Plucks and bows are not categorically perceived

STUART MICHAEL ROSEN
Department of Phonetics and Linguistics, University College London, London NW1 2HE, England

and

PETER HOWELL
Department of Psychology, University College London, London WC1E 6BT, England

Plucks and bows are not categorically perceived

STUART MICHAEL ROSEN
Department of Phonetics and Linguistics, University College London, London NW1 2HE, England

and

PETER HOWELL
Department of Psychology, University College London, London WC1E 6BT, England

Cutting and Rosner (*Perception & Psychophysics*, 1974, 16, 564-570) reported that nonspeech stimuli differing in rise time were categorically perceived in the same way as speech sounds. With two independently generated sets of stimuli essentially the same as those described by Cutting and Rosner, we were unable to replicate their finding that discrimination measured in an ABX task was best around 40 msec, the category boundary. We found discrimination always best at the shortest rise times, decreasing monotonically with increasing rise time. Oscillographic traces of Cutting and Rosner's original stimuli showed them not to have the intended rise times. Instead of starting with a rise of 0 msec and increasing linearly in 10-msec steps to 80 msec, the measured rise times were approximately 4, 6, 15, 19, 37, 43, 57, 66, and 76 msec, respectively. A set of stimuli having these rise times was generated. Two distinct patterns of response emerged from the discrimination task. Most subjects now showed best discrimination around 40 msec, but a few still performed best at the shortest rise times.

Some theorists have claimed that there are differences between the perception of speech and the perception of nonspeech sounds (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Central to this issue has been the notion of categorical perception, in which the ability to discriminate differences between sounds is determined solely by the ability to categorize them differently. This was meant to be in contrast with the perception of nonspeech continua in which the norm is that many more stimuli are able to be discriminated than identified (Miller, 1956). Practically speaking, the criteria for declaring a particular continuum categorically perceived are as follows (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970): (1) a sharp categorization function; (2) a peak in the discrimination function at the category boundary; and (3) troughs in the discrimination function within categories. It had been asserted that one of the most important ways in which the perception of speech sounds differs from the perception of "common or garden-variety" sounds was that only (but not all) speech sounds were categorically perceived, especially the stop consonants. Such assertions were first put to rest conclusively in the mid-70s by two demonstrations of categorical perception for nonspeech continua (Cutting & Rosner, 1974; Miller, Wier, Pastore, Kelly, & Dooling, 1976). Miller et al. (1976) used a noise and a buzz, with the onset of the noise varying from -10 to $+80$ msec with respect to the onset of the buzz. Discrimination was best when the noise led the buzz by about 16 msec, which was at about the same place as the category boundary in a labeling task. Cutting and Rosner (1974) used sawtooth waveforms varying in rise time from 0 to 80 msec and found best discrimination performance around 40 msec, the position of the category boundary when subjects were asked to label the stimuli as plucks and bows.

One way of explaining categorical perception put forward by Miller et al. (1976) and supported by Cutting and Rosner's (1974) study is through natural sensitivities in the auditory system. According to this notion, categorical perception occurs because, due to the structure of the auditory system, discrimination is better in certain regions, in at least some continua, than in others. This idea became especially popular because both Miller et al. (1976) and Cutting and Rosner (1974) demonstrated peaks in the discrimination function at the category boundary before any categorization training was done. Since the stimuli used in these studies would not ordinarily be labeled, it seemed unlikely that the discrimination peaks were due to implicit categorization

0031-5117/81/060156-13$01.55/0

by the subjects, as might be the case in studies involving speech sounds—thus, the notion of *natural* boundaries.

We began work in this area with the intention of exploring range effects in such continua. We felt that the magnitude of such effects in categorization, especially as compared with those obtained with stop consonants (Brady & Darwin, 1978; Rosen, 1979), as well as possible range effects in discrimination (which have not been studied), would shed light on the "naturalness" of the categories and the role of previously learned labels. We chose to use Cutting and Rosner's (1974) rise-time continuum, partially arbitrarily, but also because we felt that it might be more liable to range effects, since there were no psychoacoustical results that pointed to natural sensitivities at 40 msec. This is in contrast to the way in which Hirsh's (1959) work related to Miller et al.'s (1976) results. In fact, at least one study of the perception of rise time (van Heuven & van den Broecke, 1979) indicated best discrimination performance at the shortest rise times, decreasing monotonically with rise time. Before setting off on range effects, then, the obvious need was to construct a set of stimuli like Cutting and Rosner's (1974) to verify their findings.

EXPERIMENT 1:
CATEGORIZATION AND DISCRIMINATION WITH A RISE-TIME CONTINUUM

Method

Subjects. Three naive undergraduate students, two females and one male, served as subjects. They were paid for their participation.

Stimuli. Because van Heuven and van den Broecke (1979) reported difficulty in the accurate reproduction of rise times from audio tape, our sounds were generated by analog equipment under computer control.

Nine stimuli were constructed in which only the rise time of the amplitude envelope differed. The nominal stimulus envelope consisted of a linear rise (0 to 80 msec in 10-msec steps) followed by a 20-msec steady state and a linear decay to zero amplitude over 1.25 sec. These were synthesized in a DEC PDP-12 computer and output at run time through a 10-bit D-A converter with a sampling frequency of 8 kHz. This envelope was filtered with a simple RC filter with a time constant of .5 msec and fed to one input of a precision linear multiplier (Burr-Brown 4213BM), at the other input of which was a 312-Hz square wave from a hardware signal generator. Since leakage caused the square wave to be heard faintly even when it was nominally off, two multipliers were cascaded. The output of the first multiplier was fed to the input of another multiplier, at which point it was multiplied again by the filtered stimulus envelope from the computer. This prevented audible leakage. In order to obtain the proper stimulus envelope through this cascade arrangement, the actual envelope output by the computer was not the desired envelope, but its square root. The actual rise times differed slightly from the nominal, primarily due to the low-pass filtering of the envelopes. Inspection of the generated stimuli showed the nominal 0-msec rise to be 2.3 msec. All the other rise times were slightly longer than specified (within 1.6 msec). The final output was amplified by a

Revox B77 tape recorder and fed at a comfortable constant level to HD414X headphones in a quiet room. The frequency response of the headphones measured on real ears with a probe tube was flat to ±7 dB from 250 Hz to 10 kHz. These stimuli are slightly different from those used by Cutting and Rosner in the existence of the 20-msec plateau, a decay of 1.25 sec instead of 1.02 sec, and the use of 312-Hz square waves rather than 294- and 440-Hz sawtooth waves.

Procedure. Subjects performed all sessions individually in a quiet room. They indicated their judgments by way of two response keys. The computer handled all details of stimulus presentation, response handling, and data analysis.

Two basic tasks were performed by the subjects. In the categorization task, the subjects were asked to label the stimuli as "pluck" or "bow". No exemplars were played beforehand. The nine stimuli were presented 10 times each in a random order subject to the condition that all stimuli were presented before any were repeated. The discrimination task was of the ABX type. The subjects heard a triplet of sounds, the first two being different, and were asked to indicate whether the third sound was more like the first or the second. No feedback was given. In an experimental session, three runs of the four possible orders of each of the pairs to be discriminated were presented, again in a random order subject to the constraint that all triplets occur before any were repeated. This gave 12 observations per session per stimulus pair.

The subjects performed one session of categorization followed by one session of discrimination per day for 4 days. The discrimination was two-step (20 msec difference between stimuli to be discriminated) with a 500-msec interstimulus interval (offset to onset) between sounds of the triplet. The first day's responses were considered practice and were discarded.

Results and Discussion

Figure 1 shows the mean categorization and discrimination curves as a function of rise-time duration. Categorization is relatively sharp with the category boundary at about 30 msec. There is no hint, however, of a peak in the discrimination function at this point. Discrimination is always best at the shortest rise times, decreasing monotonically with increasing rise time.

These results are consistent with those of van Heuven and van den Broecke (1979), who measured the variability of settings in a rise time reproduction task. They found that the standard deviation of the adjustments was an increasing linear function of rise time. Since it is normally the case that standard deviations in a reproduction task are linearly related to the difference limens measured in a two-interval two-alternative forced choice (2I-2AFC) task (Cardozo, 1965; Wier, Jesteadt, & Green, 1976) and that the transformation of percent correct in a 2I-2AFC task to percent correct in an ABX task is monotonically increasing (Macmillan, Kaplan, & Creelman, 1977), the constant difference between stimuli in the ABX task should be easiest to discriminate at the shortest rise times, decreasing monotonically with increasing rise time. Van Heuven and van den Broecke (1979) felt that the differences between their results and Cutting and Rosner's could be attributed to the differences in stimulus generation techniques. Their
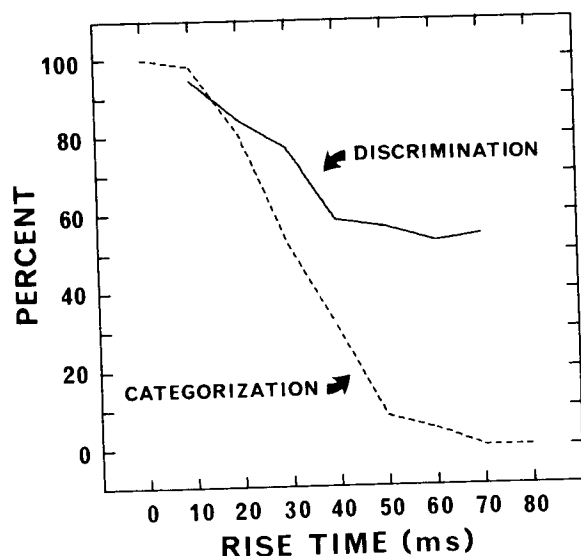
Figure 1. Categorization and discrimination functions for a rise-time continuum. For the categorization curve, ordinate values give the percentage of "pluck" responses for each stimulus. For the discrimination curve, the values give percent correct. The discrimination was two-step ABX (500-msec interstimulus interval), with the results for a particular stimulus pair plotted at the midpoint of the pair. Each point in the categorization curve represents 90 observations, while those in the discrimination curve represent 108 observations.

stimuli were generated on the spot, whereas Cutting and Rosner's were first recorded on audiotape, filtered, digitized, reconverted to analog waveforms, and filtered again for the final recording on audiotape.

EXPERIMENT 2:
TWO FURTHER DISCRIMINATION CONDITIONS

We attempted to determine the reasons for the discrepancy between the results of Cutting and Rosner and our own regarding the shape of the discrimination function. In order to test van Heuven and van den Broecke's (1979) claim that Cutting and Rosner's method of stimulus presentation was responsible for the difference in results, we included a condition in which the stimuli were recorded before presentation to the subjects. Also, since lengthening the interstimulus interval has been found to make the discrimination function more categorical[1] (Pisoni, 1973), we changed the duration between the sounds of the triplet to 2 sec.

Method

Subjects. The same three subjects as in Experiment 1 participated.

Stimuli. The stimuli were the same as those described for Experiment 1.

Procedure. A modification of the ABX discrimination procedure described in Experiment 1 was used. The interstimulus interval between the triplets of a trial was lengthened to 2 sec.

Discrimination was one-step (10 msec difference between stimuli to be discriminated) to prevent ceiling effects as the subjects became more practiced. The subjects ran under two conditions. First, in the "direct" condition, they heard the stimuli as described above for two sessions. After this, they ran one session in the "recorded" condition. Instead of routing the stimuli directly to the tape recorder we were using as an amplifier, we low-pass filtered them twice in cascade at 4 kHz with a combined roll-off of 96 dB/octave. Furthermore, instead of using the tape recorder simply as an amplifier, we set a blank tape recording and monitored the now-recorded stimuli. The computer still controlled the experiment as before, the only difference being that there was a slight delay (about 200 msec) between the time the computer actually played the stimulus and when the subject heard it.

Results and Discussion

The mean results for both conditions, averaged over the three subjects, are shown in Figure 2. Neither condition shows a peak in the middle of the discrimination function. As before, discrimination is best for the shortest rise times and worsens with increasing rise time. The outcome of the "recorded" condition is especially important. Although Cutting and Rosner's stimuli were recorded not once but twice, these results are at least an indication that the method of stimulus generation and presentation was not responsible for the discrepancies between their results, on the one hand, and ours and
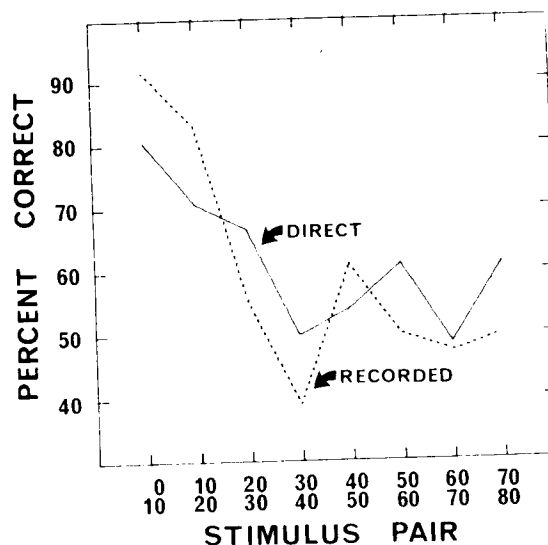

Figure 2. Discrimination results for a rise-time continuum in a one-step 2-sec interstimulus interval ABX task. The curve marked "Direct" is from stimuli presented directly after amplification to the subjects' headphones. Stimuli used for the "recorded" results were low-pass filtered at 4 kHz and monitored from a recording tape recorder. There were 72 observations per stimulus pair in the "direct" condition and 36 in the recorded condition.

van Heuven and van den Broecke's (1979), on the other.

EXPERIMENT 3: CATEGORIZATION AND DISCRIMINATION WITH ANOTHER RISE-TIME CONTINUUM

In view of the importance of Cutting and Rosner's results, we felt it necessary to make another attempt at replicating their findings using a completely different method of generating the stimuli. One reason was to follow the specifications of their stimuli much more closely in order to make sure that no slight differences in stimulus construction were affecting the results, though this seemed unlikely. Furthermore, although the stimuli of Experiment 1 were reasonably close to the ideal, we thought that a completely digital synthesis would be even better.

Method

Subjects. Eight subjects, all staff, students, or visitors at University College London, participated in this experiment. No selection was performed. All willing and available subjects were used.

Stimuli. Unlike the previous set of stimuli, these stimuli were generated completely under software control. A constant-amplitude periodic waveform made up of the first 13 harmonics of a 300-Hz sawtooth wave (amplitude of a particular harmonic inversely proportional to its harmonic number) was generated by adding together the appropriate sinusoids in phase on a DEC PDP-12 computer. Nine stimuli varying in rise time from 0 to 80 msec in 10-msec steps, followed by a 1.02-sec linear decay, were created from this basic stimulus by multiplying by the appropriate ramps.

Stimuli were output at run time at 10 kHz through a 10-bit D-A converter and filtered through two cascaded analog low-pass filters at 4,500 Hz with a total 96 dB/octave roll-off. Finally, these stimuli were amplified to a comfortable level, constant for all subjects, and routed to HD414X headphones in a quiet room. Oscillographic traces showed both the rise and decay to be linear, with all rise times within 1 msec of the desired values. These stimuli are essentially the same as those described by Cutting and Rosner.

Procedure. Again, a computer handled all details of stimulus presentation, response handling, and data analysis. Subjects sat in a quiet room and responded by pressing one of two buttons on a response box. The subjects were assigned randomly to one of two groups, four in each group. One group performed the categorization task first and one performed the discrimination task first.

In the categorization task, the subjects were asked to label the stimuli as "pluck" or "bow." The nine stimuli were presented 10 times each in a random order subject to the constraint that all stimuli were presented before any were repeated. At the beginning of the session, the subjects heard the 0- and 80-msec stimuli three times alternately as exemplars, while the buttons on the response box, marked "pluck" and "bow," lit appropriately.

The discrimination task was two-step ABX with approximately 1.95 sec between the onsets of the triplet (varying slightly with rise time). Two runs of 28 trials (7 comparisons × 4 orders) in a random order, again subject to the condition that all be presented before any were repeated, gave eight observations per stimulus pair per subject.

The labels "pluck" and "bow" were not mentioned to the subjects who performed the discrimination task first. All subjects ran the two conditions one after the other.

Results and Discussion

The mean results over the eight subjects are shown in Figure 3. There were no statistically significant differences between the two groups, in either their categorization or their discrimination performance. The categorization is reasonably sharp; there is no peak in the discrimination function at the category boundary. As before, best performance is for the shortest rise times, decreasing monotonically with increasing rise time.

Although we have focused attention on peaks in the discrimination function as a criterion for categorical perception, the traditional method has been to assess how well discrimination performance can be predicted from identification performance. In Table 1, we show for Experiments 1 and 3 the values predicted by:

$$P(C) = \frac{1}{2}[1 + (P_1 - P_2)^2], \qquad (1)$$

where $P(C)$ is the probability of a correct discrimination, and P_1 and P_2 are the probabilities of labeling the two test stimuli with the same name. This is known as the Haskins prediction formula, and assumes that stimuli can be discriminated only when they are labeled differently (Pisoni, Note 1). The fit is bad, primarily because the formula predicts very poor performance at short rise times, where the subjects are best.

For comparison purposes, we have also predicted performance on the basis of a simple-minded model which assumes that Weber's law operates over the
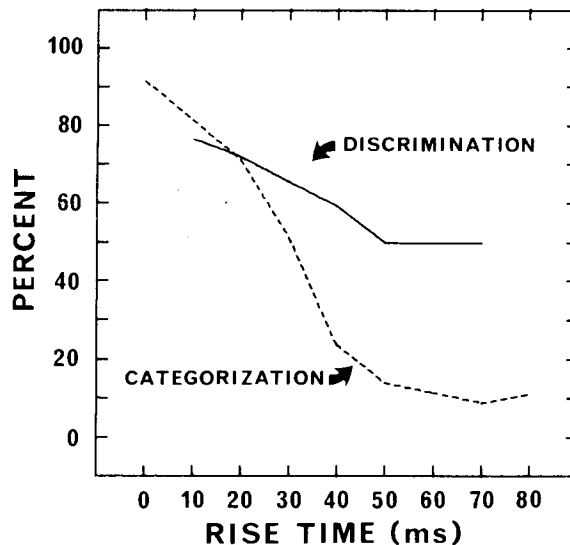

Figure 3. As for Figure 1, but for a second rise-time continuum. The interstimulus interval in the ABX task was about 900 msec. Each stimulus was categorized 80 times; each ABX pair was observed 64 times.

Table 1
Observed Discrimination Performance Compared With Theoretical Predictions for Experiments 1 and 3

| | Stimulus Pair | | | | | | | D |
	0/20	10/30	20/40	30/50	40/60	50/70	60/80	
				Experiment 1				
O	.94	.84	.77	.58	.57	.53	.55	
H	.53	.56	.67	.65	.55	.50	.50	.200
W	.94	.77	.66	.60	.57	.55	.54	.050
				Experiment 3				
O	.77	.72	.66	.59	.50	.50	.50	
H	.54	.56	.64	.62	.53	.51	.51	.108
W	.77	.59	.54	.53	.52	.51	.51	.071
V	.79	.69	.64	.61	.59	.58	.57	.055

Note—The Haskins predictions (H) were calculated subject by subject. The Weber's law predictions (W) were constrained to be equal to the observed discrimination performance (O) at the 0/20 pair. The variable Weber fraction predictions (V) were from a best fit to the Weber fractions calculated directly from the data. The calculations for Experiment 1 used the measured, not the nominal, rise times (2.3, 11.6, 21.6, etc.). Since the rise times for Experiment 3 were within 1 msec of the nominal, the nominal values were used. D = root-mean-square difference.

entire stimulus range. The transformation from performance in a two-alternative forced choice task to ABX performance was done using Macmillan et al.'s (1977) model. For further details, see the Appendix. The root-mean-square deviations show the predictions of this model to be much superior to those of the Haskins formula for both experiments. Most importantly, the monotonic decrease of performance with increasing rise time is captured only by the Weber law model.

It is also possible to play this game backwards and to calculate Weber fractions from ABX performance and check if Weber's law actually does hold (details of this procedure also in the Appendix). Table 2 shows these fractions. The data seemed reasonably linear on a log-log scale, so we fitted least squares straight lines to them on such a plot. The slope of the best-fitting line for Experiment 1 was not statistically different from 0 [t(5) = .64], indicating that Weber's law holds. The same analysis on the results of Experiment 3 showed a significant de-

crease in the Weber fraction with increasing rise time [t(2) = 4.42, p < .05]. These new estimates for the Weber fraction were used to make predictions for Experiment 3, also shown in Table 1 as the variable Weber fraction predictions. Allowing the Weber fraction to vary clearly improves the fit. Variable Weber fraction predictions for Experiment 1 were slightly worse than the Weber's law predictions, further confirming that Weber's law holds for this set of data.

Van Heuven and van den Broecke's (1979) results should bear on the issue of the shape of the Weber function. Let us assume, as we have done before, that the standard deviations in a reproduction task are proportional to the DLs in a 2I-2AFC task. The finding that the standard deviations were an increasing linear function of rise time is equivalent to saying that Weber's law holds. Unfortunately, their procedure precludes estimating the Weber function below about 20 msec, below which our second set of subjects showed the biggest relative change in the Weber fraction. In addition, van Heuven and van den Broecke's subjects, as well as the subjects in Experiment 1, were more practiced than the subjects of Experiment 3, and this may account for the differences in results. We return to this issue later.

EXPERIMENT 4:
MEASUREMENTS OF THE ORIGINAL PLUCK-BOW CONTINUUM

The experimental evidence just presented convinced us that there was, in fact, a serious discrepancy between those results and Cutting and Rosner's which was not attributable to differences in the specifications of method or stimulus generation techniques. No further progress could be made without examining Cutting and Rosner's original stimuli. James E. Cutting kindly supplied us with the original tapes.

We ran an experiment as similar as possible to Cutting and Rosner's Experiment 1. Nine subjects categorized and then discriminated the stimuli from a copy of the original tapes, listening over loudspeakers. We obtained results nearly identical to

Table 2
Weber Fractions (Percentages) Derived From Discrimination Performance in Experiments 1 and 3

| Experiment | Stimulus Pair | | | | | | | Slope | Intercept | r |
	0/20	10/30	20/40	30/50	40/60	50/70	60/80			
1	60	43	35	56	51	67	44	.0005	3.91	.002
3	106	61	51	52				-.5400	5.84	-.950

Note—Some values are missing because the calculations are not possible for chance performance. As explained in the footnote to Table 1, the measured rise times were used for the calculations of Experiment 1, whereas the nominal values were used for Experiment 3.

theirs; most importantly, the peak in the discrimination function appeared at 40 msec, the category boundary. We proceeded to measure the original stimuli.

Method

Oscillographic recordings were made of selected stimuli from the original categorization tape on an ultra-violet recorder running at 1,000 cm/sec. The first four occurrences of each of the nine low-pitched stimuli were recorded to check on the variability of the rise times from stimulus to stimulus. None was noted, so only one of each of the nine high-pitched stimuli was recorded. The rise times of all of these stimuli were measured in the following way. First, two horizontal lines were drawn to the peaks and valleys of the signal where its rise was completed and had become quasi-steady-state. Linear rises were then fitted by eye to the initial portion of the stimuli, and their intersections with the horizontal lines were noted. The positive- and negative-going halves of the best fits to the rises often intersected the parallel lines in different places, so their mean position was marked.

Table 3
Rise Times of Cutting and Rosner's Original Stimuli Measured From Oscillographic Traces of the Original Experimental Tape

Measured	Nominal								
	0	10	20	30	40	50	60	70	80
High Pitched	5	6	15	20	39	43	59	66	80
Low Pitched	4	6	15	18	35	43	56	67	72
Mean	4	6	15	19	37	43	57	66	76

Note—Times are shown to the nearest millisecond.

The distance from this point to the beginning of the stimulus was measured and converted to time.

Results and Discussion

Mean rise times are given in Table 3, separately for the low- and high-pitched stimuli. Figure 4 shows oscillograms of Cutting and Rosner's low-pitched
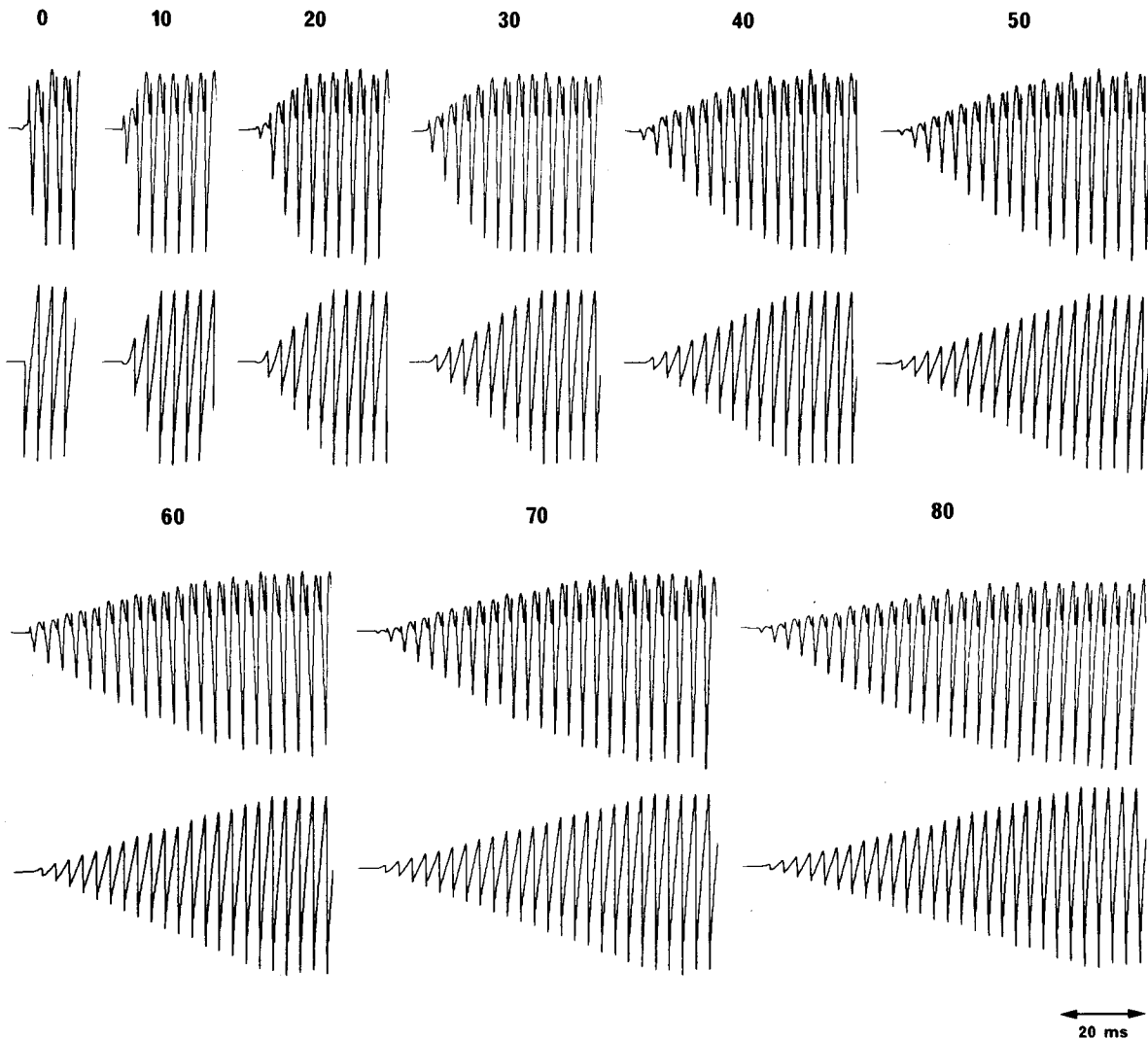

Figure 4. Initial portions of oscillographic traces of Cutting and Rosner's (1974) stimuli from their original tapes (first and third rows) compared with traces of the software sawtooth stimuli of Experiment 3. The latter may be used as a convenient time scale, since three cycles is 10 msec.
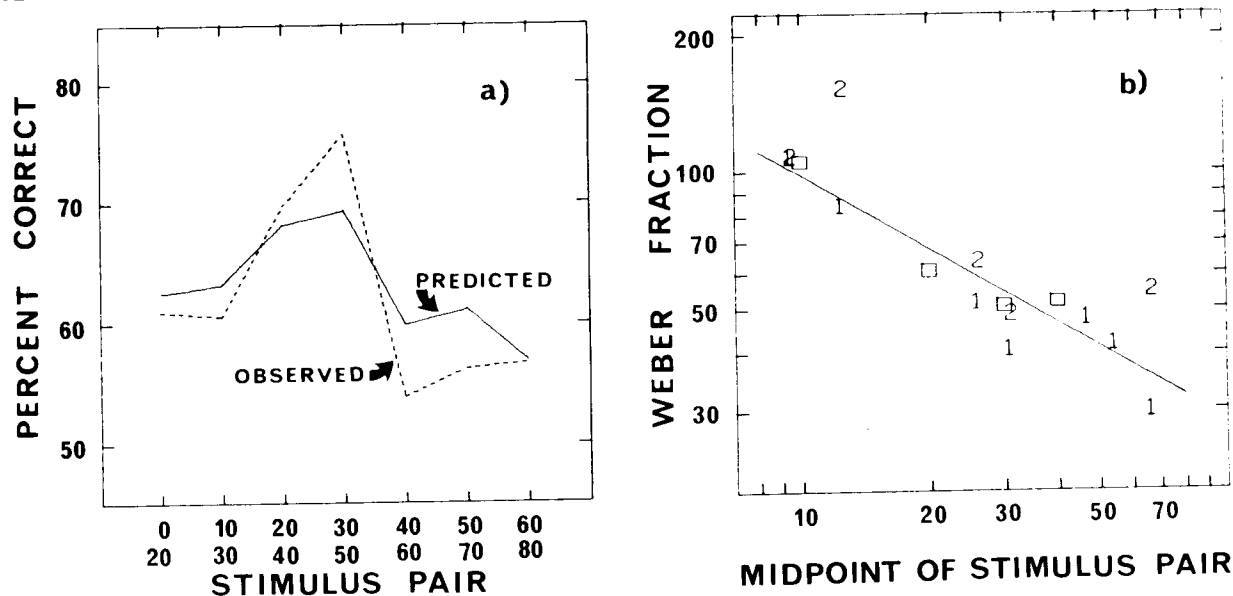
Figure 5. In (a) are shown Cutting and Rosner's mean results (Experiments 1 and 3, weighted according to subject number) compared with a prediction for performance from a variable Weber fraction model. The prediction is made from the rise-time measurements of the original stimuli and a Weber function derived from the results of Experiment 3. The critical midcontinuum peak is captured by the model. The overall fit is good because Cutting and Rosner's subjects show the same trend in the Weber fraction as the subjects in Experiment 3, as seen in (b). The squares represent the Weber fractions derived from our Experiment 3, and 1 and 2 refer to the fractions derived from the results of Cutting and Rosner's Experiments 1 and 2, respectively. The straight line is a least squares fit to the squares only. Some points are missing because it is not possible to derive Weber fractions from performances less than or equal to chance.

stimuli compared with our software sawtooth stimuli. As seen clearly in the figure and the table, the measured rise times for Cutting and Rosner's stimuli are quite different from the nominal values. Furthermore, they are wrong in a way consistent with Cutting and Rosner's discrimination results, as seen in Table 4. Here are calculated, using the values from Table 3, both the nominal and the measured difference in rise time for the seven two-step ABX pairs. The physical difference between the stimuli is actually greatest at the point Cutting and Rosner find their discrimination peak. This nonlinear spread of the stimuli along the rise-time continuum also sharpens up the categorization function.

We can use the Weber fraction models developed in the previous section for predicting performance with stimuli having these rise times. If Weber's law holds, then:

$$\text{true } d' = \frac{\text{distance between stimuli}}{\text{Weber fraction} \times \text{midpoint}}, \quad (2)$$

which is a monotonically decreasing function of rise time for these measurements. Since percent correct in an ABX task is a monotonically increasing function of d' (Macmillan et al., 1977), performance should also decrease with increasing rise time. Cutting and Rosner's subjects were, however, unpracticed and thus more like the subjects in Experiment 3. Using the Weber function derived for those subjects, we obtain a good fit to Cutting and Rosner's results, seen in Figure 5a. This correspondence holds, in fact, because the two groups of subjects show similar trends in their Weber fractions when the actual stimulus rise times are used in the calculations. This is seen in Figure 5b. Thus, *all* the unpracticed subjects are behaving in virtually the same way when the stimulus differences are taken into account.

Table 4
Difference in Rise Times, Measured in Milliseconds,
Between the Seven Two-Step ABX Pairs for
Nominal and Measured Values

	Stimulus Pair						
	0/20	10/30	20/40	30/50	40/60	50/70	60/80
Nominal	20	20	20	20	20	20	20
High Pitched	11	13	24	24	20	23	21
Low Pitched	11	11	20	26	21	24	17
Mean	11	12	22	25	20	23	19

EXPERIMENT 5:
MIMICKING THE OLD CONTINUUM
WITH THE NEW

Our analysis of the subjects' performance in Experiment 3 along with the measurements of Cutting and Rosner's stimuli enabled us to make a fairly convincing case for why Cutting and Rosner got the results they did. If the argument is correct, then

it should be possible to synthesize a new set of stimuli with Cutting and Rosner's rise times to obtain a peaked discrimination function.

Method

Subjects. All subjects from Experiment 3 who were available for retesting were used, a total of six subjects.

Stimuli. Using the same system described in Experiment 3, a new set of stimuli were synthesized, identical to those of Experiment 3 except in the rise times. These were 4, 6, 15, 17, 19, 43, 57, 66, and 76 msec, the mean rise times measured from Cutting and Rosner's stimuli.

Procedure. The subjects were run in one session of two-step ABX discrimination, as in Experiment 3, with an interstimulus interval of approximately 2.3 sec (onset to onset, matched to Cutting and Rosner's tape) using the altered continuum. They were then run on the Cutting and Rosner ABX tape, listening to the stimuli through the same headphones.

Results and Discussion

Figure 6 shows the discrimination functions, averaged over the six subjects, for all stimulus conditions. The discrimination function labeled "PROPER" refers to the results for these six subjects in Experiment 3. As seen before, discrimination is best for shortest rise times on this continuum. The Cutting and Rosner continuum and the copied stimuli, however, give quite a different result. In both cases, best discrimination performance is at 40 msec. Analyses of variance showed that the results using the two wrongly constructed continua were not significantly different from one another (F < 1), but both gave

results different from those obtained with the properly constructed stimuli [F(6,30) = 5.4, p < .001, between the proper continuum and Cutting and Rosner's, and F(6,30) = 3.7, p < .01, between the proper and the copied continuum].

EXPERIMENT 6:
CATEGORIZATION AND DISCRIMINATION
WITH THE MIMICKED CONTINUUM

As we have noted before, the nonlinear spacing of Cutting and Rosner's stimuli should make the categorization function sharper than it would be otherwise. To check this, and also to obtain more discrimination data, we ran a further group of subjects on both tasks.

Method

Subjects. The 12 subjects were staff, students, or visitors at University College London.

Stimuli. The stimuli were the same as those in Experiment 5.

Procedure. The same procedure as that of Experiment 3 was used, except that all subjects performed the ABX discrimination task before the categorization task.

Results and Discussion

The mean results for both tasks are shown in Figure 7. The categorization function is steep, with the category boundary at about 35 msec. The discrimination function, however, does not peak at 40 msec, due to good performance at short rise times. Inspection of the individual results (Table 5) shows that, although most subjects performed best near the middle of the continuum, some subjects showed peaks at the shortest rise times. These were, on the whole, the subjects with higher overall scores. Figure 8 shows the results for the best six subjects compared with those for the worst six. The two groups of subjects differ markedly in their ability to discriminate differences between short rise times but not in discriminating relatively long rise times from one another.

Consideration of both subjective impressions and the physical makeup of the stimuli suggests that there are different cues that are used for discrimination in different regions of the continuum. Physically speaking, a 20-msec change in the rise time has a large effect on the spread of spectral energy ("splatter") at onset for short rise times and a negligible effect on the spectrum for longer rises. The subjective impression is that at long rise times one must listen for the rise times per se, while at short rise times one can listen for the relative amplitude of a "thump" at the beginning of the sound. This presumably is why these stimuli sound like "plucks."

The contrasting pattern of results for the two groups of subjects implies that the better subjects make more use of the spectral cue for discrimination. Roughly speaking, subjects B.M. and D.B.
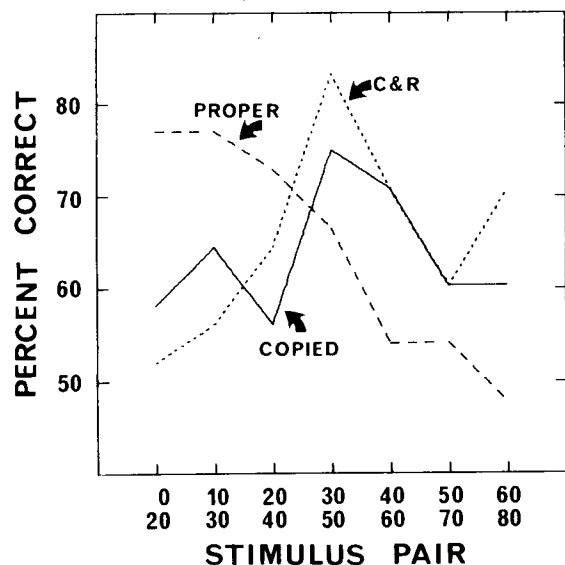

Figure 6. ABX discrimination results for three different rise-time continua. The "proper" stimuli had the correct rise times, "C & R" stimuli were from the original tapes of Cutting and Rosner (1974), and the "copied" stimuli had the same rise times as measured from the original Cutting and Rosner (1974) stimuli shown in Table 3. Each point represents 48 observations.
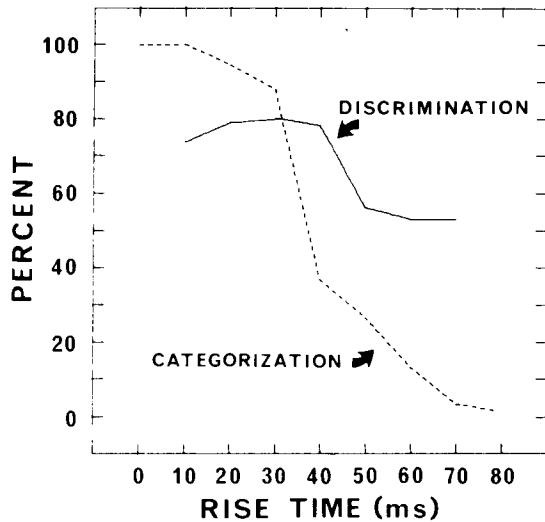
Figure 7. As in Figures 1 and 3, but for the "copied" stimuli. The interstimulus interval was about 1.2 sec. Each stimulus was categorized 120 times; each ABX pair was observed 96 times.

Figure 8. ABX discrimination results from Figure 6, plotted for the best and the worst subjects separately. Each point represents 48 observations.

seem to listen for the "thump" (as reported by B.M.) and ignore the other cue, most of the other subjects listen for the rise time per se and ignore the thump, and R.M. seems able to get both. (Of course, this must be a matter of relative and not absolute, sensitivity, since even the "thump-ignoring" subjects show best discrimination at shortest rise times when the stimuli are properly constructed). Simply taking out the three subjects who performed perfectly at the shortest rise times (R.M., B.M., and D.B.) makes the overall results look more like those of Cutting and Rosner, as seen in Figure 9. Here

Table 5
Individual Subject Results for the Two-Step ABX Discrimination Task Using the "Copied" Continuum, Ordered According to Overall Performance

| | Stimulus Pair | | | | | | |
	0/20	10/30	20/40	30/50	40/60	50/70	60/80	Total
RM	8	8	8	8	4	4	4	44
BL*	7	7	8	5	7	5	5	44
BM	8	7	8	4	5	5	5	42
JS	6	6	7	7	5	6	5	42
BC	6	8	7	8	4	5	3	41
DB	8	8	6	4	4	4	6	40
JM	7	7	6	8	3	6	2	39
BH	6	5	7	8	4	5	2	37
MP*	4	7	5	6	6	5	4	37
PB*	4	6	6	6	5	5	5	37
DH	3	7	7	4	6	5	3	36
BW	4	3	8	8	5	2	5	35
YS	5	7	4	7	5	1	6	35
GH*	5	3	3	8	5	7	4	35
DG*	5	5	4	5	5	5	5	34
SG	4	5	5	5	5	5	4	33
BO	6	5	4	4	4	2	6	31
BD*	3	3	1	6	6	2	6	27

Note—Values are number correct out of eight.
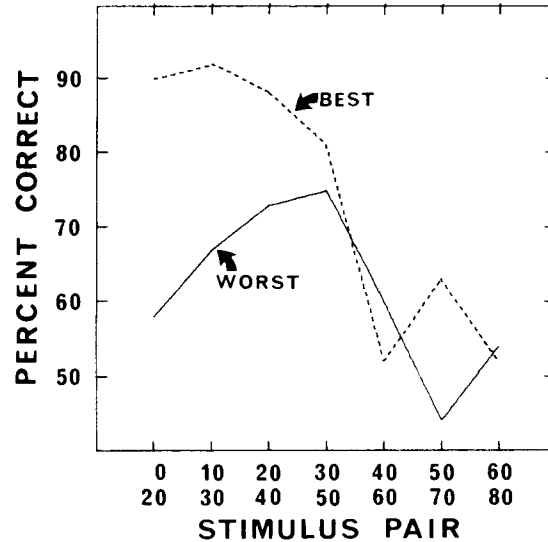*Participated in Experiment 5 using all three continua.

the results from Cutting and Rosner's original paper (mean of the two discrimination experiments weighted according to subject numbers) are compared with the mean of the 15 "ordinary" subjects, 6 from Experiment 5 and 9 from Experiment 6. The similarity between the two curves is striking.

EXPERIMENT 7:
COMPARING THE NEW AND THE OLD

We have implied that the unusual subjects of Experiment 6 should also be unusual on Cutting and Rosner's original stimuli. All three of them were available for one session using Cutting and Rosner's tape, as in Experiment 5. In Figure 10, the mean results are compared with those obtained in Experiment 6 with the copied continuum. Although the correspondence between the two discrimination functions is not overwhelmingly good, because of better performance at short rise times and worse performance at long rises on the copied continuum, the overall pattern is similar: good discrimination at the shortest rise times with little or no indication of a midcontinuum peak.

EXPERIMENT 8:
EXPLORING SUBJECT DIFFERENCES

The two previous experiments suggest that there are large differences between subjects in their ability to discriminate changes in rise time. A similar difference was seen between the results of Experiments 1 and 3 in which the former group of subjects was found to be much more sensitive at the short-rise-time end of the continuum, as evidenced
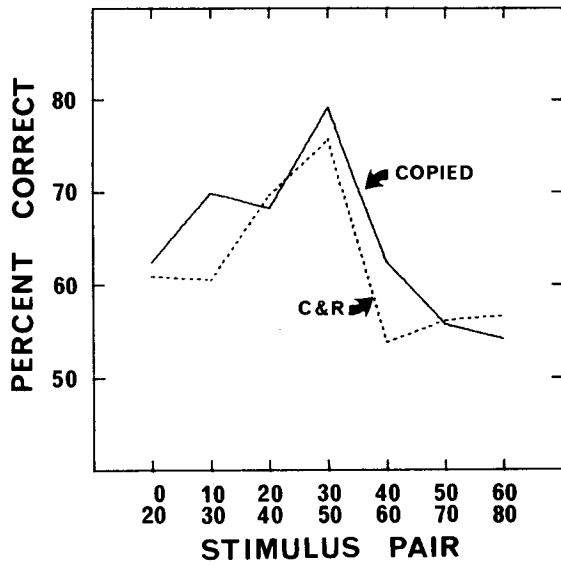
Figure 9. ABX discrimination results from Experiments 5 and 6, excluding three "odd" subjects compared with the discrimination results for sawtooth stimuli from Cutting and Rosner (1974). Each point in the "copied" curve represents 120 observations, and each in the "C & R" curve, 256 observations.
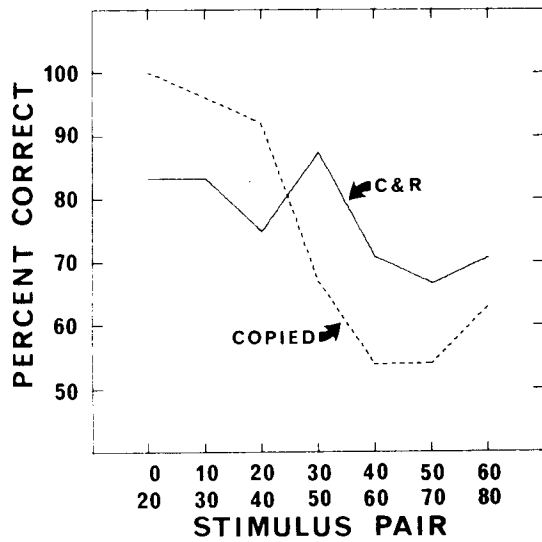

Figure 10. ABX discrimination results on the original Cutting and Rosner (1974) tapes compared with those obtained with the "copied" continuum for the three subjects who performed uncommonly well at the shortest rise times. Each point represents 24 observations.

by their smaller Weber fractions there. We speculated that this difference might be due to practice effects, but the unusual subjects of Experiments 6 and 7 had practiced no more than anyone else. Perhaps the differences are not due to practice but simply to chance selection of such unusual subjects.

We approached this issue from both sides. First, we tried to make an ordinary subject into an unusual one with extended practice, and second we looked at the performance of the subjects of Experiment 1 on the copied continuum.

Method

Subjects. Two of the three subjects of Experiment 1 were available for further testing, S.S. and S.B. One of the subjects of Experiments 3 and 5 who showed the "ordinary" pattern of performance, G.H., was recruited for testing over an extended period of time. A subject who showed the "unusual" pattern of performance, S.R. (the first author), was also run for an extended period of time. This is the first time data from either author have been reported.

Stimuli. There were two sets of stimuli: the original Cutting and Rosner stimuli, used in Experiments 5 and 7, and the copied continuum of Experiments 5, 6, and 7.

Procedure. Subjects G.H. and S.R. performed the ABX task with the copied continuum and with Cutting and Rosner's stimuli four times each in a counterbalanced order. Subjects S.S. and S.B. worked only with the copied continuum. Each of them ran two sessions of 56 trials each, one after the other. All procedural details were otherwise as in Experiment 5.

Results and Discussion

The mean results for G.H. and S.R. are shown in Figure 11 separately for the two subjects. For rise times of 40 msec or longer, the subjects do not differ, although the different continua give different results, for reasons discussed below. For short rise times, S.R. shows better performance on both continua, although each subject performs similarly on the two different continua. This is further support for the assertion that subjects vary in their
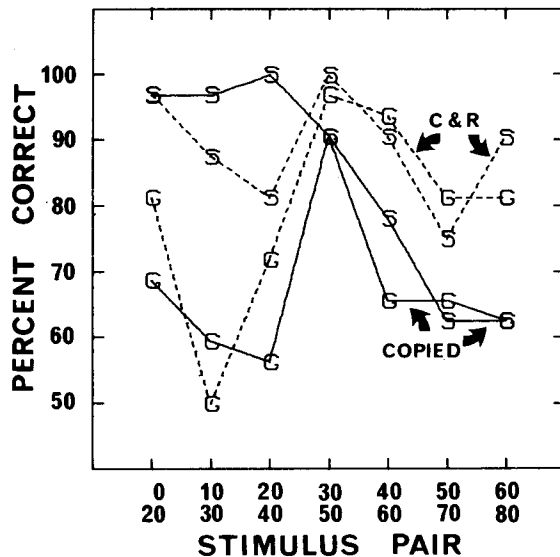

Figure 11. ABX discrimination results for two subjects on the "copied" continuum and the original Cutting and Rosner (1974) tapes. G refers to Subject G.H. and S to Subject S.R., who ran in each condition four times each. There were 32 observations per stimulus pair per condition.

Table 6
Weber Fractions (Percentages) Derived From the Results of Subjects S.R. and G.H. on the "Copied" Continuum in Experiment 8

| | Stimulus Pair | | | | | | | Slope | Intercept | r |
	0/20	10/30	20/40	30/50	40/60	50/70	60/80			
S.R.	30	27		27	22	37	25	−.03	3.4	−.142
G.H.*	79	108	109	27	33	33	25	−.71	6.3	−.790
G.H.†	109	120	61	36	33	29	25	−.83	6.7	−.970

*Four sessions.   †Last eight sessions.

sensitivity to different cues. Furthermore, moderate amounts of practice do not seem to change subjects' performance. G.H. participated in a further eight sessions of ABX discrimination and gave essentially the same results.

We derived the Weber fractions on the "copied" continuum for both subjects as described in Experiment 3. These are shown in Table 6, along with the results of fitting these values with straight lines on log-log plots. The slope of the best-fitting line for S.R.'s data is not significantly different from 0, while G.H.'s best-fitting line has a significant negative slope $[t(5) = 2.88, p < .05]$. This parallels the difference in results between Experiments 1 and 3. Also shown in Table 6 are the Weber fractions derived from the further sessions G.H. ran. Again, there is a significant negative slope $[t(5) = 8.77, p < .001]$. If the differences between subjects were solely due to practice, we would expect G.H.'s data to become more like S.R.'s, that is, that the Weber function would flatten out. In fact, it becomes marginally steeper.

As we have pointed out before, if Weber's law holds for the discrimination of rise time, performance should be a monotonically decreasing function of rise time even for stimuli having the rise times Cutting and Rosner's did. Since Weber's law did hold for the subjects of Experiment 1, there should be no midcontinuum peak in their discrimination function on the "copied" continuum. This is, indeed, the case, as seen in the mean results for subjects S.S. and S.B. in Table 7.

All the results we have reported, then, are consistent with the notion that subjects show different degrees of sensitivity to different cues, and that these differences are not, at least primarily, due to practice.

Table 7
Mean Discrimination Performance for Subjects S.S. and S.B.
in Experiment 8

| | Stimulus Pair | | | | | | |
	0/20	10/30	20/40	30/50	40/60	50/70	60/80
PC	.81	.78	.63	.78	.41	.53	.47

Note—PC = proportion correct.

Finally, one problem we have not mentioned is that some of the subjects who listened to Cutting and Rosner's original tape through headphones reported artifacts in the stimuli that helped their discrimination performance. These took two forms. The high-pitched 0- and 20-msec stimuli seemed more discriminable because of "ringing" throughout the 20-msec stimulus. This was visible on a spectrogram as amplitude modulation in some of the lower harmonics. Other stimuli had noises "underneath" the main sound. The 0- to 40-msec stimuli seemed to have no such artifacts, while the longer stimuli did. These were seen as extra components in a frequency-amplitude spectrum. Also, what artifacts there were differed from stimulus to stimulus. These latter artifacts are responsible for both G.H.'s and S.R.'s performing better at rise times 40 msec or longer with the Cutting and Rosner stimuli than with the copied continuum. Figure 9, of course, implies that Cutting and Rosner's subjects did not make much use of these extra cues.

GENERAL DISCUSSION

Although subjects can classify a continuum of sawtooth waveforms varying in rise time into two sharply defined categories with the boundary around 30-40 msec,[2] discrimination performance for equally spaced stimuli is always best for the shortest rise times. Cutting and Rosner (1974) found otherwise because their stimuli were not equally spaced on the continuum. Unfortunately, a fair number of other studies have as their sole *raison d'être* the supposed categorical nature of the perception of rise time. This applies to Cutting, Rosner, and Foard's (1976) studies on selective adaptation and varying interstimulus interval, the reaction time work of Blechner, Day, and Cutting (1976), the study with infants of Jusczyk, Rosner, Cutting, Foard, and Smith (1977), and the cross-continuum selective adaptation experiment by Remez, Cutting, and Studdert-Kennedy (1980). Especially unfounded are the speculations in the latter paper about the categorical nature of the perception of rise time being related to the categorical nature of sound production mechanisms in the violin, a fanciful extension

of motor theory which is having enough troubles of its own (see Howell & Harvey, in press, for a recent review).

Interpretation of one of the studies mentioned above depends strongly on the rise times of the stimuli's being correct. Jusczyk et al. (1977) claimed that infants, like adults, perceived the difference between stimuli with 30- and 60-msec rise times but not between stimuli with 0- and 30-msec or 60- and 90-msec rises. They did, in fact, use Cutting and Rosner's original stimuli. Assuming that the 90-msec stimulus has the correct rise time, as do the longer stimuli in the original set, we can use the variable Weber fraction model to predict performance for these stimulus pairs. Taking the Weber function to be that derived from the results of Experiment 3, the true d's between the 0-30, 30-60, and 60-90 pairs are 1.42, 2.08, and 1.19, respectively. Thus, the infants are "categorically perceiving" for the same reasons as the adults are. If the rise times had been correct, then the infants would be perceiving in a way completely different from that of adults, the opposite conclusion to the one reached by Jusczyk et al. (1977).

Finally, Cutting and Rosner's (1974) results have played a strong supporting role in some theorizing about the general nature of speech perception. Stevens (in press), for example, uses these data to support his claim that natural sensitivities in the auditory system are responsible for categorical perception. Given that the rise-time continuum is not categorically perceived, Stevens' ideas now rest on a single concordance between results from a nonspeech continuum—that involving temporal order judgments and VOT (Miller et al., 1976; Pisoni, 1977). Even this concordance seems doubtful when looked at in detail, according to Summerfield (Note 2). He claims that, while speech VOT continua show shifts in their category boundaries as a result of shifts in their first formant onset frequencies, nonspeech VOT analogs do not show similar shifts when their analogous "first-formant" onset frequencies are shifted.

Perhaps the most interesting positive outcome of this study is the notion that the discrimination of changes in rise time depends on different cues, spectral and temporal, for different magnitudes of rise time and that subjects vary in their relative sensitivity to the different cues. Even though these stimuli are not as complex as speech, they are more complex than the typical continua used in psychophysical experiments, and we suggest that such individual variations in perception may also be important in the perception of speech where most phonetic contrasts are multiply cued. Different perceivers may direct their attention to different aspects of a particular stimulus. These differences in strategy may well be reflected in the differences in an individual's ability, after losing some hearing or being fitted with a new hearing aid, to utilize any residual hearing or make efficient use of an aid. Techniques to train patients to attend to the still perceivable contrast-signaling parts of the speech signal could have great importance in rehabilitative schemes.

REFERENCE NOTES

1. Pisoni, D. B. On the nature of categorical perception of speech sounds. Supplement to *Haskins Laboratories Status Report on Speech Research*, 1971.
2. Summerfield, Q. *Does VOT equal TOT or NOT?* Paper presented at the meeting of the Experimental Psychology Society, Cambridge, England, July 1980.

REFERENCES

BLECHNER, M. J., DAY, R. S., & CUTTING, J. E. Processing two dimensions of nonspeech stimuli: The auditory-phonetic distinction reconsidered. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, 2, 257-266.

BRADY, S. A., & DARWIN, C. J. Range effect in the perception of voicing. *Journal of the Acoustical Society of America*, 1978, 63, 1556-1558.

CARDOZO, B. L. Adjusting the method of adjustment: SD vs DL. *Journal of the Acoustical Society of America*, 1965, 37, 786-792.

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.

FUJISAKI, H., & KAWASHIMA, T. A quantitative model for the mechanisms of discriminating speech sounds. *Reports of the Seventh International Congress on Acoustics*, 1971, Paper 21H2, 433-436.

GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1964.

HIRSH, I. J. Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 1959, 31, 759-767.

HOWELL, P., & HARVEY, N. Perceptual equivalents and motor equivalents in speech. In B. Butterworth (Ed.), *Language production* (Vol. 2). London: Academic Press, in press.

JUSCZYK, P. W., ROSNER, B. S., CUTTING, J. E., FOARD, C. F., & SMITH, L. B. Categorical perception of nonspeech sounds by 2-month-old infants. *Perception & Psychophysics*, 1977, 21, 50-54.

KAPLAN, H. L., MACMILLAN, N. A., & CREELMAN, C. D. Tables of d' for variable-standard discrimination paradigms. *Behavior Research Methods & Instrumentation*, 1978, 10, 796-813.

LEVITT, H. Decision theory, signal-detection theory, and psychophysics. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view.* New York: McGraw-Hill, 1972.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, 74, 431-461.

MACMILLAN, N. A., KAPLAN, H. L., & CREELMAN, C. D. The psychophysics of categorical perception. *Psychological Review*, 1977, 84, 452-471.

MILLER, G. A. The magical number seven plus or minus two, or, some limits on our capacity for processing information. *Psychological Review*, 1956, 63, 81-96.

MILLER, J. D., WIER, C. C., PASTORE, R. E., KELLY, W. J., & DOOLING, R. J. Discrimination and labeling or noise-buzz

sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, 1976, **60**, 410-417.

PARDUCCI, A. Category judgement: A range-frequency model. *Psychological Review*, 1965, **72**, 407-418.

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

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

REMEZ, R. E., CUTTING, J. E., & STUDDERT-KENNEDY, M. Cross-series adaptation using song and string. *Perception & Psychophysics*, 1980, **27**, 524-530.

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

STEVENS, K. N. Constraints imposed by the auditory system on the properties used to classify speech sounds: Data from phonology, acoustics and psychoacoustics. In T. F. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech*. Amsterdam: North-Holland, in press.

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.

WIER, C. C., JESTEADT, W., & GREEN, D. M. A comparison of method-of-adjustment and forced-choice procedures in frequency discrimination. *Perception & Psychophysics*, 1976, **19**, 75-79.

NOTES

1. This was not true for stop consonants, which showed no change in performance with changes in the interstimulus interval (ISI). Vowels, however, and short-duration vowels in particular, showed a much higher decrement for within-category than between-category discriminations for increasing ISI. This increased the peak-to-trough distance in the discrimination function, and hence its "categoricalness." Such effects are also to be expected from the model for categorical perception of Fujisaki and Kawashima (1971).

2. The boundary is, as one might expect, extremely labile with respect to changes in overall stimulus range. One subject was run in a labeling experiment using the computer-constructed software sawtooth stimuli of Experiment 3. On a given day, she categorized as pluck or bow the entire set of stimuli (9 stimuli × 5 observations per stimulus) followed by a subrange of five adjacent stimuli (20 observations per stimulus). There were three subranges: 0-40 msec, 20-60 msec, and 40-80 msec. Category boundaries were 25, 40, and 64 msec, all close to the midpoints of their respective ranges. (See Parducci, 1965, for a theoretical account of such boundary shifts.) That Cutting and Rosner (1974) found a speech and a nonspeech continuum, both varying in rise time, to have the category boundaries approximately equal in absolute value is primarily due to the two continua's having the same stimulus ranges.

APPENDIX

Given the Weber fraction (WF), we want to calculate performance in an ABX discrimination task for any given stimulus pair. We start by calculating the difference limen (DL), the minimum change in the stimulus necessary for detection by a subject. This is normally taken to mean that the subject discriminates the two stimuli correctly 75%-80% of the time in a two-interval two-alternative forced-choice (2I-2AFC) task. By definition,

$$WF = \frac{DL}{\text{stimulus magnitude}}, \qquad (3)$$

where we take the stimulus magnitude to be the midpoint of the pair of stimuli to be discriminated, so that:

$$DL = MS \times WF, \qquad (4)$$

where MS is the midpoint of the stimuli. For an unbiased observer performing at 76% correct in a 2I-2AFC task, the d' between the stimuli is $\sqrt{2}$ (Green & Swets, 1966). Thus, when two stimuli are one DL apart, the d' between them is $\sqrt{2}$, measured in a 2I-2AFC task. Both empirically and (under some assumptions) theoretically, this corresponds to a d' of 1 in a yes-no task, $d'_{yn} = 1$ (Levitt, 1972). In short, we assume that stimuli one DL apart have a $d'_{yn} = 1$. We now assume that the d'_{yn} between stimuli is proportional to the distance between them (which is another way of saying that the psychometric function is a cumulative normal), so that

$$d'_{yn} = \frac{DIS}{DL} = \frac{DIS}{MS \times WF}, \qquad (5)$$

where DIS is the physical difference between the stimuli.

Finally, we consult the tables compiled by Kaplan, Macmillan, and Creelman (1978) from which we can calculate the proportion correct for the discrimination of stimuli in an ABX task having that particular d'_{yn} under the assumption that the number of hits = 1 − number of false alarms.

Let's work some examples for illustration. What would be the performance of a subject in an ABX discrimination task between the 0- and 20-msec stimuli if the Weber fraction at 10 msec was 60%? Since the DL is $.6 \times 10 = 6$ msec and the distance between the stimuli is 20 msec, $d'_{yn} = 3.33$ according to Equation 5. From Kaplan et al.'s (1978) tables, then, $d'_s = 3.17$, and, by the arguments in Macmillan et al. (1977), performance would be 94.4% correct. What if Weber's law held? How well would this subject do with the 60-80 pair? The DL is now $.6 \times 70 = 42$ msec and the distance still 20 msec, so that $d'_{yn} = .48$. The tables show $d'_s = .13$, or 52.6% correct. This shows that performance decreases monotonically with increasing rise time if Weber's law holds.

To calculate the Weber fractions from ABX performance, we need only reverse these steps. For example, subjects get 76.6% correct for the 0-20-msec pair. What's the Weber fraction? First, 76.6% correct corresponds to a d'_s of 1.45 or a d'_{yn} of 1.89, according to Kaplan et al.'s (1978) tables. Rearranging Equation 5, we get:

$$WF = \frac{DIS}{d'_{yn} \times MS}. \qquad (6)$$

Substituting the proper values, the Weber fraction is 1.06.