

Categorical results do not imply categorical perception

JOSEPH M. HARY and DOMINIC W. MASSARO
University of California, Santa Cruz, California

Categorical perception refers to the ability to discriminate between- but not within-category differences along a stimulus continuum. Although categorical perception was thought to be unique to speech, recent studies have yielded similar results with nonspeech continua. The results are usually interpreted in terms of categorical, as opposed to continuous, perception of both speech and nonspeech continua. In contrast, we argue that these continua are perceived continuously, although they are characterized by relatively large increases in discriminability near the category boundary. To support this argument, the amplitude rise time of a tone was varied to produce either an increase or a decrease in the intensity during the initial portion of the tone. A bipolar continuum of onset times increasing and decreasing in amplitude yielded traditional categorical results. However, when only half of this continuum was tested, subjects perceived the same sounds continuously. The finding of traditional categorical results along the bipolar continuum, when the sounds were shown to be perceived continuously in another context, argues against the use of traditional categorical results as evidence for categorical perception.

A central issue in auditory information processing is whether certain auditory continua are perceived categorically rather than continuously. Continuous perception refers to a relatively linear relationship between changes in a stimulus and changes in the perceptual experience of that stimulus. The discrimination of differences among stimuli within most auditory continua indicates that these continua are, in fact, perceived continuously. In contrast, a large amount of research has been aimed at demonstrating the categorical perception of certain speech contrasts. Categorical perception refers to a mode of perception by which changes along a stimulus continuum are perceived, not in a continuous, but in a discrete manner (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970).

According to the theory of categorical perception, listeners are assumed to be limited in their ability to discriminate differences between different speech sounds that belong to the same phoneme category. The sounds within a category can be identified only absolutely, which limits discrimination of the sounds. Subjects are able to discriminate only sounds that are identified as belonging to different categories. Within the context of the motor theory of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), categorical perception has been assumed to result from the categorical nature of speech production. Different sounds are assumed to be mediated by the same articulatory set of commands, which leads to perceptual

equivalence between the sounds. This theory predicts major differences in the perception of speech and nonspeech sounds.

The test distinguishing between continuous and categorical perception usually involves the observed relationship between the identification and discrimination of sounds along a stimulus continuum. For example, Liberman, Harris, Hoffman, and Griffith (1957) used the pattern playback to generate a series of 14 consonant-vowel syllables progressing from /be/ to /de/ to /ge/ (/e/ as in *gate*). To produce the continuum, the onset frequency of the second formant transition of the initial consonant was changed in equal steps. In the identification task, observers identified random presentations of the sounds as b, d, or g. The discrimination task used the ABX paradigm. Three stimuli were presented in the order ABX; A and B always differed and X was identical to either A or B. Observers were instructed to indicate whether X was equal to A or B. This judgment was supposedly based on auditory discrimination, in that observers were instructed to use whatever auditory cues they could perceive.

Liberman et al. tested the hypothesis that listeners can discriminate stimuli only to the extent that they can recognize them as belonging to different phoneme categories. The hypothesis was quantified in order to predict discrimination performance from identification judgments. According to this formalization, stimuli can be discriminated only to the extent that they are identified as different. The original results, and many later findings, were interpreted as being consistent with the hypothesis of categorical perception. In contrast, the present paper argues that these traditional "categorical results" cannot be taken as evidence for categorical perception.

The preparation of this paper was supported in part by NIMH Grant MH-35334. The authors' mailing address is: Program in Experimental Psychology, Clark Kerr Hall, University of California, Santa Cruz, California 95064.

Massaro and Oden (1980) discuss two possible explanations of categorical perception results, explanations that do not depend on categorical perception. On the one hand, the well-known contribution of memory limitations in the discrimination task can lead to categorical perception results (Massaro, 1975a, 1975b; Paap, 1975; Pisoni & Lazarus, 1974; Fujisaki & Kawashima, Note 1). Results consistent with categorical perception are also obtained whenever a perceptual discontinuity is present, that is, when a single acoustical continuum gives rise to both relatively continuous and relatively discrete perceptual dimensions. For example, Pastore, Ahroon, Baffuto, Friedman, Puleo, and Fink (1977) asked subjects to identify and discriminate loudness differences in the presence or absence of a fixed background reference tone. When test tones of different amplitudes were presented in silence, there was no relationship between identification and discrimination. When the reference tone was present, the test tones had amplitudes slightly greater or smaller than that of the reference tone. The reference tone caused a sharp boundary in the identification function and a corresponding peak in the discrimination function, as predicted by categorical perception. These results indicate that, although listeners were able to discriminate whether the test tone was louder or softer than the reference tone, they were not able to discriminate the magnitude of difference. The results showed good discrimination between the categories "louder" and "softer" and poor discrimination within each category. Accordingly, sounds perceived as continuous in a background of silence appeared to be perceived categorically in the context of a reference tone. A perceptual discontinuity created by discriminating relative loudness with respect to the reference tone overshadowed the discrimination of absolute loudness.

Given the importance of perceptual discontinuities for categorical results, certain discrepancies are noticed in the work of Cutting and Rosner (1974) and Cutting, Rosner, and Foard (1976). They found results consistent with categorical perception by varying only the amplitude rise time of tones. The rise time was varied in 10-msec steps from 0 to 80 msec. Their subjects identified the sounds as "pluck" or "bow" and discriminated the sounds in an ABX task. Sounds with rise times shorter than 40 msec were predominantly identified as plucks, whereas sounds with longer rise times were identified as bows. Discrimination followed identification in that the sounds most accurately discriminated were those that crossed the 40-msec boundary. The existence of a perceptual discontinuity in the 40-msec region does not seem reasonable, since psychophysical research has shown that the best discriminability of rise time occurs at the shortest rise times (van Heuven & van den Broecke, 1979).

Although the rise-time continuum seems to be a case of categorical perception without the existence of a perceptual discontinuity, Rosen and Howell (1981),

using stimuli generated on line in real time by a computer system, could not replicate the results of Cutting and Rosner (1974) and Cutting et al. (1976). These computer-generated sounds produced best discrimination at the shortest rise times and a decrease in discriminability with increases in rise time, which accords with the prediction of Weber's law. Further investigation revealed that the rise time of Cutting and Rosner's continuum did not vary by constant increments of 10 msec across all levels. In fact, the greatest difference in rise time straddled the 40-msec level, the very point at which Cutting and Rosner found the peak in their discrimination function. In terms of the perceptual discontinuity idea, Cutting and Rosner inadvertently created such a discontinuity by introducing a large stimulus difference in the middle of the continuum. Hence, categorical perception results were found.¹

The major goal of these three experiments was to demonstrate that categorical perception results do *not* necessarily mean that perception is truly categorical. Although listeners may be sensitive to continuous changes along a stimulus continuum, they might produce categorical results if the continuum were characterized by a perceptual discontinuity. The discrimination of small differences on either side of the perceptual discontinuity is overshadowed by the large difference in discrimination across the perceptual discontinuity. In both the Cutting and Rosner (1974) and Pastore et al. (1977) experiments, the potential discrimination within a category was not apparent because of a large discriminable difference between categories. A perceptual discontinuity between categories can overshadow the continuous perception of differences within a category. If this is the case, the traditional identification-discrimination task is not a valid measure of categorical perception. Our goal was to show that categorical results could be found with a continuum of sounds shown to be perceived continuously. If we were successful, the identification-discrimination task would have to be rejected as a valid measure of categorical perception. We first evaluated the previous results of Cutting and Rosner (1974) in terms of the perceptual discontinuity concept and then modified the continuum to provide a strong test of the validity of the identification-discrimination task.

EXPERIMENT 1

In Experiment 1, we addressed the question of the categorical perception of rise time. This issue was sufficiently important to justify a second, independent replication: additional study not only might illuminate the perceptual processing of rise time, but also might further our understanding of the perception of speech and nonspeech continua. In addition, we modified the traditional discrimination procedure to reduce auditory memory contributions. Instead of using an ABX task, we used a two-interval AX task. This procedure should

have encouraged the use of the auditory dimension and, therefore, attenuated the effects of an abstract memory code (Massaro, 1976; Paap, 1975; Pisoni, 1973; Fujisaki & Kawashima, Note 2). To prevent the possible corruption of the stimuli by the recording of the sounds on tape (van Heuven & van den Broecke, 1979) we used a computer system to generate our sounds on line in real time.

The experiments tested whether the dimension of rise time was indeed categorically perceived. If it was, then our stimuli should have elicited discrimination functions similar to the classic form outlined by Studdert-Kennedy et al. (1970). However, if the results found with Cutting and Rosner's continuum were due to the inaccuracy of the stimuli and their recording method, we would have expected our data to be consistent with Rosen and Howell's (1981) results. If Weber's law describes the perception of rise time (van Heuven & van den Broecke, 1979), discrimination should be most acute around the instantaneous rise time and fall off with increases in rise time.

Method

Subjects. Eleven female and male subjects from the University of California, Santa Cruz, participated for about 1 h as one of several alternatives for credit in an introductory psychology course.

Stimuli and Procedure. The characteristics of our stimuli replicated those of Cutting and Rosner (1976, Experiment 2) as much as possible. Each sound stimulus was a 294-Hz sawtooth wave intended to simulate the sound of a stringed instrument. Nine levels of the stimulus were generated by varying the rise time from 0 to 80 msec in 10-msec intervals. The sound with a 0-msec rise time reached full amplitude within 1/4 cycle. The stimuli were set to decay to zero amplitude 1,020 msec after completion of the rise time. Each stimulus was terminated after 750 msec. A minor amplitude fluctuation in the steady-state portion of our stimuli, due to round-off error in the sound generation program was not perceptually significant.

The subjects participated in an identification session and a two-interval AX discrimination session. The sessions were counterbalanced across subjects so that five subjects first participated in the AX discrimination session and six subjects initially participated in the identification session.

The subjects were familiarized with identifying the stimuli as plucks or bows before the identification session. These practice trials began with the presentation of the end points of the continuum repeated 10 times in alternation. Immediately after the presentation of each end point stimulus, the subjects were told whether the sound was a pluck (0-msec rise time) or a bow (80-msec rise time). The end points were followed by three blocks of nine trials, sampled randomly without replacement from the nine sounds along the stimulus continuum. On each trial, the subjects were presented with the stimulus, followed by a 3-sec response interval. The subjects identified the stimulus as a pluck or a bow by pressing the appropriate key on a computer terminal keyboard.

The identification session consisted of 36 blocks of nine trials. Within each block, stimuli were sampled without replacement from the nine levels of the rise-time continuum. Unknown to the subjects, their responses to the first three trial blocks were not collected. Thus, 33 identifications were collected for each level of the sound continuum.

In the discrimination session, the stimuli were paired in a two-step AX paradigm. Nine same (AA) pairs were created by pairing each stimulus with itself. Different (AB) pairs were constructed by pairing sounds differing in rise time by 20 msec

to form 14 comparisons: 0-20, 20-0, 10-30, 30-10, 20-40, 40-20, 30-50, 50-30, 40-60, 60-40, 50-70, 70-50, 60-80, and 80-60. The presentation of the A stimulus signified the beginning of a trial; the X stimulus followed after 500 msec of silence. The subjects had 3 sec to identify the two sounds as either "same" or "different." The subjects were told that half of the time the sounds would be the same and half of the time the sounds would be different. A total of 272 trials were presented. The responses to the first 20 trials were not collected, and sampling without replacement began anew on the 21st trial. Each AA pair was repeated 14 times and every unique AB or BA pair was repeated 9 times, for a total of 126 "same" and 126 "different" trials.

Stimuli were generated by the program MUSIC-11 (Vercoe, Note 3) at a sampling rate of 10 kHz and stored in files on disk. During the experiment, stimuli were played back at 10 kHz by a 12-bit D-to-A converter (Data Translation Model 1711) and low-pass filtered (Krohn-Hite Model 3500) to 4.2 kHz. Stimuli were then amplified (McIntosh Model MC-30) and presented to the subjects at 67 dB SPL over Grason-Stadler TDH-49 headphones. All experimental events and data collection were controlled by a PDP-11/34 computer.

Results

The five subjects who participated in the discrimination session before identifying the items were eliminated from the data analysis because they could not discriminate any differences in the AX task. Figure 1 plots the percentage identification responses and the percentage of correct discrimination of the remaining six subjects. The percentage of "pluck" identifications decreases systematically with increases in rise time. In contrast with the sharp identification function characteristic of categorical perception, the identification function falls off relatively continuously as rise time increases.

Discrimination performance was measured in terms of both average percentage correct and *d'* values com-

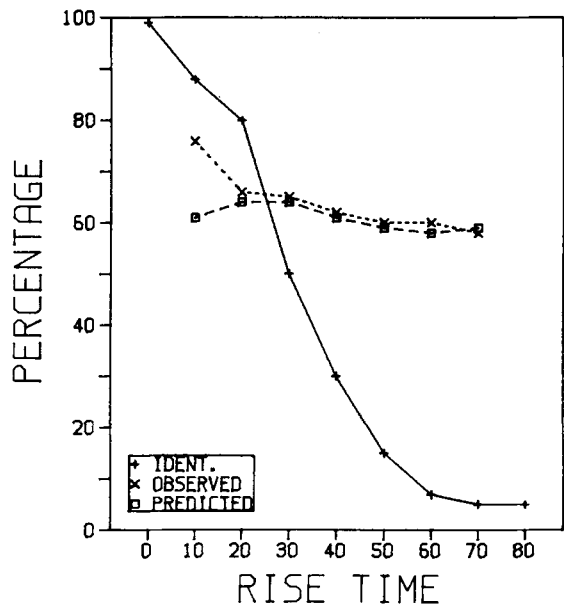


Figure 1. Percentage of "pluck" identifications and observed versus predicted percentage correct discrimination as a function of rise time in milliseconds (Experiment 1).

puted from the average response probabilities. Since the presentation of the stimuli was organized to present the same number of "same" and "different" trials, the number of replications of a given "same" comparison (i.e., 30-30) was less than the number of replications of a specific "different" comparison (i.e., 30-60). Therefore, in calculating the percentage correct, the percentage "same" and percentage "different" responses were weighted appropriately to reflect this difference. Depending on the subject's criterion to respond "same," chance performance would be between 39% and 61% correct. Therefore, d' , which provides a measure of discrimination that is relatively independent of the criterion, is an important additional measure of discrimination in this task.

The discrimination function in Figure 1 indicates that the subjects were most accurate in discriminating differences in the 0-20-msec comparison. Discrimination performance decreased continuously with increases in the rise time of the stimuli to be discriminated. Figure 2 plots the discriminability of the sounds in d' units. These data make more apparent the gradual decrease in discriminability as the duration of rise time increases. Therefore, the subjects appear to have perceived this continuum continuously in a manner roughly consistent with the predictions of Weber's law.

A test of categorical perception is the ability to predict the discrimination data from the identification data (Liberman et al., 1967; Studdert-Kennedy et al., 1970). Predicted percentages of discrimination were computed using the algorithm given by Massaro (see 1975a, pp. 513-515). In the AX task, the predicted probability of a correct response, $P(C)$, is $P(C) = P(S)*P$

(same/same) + $P(D)*P$ (different/different), where $P(S)$ is the probability of a "same" trial, P (same/same) is the probability of a "same" response to a "same" trial, and so forth, analogously for the different trials. The probability of a "same" response is predicted to be equal to the probability that the subject identifies the two stimuli with the same name, whereas the probability of a different response is equal to the probability of the two stimuli's being identified with different names. Predicted d' values were also computed by simply using the appropriate identification probabilities averaged across subjects.

The predicted percentage correct function, averaged over six subjects, is plotted in Figure 1. The algorithm predicts very poor performance at the 0-20-msec comparison, the point at which actual performance was the most accurate. The poor match between the observed and predicted discrimination is most clearly seen in the d' values shown in Figure 2. The predictions underestimated discrimination performance, especially at the short rise times. Also, the predicted improvement in discrimination near the category boundary was not present in the observed results.

The observed and predicted percentage correct discrimination values for each of the six subjects were submitted to an analysis of variance, with predicted versus observed values, the seven discrimination levels, and subjects as variables. The analysis produced a significant interaction between discrimination level and predicted versus observed values [$F(6,30) = 2.76, p < .05$]. As can be seen in Figures 1 and 2, the predicted values underestimated observed discrimination at short rise times. In contrast with the predictions of categorical perception, predicted discrimination did not match well with observed performance.

The results are consistent with the identification and discrimination functions of Rosen and Howell (1981), who failed to replicate the Cutting et al. (1976) results. Also consistent with Rosen and Howell's observations are the reports by most of our subjects that they listened to the relative sharpness of the sounds in the discrimination task and judged the sounds on this basis. In addition, the fact that discrimination was best around the 10-msec rise time is consistent with the van Heuven and van den Broecke (1979) study. Their subjects were most accurate in reproducing rise times of short durations, as expected from Weber's law.

EXPERIMENT 2

The results of the first experiment revealed that the rise-time continuum is perceived continuously. The goal of Experiment 2 was to show that the same rise-time continuum could be made to be perceived categorically when a perceptual discontinuity was introduced. Consider the negative rise-time continuum illustrated in Figure 3. Although all sounds with negative rise times reach maximal amplitude in $\frac{1}{4}$ cycle, the amplitude of

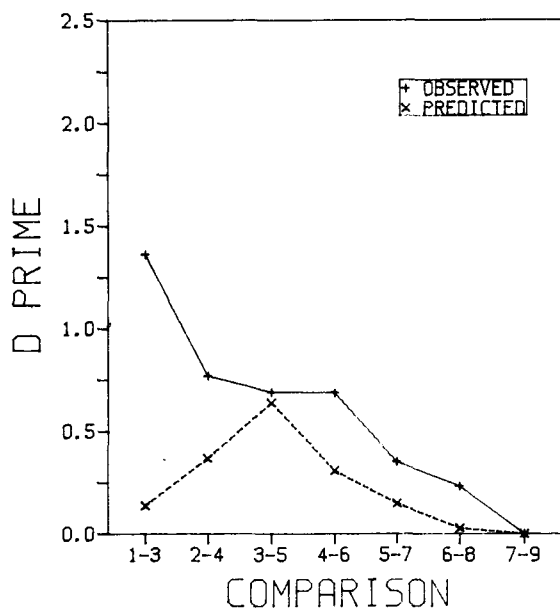


Figure 2. Observed and predicted d' values as a function of comparison levels for Experiment 1. Comparison levels range from 1-3 (0-20 msec) to 7-9 (60-80 msec).

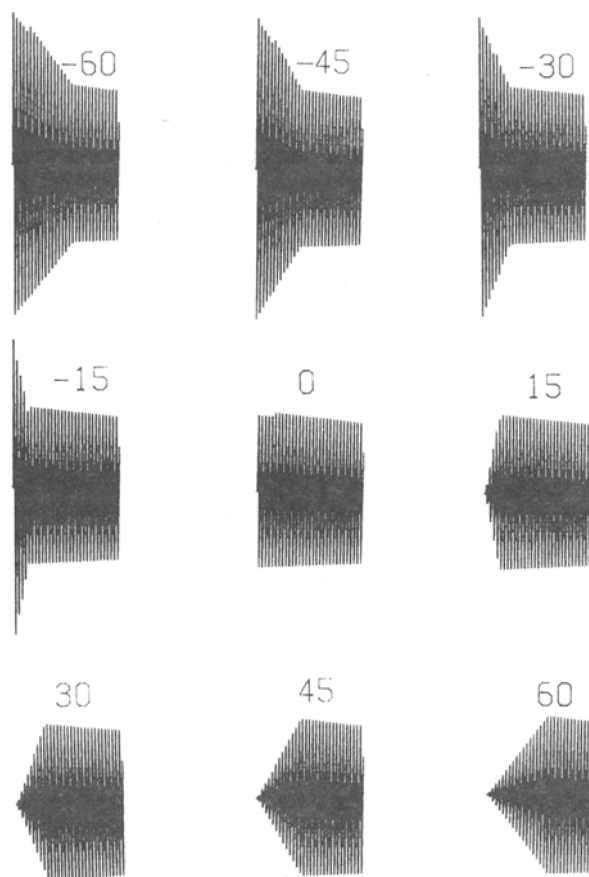


Figure 3. Illustration of the first 100 msec of each of the sounds along the bipolar rise-time continuum. The rise time in milliseconds is given at the top of each of the sounds.

the sounds falls to a steady-state amplitude at varying rates to create a continuum of negative rise times. By combining the sounds along this negative rise-time continuum with a continuum of positive rise times, we might create a perceptual discontinuity. A highly discriminable category boundary around a rise time of 0 msec should be found. Stimuli that have negative or 0-msec rise times should sound very "pluckish," whereas stimuli with positive times should sound "bowish." This discontinuity should produce a sharp category boundary and a corresponding peak in the subject's discrimination function.

Experiment 2 was a direct replication of Experiment 1, but with the rise-time continuum extended in positive and negative directions to create a "bipolar" continuum. In addition, the standard positive continuum was also tested to provide a direct comparison between these two types of continua. We predicted that categorical discrimination functions should be observed with the bipolar continuum but not with the standard continuum of rise times. This result would support the hypothesis that traditional categorical perception results can be found along a continuum of sounds that is perceived continuously.

Method

Subjects. Fifteen female and male subjects were recruited through an ad in the campus newspaper. They were paid \$5.00/h to participate for 1 h a day in the 2-day experiment.

Stimuli and Procedure. A bipolar variation of the pluck-bow continuum was generated. This continuum consisted of four tones with negative rise times, one 0-msec rise-time tone, and four tones with positive rise times. Each negative and 0-msec rise-time tone started at the zero crossing and reached a maximal amplitude within $\frac{1}{4}$ cycle. Four levels of negative rise time were created by varying the fall time between 15 and 60 msec in 15-msec intervals from a maximal amplitude of 73 to 67 dB SPL. The tones with positive rise times were similar to those used in Experiment 1, except that their rise times varied from 15 to 60 msec in 15-msec intervals to a maximal amplitude of 67 dB SPL. After the completion of the rise or fall time, the tones decayed to zero amplitude over 1,020 msec. As in Experiment 1, the stimuli were generated at 294 Hz and terminated after 750 msec. Figure 3 illustrates the first 100 msec of the nine sounds along the bipolar continuum.

The standard continuum was modeled after that in Experiment 1. Nine levels of the stimulus, varying from 0 to 120 msec in 15-msec steps, were generated. All other stimulus parameters were identical to those used in Experiment 1 and in the generation of the bipolar continuum.

All subjects were tested with both the bipolar and the standard continuum. Each continuum was tested on a different day, and the order of presentation was counterbalanced across subjects. On each day, the subjects participated in both an identification session and a discrimination session. Since the first two subjects who participated in the discrimination session before the identification session could not discriminate above chance on either continuum, we eliminated these subjects and subsequently decided to begin each day with the identification session. The parameters of the identification and discrimination sessions were identical to those in Experiment 1. Due to the 15-msec step size, there was a 30-msec rise-time difference between the tones on different comparisons in the AX task.

Of the 15 subjects who participated in the experiment, 7 had to be eliminated for various reasons: 2 because they initially had participated in the discrimination session, 1 because of equipment failure, and 4 for responding randomly or for missing a significant number of responses in any condition.

Results

Figure 4 gives the average identification, percentage correct discrimination, and predicted discrimination functions for the standard condition. These functions are highly similar to those obtained in Experiment 1, except for an overall improvement in discrimination performance for the 15-msec step size in the present experiment. Both the identification and discrimination functions indicate that the subjects perceived the continuum in a continuous manner. The subjects discriminated the 0-30-msec comparison most accurately, with their performance falling off sharply over the remaining levels. These results are most consistent with continuous perception and the applicability of Weber's law to the rise-time continuum.

Figure 5 presents the average identification, discrimination, and predicted discrimination functions for the bipolar condition. The identification function shows a very sharp drop in the percentage of "pluck" responses between the 0- and 15-msec rise times. This stands in marked contrast with the relatively continuous drop in the identification function along the standard con-

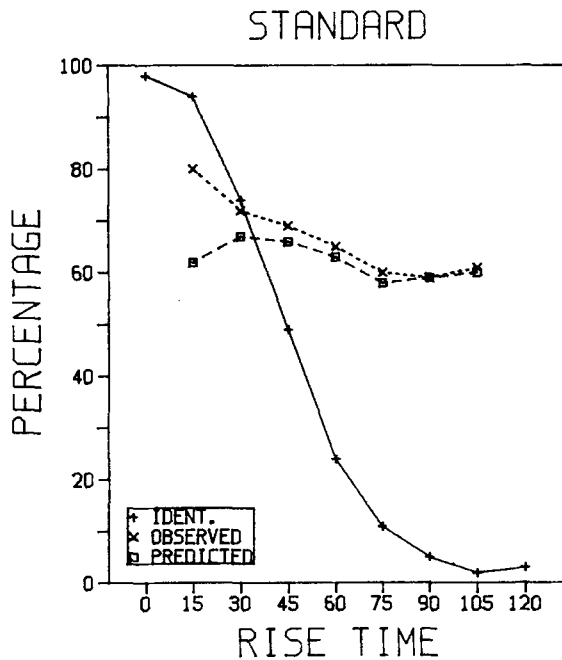


Figure 4. Percentage of "pluck" identifications and observed versus predicted percentage correct discrimination results for the standard continuum (Experiment 2).

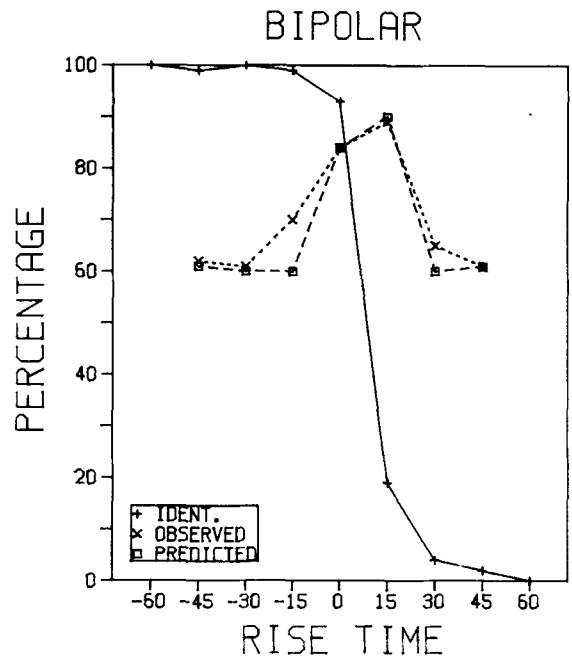


Figure 5. Percentage of "pluck" identifications and observed versus predicted percentage correct discrimination results for the bipolar continuum (Experiment 2).

tinuum. In addition, the bipolar continuum produced discrimination functions highly representative of traditional categorical perception. The requisite peaks and troughs for categorical perception are present in the discrimination function.

Figure 6 gives the observed and predicted discrimination performance in terms of d' values. The results for the standard continuum replicated what was found in Experiment 1. For the bipolar continuum, the predicted

d' values tend to underestimate the observed d' values, although both the predicted and observed functions follow the general form of categorical perception results (Studdert-Kennedy et al., 1970).

The predicted and observed percentage correct discrimination performance for the standard and bipolar continua were submitted to an analysis of variance, with the standard versus bipolar continua, discrimination levels, predicted versus observed discrimination scores,

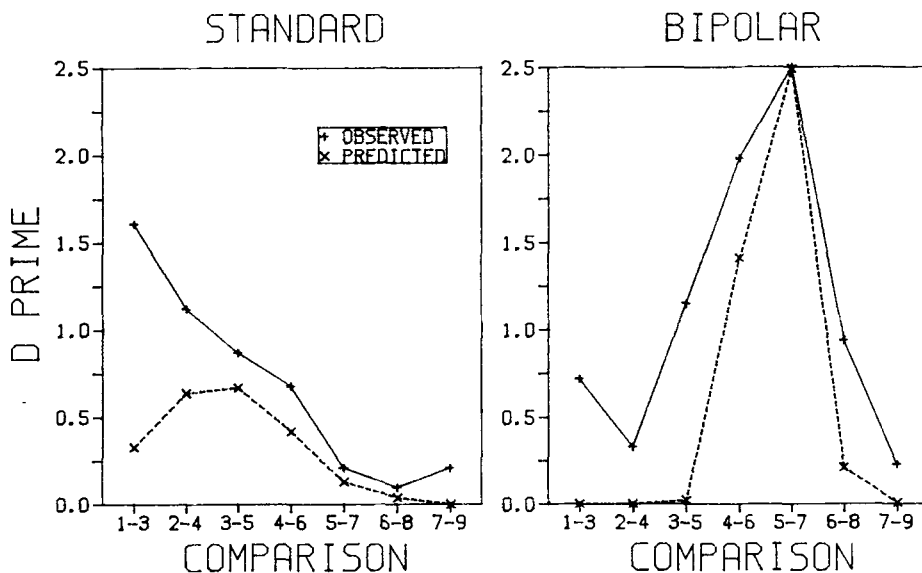


Figure 6. Observed and predicted d' values as a function of comparison level for the standard and bipolar continua of Experiment 2. For the standard continuum, comparison levels range from 1-3 (0-30 msec) to 7-9 (90-120 msec). For the bipolar continuum, comparison levels range from 1-3 (-60-30 msec) to 7-9 (30-60 msec).

and subjects as factors. Discrimination performance was significantly different for the two continua [$F(1,7) = 14.10, p < .01$]. This result and the triple interaction of continua, discrimination level, and observed versus predicted values [$F(6,42) = 5.90, p < .001$], reflects the large differences that were observed for the two continua. The form of the predicted function describes the observed discrimination much better with the bipolar continuum than with the standard continuum. We stress that the good match between predicted and observed performance for the bipolar continuum does not imply that the subjects could not hear within-category differences among stimuli. Models of continuous perception should be able to reproduce both the categorical and continuous functions found in this experiment.

EXPERIMENT 3

Although the results of Experiments 1 and 2 were consistent with our predictions, they were limited to the subjects who performed the identification task before the discrimination task. The subjects who initially participated in the discrimination session could not discriminate differences between the sounds and had to be eliminated from the data analysis. We felt that the poor discrimination was related to simply the overall difficulty of the discrimination task. Therefore, it was necessary to demonstrate that the results of subjects who initially participate in the discrimination session replicate the observed results. To decrease the difficulty of the discrimination task, we decided to provide feedback and to use a three-step difference between sounds in the discrimination task. In addition, we hypothesized that an easier discrimination task would improve discrimination within a category even with the bipolar continuum. This result would show that discrimination is possible within a category and that poor within-category discrimination results do *not* necessarily reflect categorical perception.

Method

Subjects. Fourteen female and male subjects from the university participated for about 1 h as one of several alternatives for credit in an introductory psychology course.

Stimuli and Procedure. The bipolar and standard continua from Experiment 2 were used. The identification session was identical to those used in Experiments 1 and 2.

In the discrimination session, the stimuli were paired in a three-step AX paradigm. As in Experiments 1 and 2, 9 same (AA) pairs were created by pairing each stimulus with itself. For both the standard and bipolar continua, 12 different (BA or AB) three-step pairs were created by pairing sounds differing in rise time by 45 msec. The presentation of the A stimulus signified the beginning of a trial. The second stimulus of the AX pair followed after 500 msec of silence. The subjects had 3 sec to identify the sounds as "same" or "different." After the 3-sec response interval, the subjects were informed of the nature of the sounds by flashing either the word "same" or "different" for 500 msec on the computer terminal. A new trial started 500 msec after the completion of the feedback interval. The subjects were told that 50% of the time the sounds would be the

same and 50% of the time the sounds would be different. In all, 236 trials were presented. As in the previous experiments, the first 20 trials were not included in the data analysis. Each "same" (AA) pair was repeated 12 times and each "different" (BA or AB) pair was repeated 9 times, for a total of 108 "same" and 108 "different" trials.

Seven subjects were tested under the standard condition, and seven subjects were tested under the bipolar condition. Under both conditions, all subjects participated in the discrimination session first. All other conditions of the experiment were identical to those in Experiment 2.

Results

Figures 7 and 8 display the average identification, percentage correct discrimination, and predicted discrimination functions for the standard and bipolar conditions, respectively. The major differences between the standard and bipolar conditions observed in Experiment 2 were replicated in the present comparison. For the standard continuum, "pluck" responses decreased more rapidly with increases in rise time than had been observed in Experiment 2. The subjects discriminated comparisons with short rise times better than they did those with long rise times. The relatively good discrimination for the 75-120-msec (stimulus levels 6-9) comparison might reflect a discrimination advantage for an end point stimulus along a continuum (Samuel, 1977). For the bipolar condition, the results were very similar to the bipolar condition of Experiment 2, except for some overall improvement in discrimination performance.

The observed and predicted d' values are given in Figure 9. This measure of performance makes apparent the very good discrimination within a category in the bipolar continuum. Thus, an easier discrimination task reveals that the categorical results found with the bipolar continuum in Experiment 2 cannot be taken as evidence for categorical perception.

The percentage correct and d' values sometimes gave very different measures of performance, especially in terms of the differences between predicted and observed performance within a category in the bipolar continuum. Consider the comparison between rise times of -60 and -15 msec, shown as the leftmost points in Figure 8 and in the bipolar continuum in Figure 9. The percentage correct scores were 60% and 56% for the observed and predicted values, respectively. The corresponding d' values were 1.26 and 0. Thus, the superiority of observed discrimination over predicted discrimination is highly apparent in the d' values, but not in the percentage correct values. The d' measure appears to be the more appropriate measure, since the subjects tended to respond "same" most of the time when faced with a relatively difficult discrimination task. The subjects appear to have listened for a difference and responded "same" unless they were fairly confident that they had heard a difference. If a subject had responded same 98% and 78% of the time on "same" and "different" trials, a percentage correct score would be 60% and the d' score would be 1.28. Thus, this decision bias can result in a relatively poor discrimination score when measured by a percentage correct, relative to a d' score.

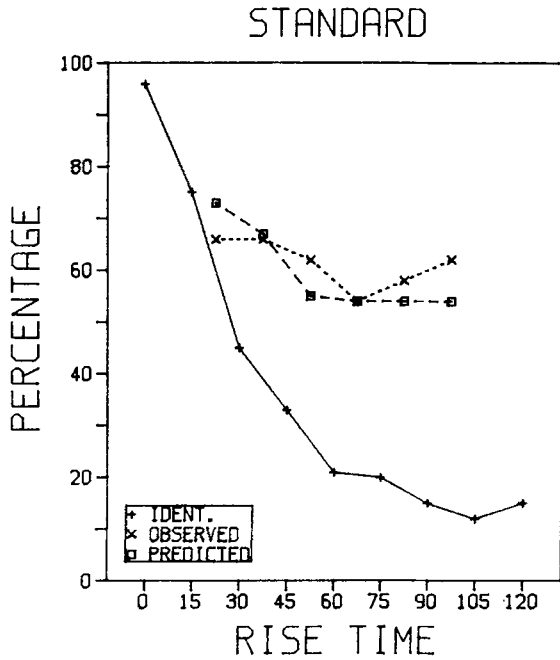


Figure 7. Percentage of "pluck" identifications and observed versus predicted percentage correct discrimination results for the standard condition (Experiment 3).

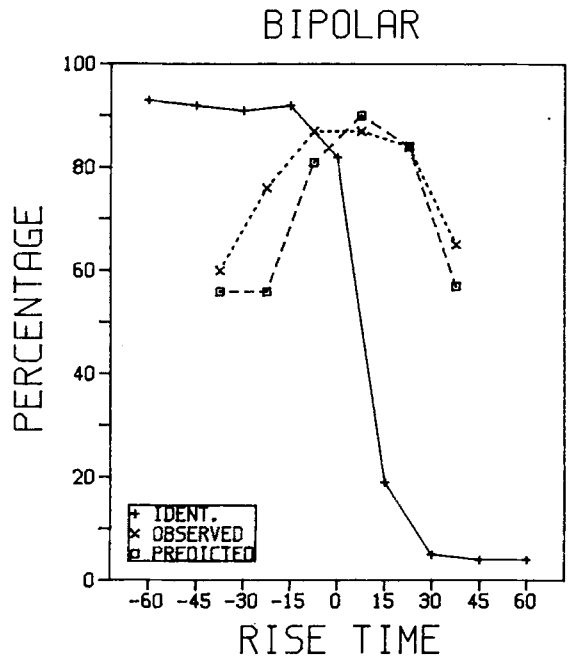


Figure 8. Percentage of "pluck" identifications and observed versus predicted percentage correct discrimination results for the bipolar condition (Experiment 3).

GENERAL DISCUSSION

We have demonstrated that a modification of Cutting and Rosner's (1974) stimuli that is guided by the principle of a perceptual discontinuity produces functions consistent with the predictions of categorical perception.

When subjects are presented with a continuum of sounds, a large change in relative discriminability along the continuum makes apparent two perceptual categories. This discontinuity along the continuum need not be absolute. Subjects might still discriminate differences within categories, but not at the level of differences between them.

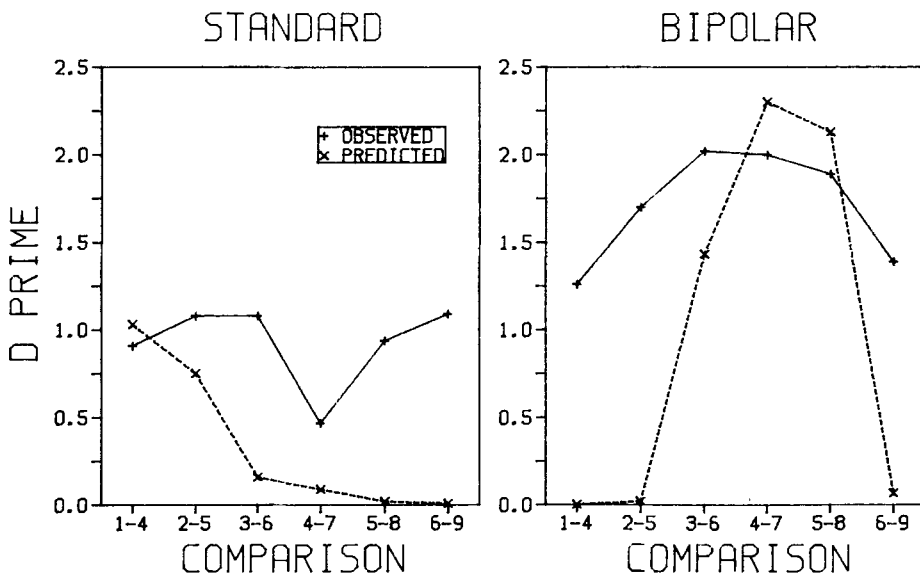


Figure 9. Observed and predicted d' values as a function of comparison level for the standard and bipolar continua of Experiment 3. For the standard continuum, comparison levels range from 1-4 (0-45 msec) to 6-9 (75-120 msec). For the bipolar continuum, comparison levels range from 1-4 (-60 -15 msec) to 6-9 (15-60 msec).

The categorical results of Cutting and Rosner's (1974) continuum might be understood in terms of a perceptual discontinuity. Due to an error in stimulus generation, the 30- and 50-msec comparison was separated by 26 msec instead of the nominal 20 msec for the low-pitched stimuli in the two-step ABX task. This comparison should have been more discriminable than any other comparison along the continuum. If so, an apparent perceptual discontinuity could have resulted, creating a boundary between two perceptual categories.

The results of previous research can be interpreted by considering the effects of a perceptual discontinuity (e.g., Miller, Wier, Pastore, Kelly, & Dooling, 1976; Pisoni, 1977). For example, in Pisoni's (1977) experiments involving the relative onset time of two tones, the perception of onset differences might be expected to have been relatively continuous. The subjects might also have distinguished a complex quality of the tone stimuli. Tones beginning very close together would have been heard as a single complex sound. Tones separated by a larger onset time interval would have been heard as a simple sound preceding a complex sound. The quality differences between relatively simultaneous tones and tones with very different onset times would be responsible for relatively easy discrimination between these two classes of items. The results were consistent with this description. The subjects showed good discrimination between categories and relatively poor discrimination within these categories.

In both Experiment 1 and Experiment 2, all subjects who performed in the discrimination session before the identification task performed at chance level across both the standard and bipolar continua. This floor effect implies that initial discrimination among these complex tone stimuli was relatively poor. The practice and identification sessions could have focused the observers' attention on the appropriate dimension of the particular "pluck" or "bow" quality of the sound. However, continuous and categorical discrimination results in the standard and bipolar continua (of Figures 7 and 8) were also found when subjects were not first exposed to the identification task. Making the discrimination task easier was sufficient to produce the same functions in Experiment 3.

As might be expected, a comparison of the identification functions for the standard and bipolar continua reveals a context dependency. Sounds with positive rise times were identified as plucks more often in the standard continuum than in the bipolar continuum. With a perceptual discontinuity between negative and positive rise times, about half of the sounds are heard as being "pluckish" in the bipolar continuum. In the standard continuum, most of the sounds are heard as "bowish." If subjects have some tendency to respond equally with the two response alternatives, it is not surprising that a fairly "bowish" sound will be identified as a pluck much of the time in the standard continuum.

The present results show that discrimination between two sounds along a continuum is highly dependent on the context of the stimulus dimension that is used in the experiment. In Experiment 2, the discrimination between 0- and 30-msec rise times averaged 80% correct within the standard continuum. The same two stimuli were discriminated accurately 89% of the time within the bipolar continuum. Similarly, the 30-60-msec stimuli were discriminated accurately 69% of the time within the standard continuum and only 61% of the time within the bipolar continuum. It seems that experience with other sounds along a continuum influences the discrimination performance between two given sounds. If subjects have auditory memory limitations, they could respond on the basis of category labels in the discrimination task. This strategy would improve discrimination between two sounds from different categories relative to two sounds from the same category. Two given sounds could change their category membership, depending on the context of the stimulus continuum. In our experiment, a sound called predominantly "pluck" in the standard continuum was called predominantly "bow" in the bipolar continuum. The differences observed between the discrimination of two given sounds as a function of the context continuum is consistent with this category membership idea. Discrimination between two sounds was better when these sounds tended to be categorized differently in the identification task.

In our experiments, we focused on a qualitative change along a continuum in order to create a traditional categorical result with sounds that can be perceived continuously. Subjects perceived these sounds continuously when rise time was varied in a single direction. Similarly, there is strong evidence that speech continua, which give rise to traditional categorical perception results, are perceived continuously (Carney, Widin, & Viemeister, 1977; Samuel, 1977; Massaro & Cohen, Note 4).

Massaro and Cohen (Note 4) asked subjects to rate the degree to which a given speech sound resembled one alternative or another. In one experiment, subjects rated consonants along a voicing continuum from /ba/ to /pa/. The task was to rate the degree to which a stimulus from the continuum was ba-like or pa-like. If the continuum were perceived categorically, a stimulus would be more likely to produce the percept /ba/ to the extent it is away from the category boundary and close to the /ba/ end of the continuum. If the continuum were perceived continuously, the percept of a stimulus close to the /ba/ end point would be more ba-like than its neighbor closer to the /pa/ end of the continuum. Both categorical and continuous perception viewpoints predict a continuous change in the rating responses with changes along the stimulus continuum. However, the two viewpoints can be distinguished on the basis of the distribution of rating responses to each stimulus along

the continuum. The distribution of rating responses indicated that subjects appeared to perceive these continua in a continuous manner. We predict that the bipolar rise-time continuum would also produce similar results when incorporated in a rating study. Given an opportunity to rate the stimuli, subjects would be able to indicate, within each of the two perceptual categories, different degrees of "pluckishness" or "bowishness."

Categorical perception has been central to the description of the processing of a number of auditory continua. Originally, categorical perception was believed to be unique to speech continua (Liberman et al., 1957). More recently, studies with nonspeech sounds have been used to demonstrate that categorical perception also occurs for certain nonspeech continua (Miller et al., 1976; Pastore et al., 1977; Pisoni, 1977). In contrast with both of these viewpoints, our argument is that categorical results found in the identification-discrimination task do not necessarily imply that perception is categorical. Given the evidence for continuous perception of auditory continua, we argue that true categorical perception is questionable for both speech and nonspeech continua.

REFERENCE NOTES

1. Fujisaki, H., & Kawashima, T. *On the modes and mechanisms of speech perception* (Annual Report of the Engineering Research Institute, Vol. 28). Tokyo: Faculty of Engineering, University of Tokyo, 1969.
2. Fujisaki, H., & Kawashima, T. *Some experiments on speech perception and a model for the perceptual mechanism* (Annual Report of the Engineering Research Institute, Vol. 29). Tokyo: Faculty of Engineering, University of Tokyo, 1970.
3. Vercoe, B. *Music II*. Cambridge, Mass: The studio for Experimental Music, M.I.T., 1976.
4. Massaro, D. W., & Cohen, M. M. *Categorical or continuous speech perception*. Unpublished manuscript, 1981. (Available from Program in Experimental Psychology, University of California, Santa Cruz, Calif., 95064.)

REFERENCES

- CARNEY, A. E., WIDIN, G. P., & VIEMEISTER, N. F. Non-categorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 1977, **62**, 961-970.
- CUTTING, J. E. Plucks and bows are categorically perceived, sometimes. *Perception & Psychophysics*, 1982, **31**, 462-476.
- CUTTING, J. E., & ROSNER, B. S. Categories and boundaries in speech and music. *Perception & Psychophysics*, 1974, **16**, 564-570.
- CUTTING, J. E., ROSNER, B. S., & FOARD, C. F. Perceptual categories for musiclike sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology*, 1976, **28**, 361-378.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. S., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 431-461.
- LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S., & GRIFFITH, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, **54**, 358-368.
- MASSARO, D. W. *Experimental psychology and information processing*. Chicago: Rand McNally, 1975. (a)
- MASSARO, D. W. *Understanding language: An information-processing analysis of speech perception, reading, and psycholinguistics*. New York: Academic Press, 1975. (b)
- MASSARO, D. W. Auditory information processing. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Attention and memory* (Vol. 4). Hillsdale, N.J: Erlbaum, 1976.
- MASSARO, D. W., & ODEN, G. C. Speech perception and framework for research and theory. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 3). New York: Academic Press, 1980.
- MILLER, J. D., WIER, C. C., PASTORE, R. E., KELLY, W. J., & DOOLING, R. J. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, 1976, **60**, 410-416.
- PAAP, K. Theories of speech perception. In D. W. Massaro (Ed.), *Understanding language: An information processing analysis of speech perception, reading and psycholinguistics*. New York: Academic Press, 1975.
- PASTORE, R. E., AHROON, W. A., BAFFUTO, K. J., FRIEDMAN, C., PULEO, J. S., & FINK, E. A. Common-factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, **3**, 686-696.
- PISONI, D. B. Auditory and phonetic codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 1973, **13**, 253-260.
- PISONI, D. B. Auditory 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.
- PISONI, D. B., & LAZARUS, J. H. Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, 1974, **55**, 323-334.
- ROSEN, S. M., & HOWELL, P. Plucks and bows are not categorically perceived. *Perception & Psychophysics*, 1981, **30**, 156-168.
- SAMUEL, A. G. The effect of discrimination training on speech perception: Noncategorical perception. *Perception & Psychophysics*, 1977, **22**, 321-330.
- 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.
- VAN HEUVEN, V. J. J. P., & VAN DEN BROECKE, M. P. R. Auditory discrimination of rise and decay times in tone and noise bursts. *Journal of the Acoustical Society of America*, 1979, **66**, 1308-1315.

NOTE

1. After the present research had been completed, Cutting (1982) acknowledged the problems with the original Cutting and Rosner (1974) experiment. In an independent replication, perception was found to be relatively continuous for a linear pluck-bow continuum. In two additional experiments, however, results approaching categorical perception were found when the sounds along the pluck-bow continuum were spaced logarithmically. Following the logic of the present paper, we do not accept these results as evidence for the categorical perception of the rise-time dimension. Changes along this dimension can be perceived continuously, and the finding of relatively good discrimination between categories and poor discrimination within a category in an ABX task is not evidence to the contrary. Similarly, an agreement between identification and discrimination performance also is not evidence for categorical perception.

(Manuscript received April 6, 1982;
revision accepted for publication September 24, 1982.)