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). In order to control for this response bias, values of d' were obtained from the tables provided for the AX paradigm by Kaplan, Macmillan, and Creelman (1978). To obtain relatively 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. 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 interaction of stimulus type and task, F(3,33) = 48.0, p < .001. The performance level in the discrimination
Figure 1. Labeling functions for the four stimulus continua in the simple identification task.
Figure 2. Percent "different" responses in the AX discrimination task (bottom row), in the AX labeling task (middle row), and as predicted from simple identification (top row).
task exceeded that in the AX labeling task for timbres, $F(1,11) = 7.5, p = 0.019$, for vowels, $F(1,11) = 21.4, p = .001$, and especially for fricative noises, $F(1,11) = 131.8, p < .001$, whereas AX labeling performance actually exceeded discrimination performance for CV syllables, although only with marginal significance, $F(1,11) = 4.5, p = .056$. The reversal for CV syllables suggests that listeners, in their (unsuccessful) attempt to make fine discriminations among CV syllables, made less effective use of category labels than in the labeling task. It also suggests that the commonly observed advantage of obtained CV syllable discrimination over scores predicted from single-item identification tests may indeed be due to context effects in the discrimination paradigm (see below)—i.e., that the advantage is an artifact of using inappropriate predictions. For vowels, the significant advantage of discrimination over labeling performance indicates that, contrary to the preliminary conclusions of Repp et al. (1979), the discrimination of isolated steady-state vowels is not phonetically mediated to the same extent as the discrimination of CV syllables. Phonetic mediation seems to play little or no role in fricative noise discrimination, where performance was exceedingly high even within categories.

\begin{table}[h]
\centering
\caption{Table 2}
\begin{tabular}{lccc}
\multicolumn{4}{l}{Average values of $d'$ as a function of task and step size for each stimulus type} \\
\hline
\multicolumn{1}{l|}{Step Size} & 1 & 2 & 3 \\
\hline
\hline
CV Syllables & & & \\
Labeling & 1.20 & 2.14 & 2.90 \\
Discrim & 0.93 & 1.75 & 2.90 \\
D-L & -0.27 & -0.39 & 0.00 \\
\hline
Vowels & & & \\
Labeling & 1.24 & 2.41 & 3.15 \\
Discrim & 1.57 & 3.32 & 4.38 \\
D-L & 0.33 & 0.91 & 1.23 \\
\hline
Fricative Noises & & & \\
Labeling & 1.90 & 2.90 & 3.59 \\
Discrim & 4.69 & 5.80 & 5.78 \\
D-L & 2.79 & 2.90 & 2.19 \\
\hline
Timbres & & & \\
Labeling & 0.82 & 1.75 & 2.54 \\
Discrim & 1.30 & 2.36 & 3.39 \\
D-L & 0.48 & 0.61 & 0.85 \\
\hline
\end{tabular}
\end{table}

Clearly, the magnitude of the overall difference between discrimination and labeling performance cannot be taken as a direct indicator of whether or not discrimination responses are mediated by category labels. Even if
category labels play no role, discrimination performance will approach labeling performance when discrimination is made sufficiently difficult. To assess the possible role of mediation by category labels, the shapes of the obtained discrimination and labeling functions 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. Thus, discrimination scores should show peaks at the same points as AX labeling scores. (Compare the figures in the bottom row with those in the middle row of Fig. 2.)

Such peaks are clearly present in the discrimination functions for CV syllables. The vowels show small peaks in the boundary region, especially in the 1-step function, indicating that category labels did play some role. Performance with fricative noises was too close to the ceiling, at least for 2- and 3-step functions, for any clear peaks to be exhibited. The timbre results are puzzling: The discrimination functions (especially 1-step and 2-step) do exhibit peaks in the category boundary region, even though it might seem implausible that the subjects relied on the arbitrary category labels, "high" and "low," in making their discriminations. However, there is no obvious psychoacoustic reason 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, the question of whether mediation by category labels played a role in discrimination is to be answered as follows: CV syllables—yes; vowels—in part; fricative noises—can't tell (if yes, category labels had little to contribute); timbres—in part (surprisingly).

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, both labeling and discrimination d' scores increase with step size. However, to reflect a true increase in auditory information, discrimination scores should increase more than labeling scores—i.e., the difference between labeling and discrimination scores should increase as a function of step size. Such an increase can indeed be observed for vowels [the interaction of task and step size was significant, 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 increase between step sizes 2 and 3 (Table 2) was not significant. This pattern of results further establishes that additional auditory information is available for vowels, timbres, and most likely fricative noises, but not for CV syllables.

Context independence. In order to assess the effects of stimulus context on identification in the AX labeling tasks of the present experiment, we tabulated the labeling response frequencies separately for stimuli occurring first and those occurring 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 were considered, since the other stimuli could not be paired with both higher and lower stimuli one, two, and three steps apart on a given continuum. The results are shown in Figure 3: The percentage of responses in the "lower"
Figure 3. Context effects in the AX labeling task: Percent 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.
response category (the category associated with stimulus 1) is shown, separately for each target stimulus, 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, $F(1,11) = 46.4, p < .001$. However, the magnitude of the effect varies with stimulus type—the interaction of stimulus type and position of context stimulus relative to target (lower versus higher) was significant: $F(3,33) = 3.7, p = .022$. This interaction may be due in part 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. Separate analyses conducted on each stimulus type 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), 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, $F(2,22) = 82.9, p < .001$. This increase was greater for some stimulus types than for others, as revealed in a significant interaction of context stimulus position and stimulus type, $F(6,66) = 4.7, p = .001$. This interaction may also be due in part 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 these results, timbres are highest in context independence (quite unexpectedly), with considerable contrast effects for fricative noises, CV syllables, and especially vowels. Note that the context effects obtained for the various stimulus types do not always take the same form. For example, retroactive contrast effects are larger than proactive effects for vowels, but retroactive and proactive contrast effects are essentially equal for CV syllables. The effects of stimulus context therefore depend on the nature of the stimulus, and a simple explanation of these effects will not hold across different stimulus types.

**GENERAL 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.
("predictability"). One of the aims of the present study was to separate these two aspects, by examining to which 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 of our fricative noises). Both vowels and fricative noises are noncategorically perceived, but apparently for different reasons—vowels primarily due to context sensitivity, fricative noises primarily due to lack of predictability.

Using the methodology proposed by Repp et al. (1979), we demonstrated that discrimination performance for CV syllables does not exceed labeling performance when context effects on labeling 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. However, it has been shown 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 predictability. 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. However, discrimination performance was much higher for fricative noises than for vowels. Discrimination performance for 2-step vowel pairs was similar to that for 1-step fricative noise pairs (cf. Figure 2), so a fair comparison can be made between those portions of the results. However, even when the obtained performance levels are thus equated, it is still true that vowels are more predictable (i.e., a larger portion of the discrimination scores can be accounted for by the use of category labels), whereas fricative noises are less context-sensitive. How are these differences to be explained?

The difference in predictability could arise from either or both of two sources: a difference in auditory distinctiveness, or a difference in the use of category labels in discrimination. The much higher discrimination scores for fricative noises may reflect the greater auditory distinctiveness of these stimuli; in addition, however, 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, the noises, being less speechlike, may have facilitated an auditory mode of processing.

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 (1-step fricative noises vs. 2-step vowels). Some investigators have argued that contrast effects arise only after categorization of the stimuli (Fujisaki & Shigeno,
1979), but there is evidence that this argument is not correct. Specifically, 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. Therefore, we must look at the auditory properties of the stimuli in order 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 the category labels or that the category labels would be stable across changes in stimulus context. On the contrary, we found a fair amount of predictability for timbres. In fact, the labeling performance for timbres matched the discrimination performance more closely than was the case for vowels (but less closely than for CV syllables). In addition, peaks at the category boundary region were found in the discrimination functions, although these peaks were considerably smaller than those found for CV syllables. Moreover, the magnitude of the context effects on labeling was smaller for timbres than for any of the other stimulus classes studied. Therefore, timbres tended to satisfy 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 employed were 250 msec long, the timbres were only 50 msec. This short duration was necessary in order to insure that our timbres would not be mistaken for vowels. Fujisaki & 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 predictable CV syllables was quite short in duration, although the entire stimulus was 250 msec long.

An explanation must still be found for the fact that timbres were high in context independence as well as predictability. 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. However, duration per se 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 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.

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 the magnitude of the contrast effects is decreased when the stimuli being compared are highly similar. (See also Crowder, 1980, 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.) However, this line of reasoning would lead one to expect the largest contrast 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.

In conclusion, stimulus continua rarely, if ever, perfectly satisfy the standard predictability test, 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 due entirely 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 surprisingly proved to be more categorically perceived than either fricative noises or vowels, due both to smaller context effects and to 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**


FOOTNOTES

1Unequal 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).

2Analyses of variance performed on the discrimination data yielded significant effects of stimulus location (p < .01) for each step size of the timbres.

3For the purpose of this analysis, responses to pairs of identical stimuli (indicated by squares in Figure 3) were not included.
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 "different" responses were obtained in both discrimination and labeling tasks when the first stimulus in a pair had a higher position on the continuum than 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 et al. (1979), although the reversal occurs at an earlier point on the vowel continuum in the present study.

CV 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.