Plucks and bows are categorically perceived, sometimes

JAMES E. CUTTING
Cornell University, Ithaca, New York

Cutting and Rosner (*Perception & Psychophysics*, 1974, **16**, 564-570) reported that sawtooth wavetrains were perceived categorically when an array was varied in linear rise time increments. That is, rapid rise time stimuli were identified as plucked strings and slower rise time stimuli as bowed strings, and pluck-bow comparisons were relatively easy to discriminate but pluck-pluck and bow-bow comparisons were not. Such results indicate the general equivalence of identification and discrimination tasks. Rosen and Howell (*Perception & Psychophysics*, 1981, **30**, 156-168), however, report that categorical perception does not occur for these sounds for two reasons. First, the original stimuli did not have the rise times reported by Cutting and Rosner. Second, the perception of these stimuli more closely follows a prediction from a Weber fraction. Acknowledging the first fact to be true, in part due to difficulties in digital-to-analog conversion, I set out to replicate and extend the results of Rosen and Howell. In Experiment 1, I found that stimuli with equal linear increments of rise time are not perceived categorically; but they are not perceived to follow closely a logarithmic relation either. In Experiment 2, I found that stimuli with equal logarithmic increments of rise time were generally perceived categorically. Experiment 3 replicated the results of both experiments. Thus, plucked and bowed music-like sounds can sometimes be found to be perceived in a categorical manner. However, categorical perception is not found with stimuli generated in the manner of Cutting and Rosner (1974) or Rosen and Howell (1981), and the phenomenon generally seems subject to rather stringent, if not curious, stimulus conditions. Moreover, and more deeply, categorical perception seems hardly the bedrock phenomenon it once appeared to be, whether in speech or in any other domain.

The purpose of this paper is twofold. First, it is a replication of Rosen and Howell (1981), a paper that falsified the conclusions of Cutting and Rosner (1974). The earlier paper claimed that sawtooth wave trains, when given systematic variation in their onset characteristics (or rise time), were perceived in a nearly categorical manner. That is, generally speaking, listeners could discriminate only as well as they could identify: they could identify the sounds as either plucked strings (rapid rise-time stimuli) or bowed strings (more gradual rise-time stimuli) and discriminate between plucked and bowed sounds, but they could not discriminate among different plucked sounds or different bowed sounds. Rosen and Howell concluded, however, that categorical perception did not occur. Second, this paper presents some new data used to assess the alleged categorical perception of plucked and bowed music-like sounds. In essence, an array of stimuli

generated with equal linear increments of rise time is not perceived categorically, but an array generated with equal logarithmic increments apparently is perceived categorically.

Credit to Rosen and Howell

Rosen and Howell (1981) present several findings. From my perspective, there are three important aspects to their report. First, they synthesized a new continuum of sawtooth waves differing in linear increments of rise time, analogous but not identical to the array reported by Cutting and Rosner (1974). They *did not* obtain results consistent with categorical perception. In particular, although they obtained the usual quantal identification function, they did not obtain a nonlinear, nonmonotonic discrimination function, as is shown in their Figures 1 and 3. Instead, they found a discrimination function that might be predicted better on the basis of a Weber fraction for rise time (see also van Heuven & van den Broecke, 1979). Second, using the original tapes of Cutting and Rosner (1974), they replicated our results, finding what appeared to be categorical perception. Third, to reconcile the difference in the two findings, they measured the original stimuli and found them quite discrepant from the rise times reported in the Cutting and Rosner paper. Moreover, the discrepancies are such that they predict the non-

linearities in the discrimination results. Thus, they concluded that plucked and bowed music-like sounds are not perceived categorically, and that the original claim concerning categorically is erroneous. Such systematic sleuthing is to be applauded; such a conclusion, however, warranted further investigation.

Two facts become pertinent. First, after reinspecting the stimuli, I concur with Rosen and Howell on the discrepancies in rise times in the original stimuli. Moreover, the error is mine. Discussion of these discrepancies and of the manner in which they occurred is given in the Appendix. Second, although Rosen and Howell report noncategorical perception of plucks and bows, several other investigators have found various plucked and bowed stimuli to be categorical. For example, Macmillan (Note 1, Note 2), using analog-generated stimuli of considerably lower fundamental frequency, did find categorical perception. In fact, his results, if anything, were more categorical than those of Cutting and Rosner (1974). The two sets of results differed only in that the identification and discrimination boundaries fell at about 25 msec of rise time for Macmillan's data rather than at about 35 msec. Also, Remez (1978) created a plucked-to-bowed continuum by tailoring natural tokens of musical sounds played on a bass viol. These, too, were perceived categorically. His continuum, unlike that of other investigators, was a rise-time-by-amplitude-at-onset continuum rather than simply a rise-time continuum. That is, his array had covariation in amplitude at onset as rise times became shorter. This was achieved by incrementally trimming off the onset portions of the stimuli. I take the results of Macmillan and of Remez to indicate that universal *non*categorical perception of plucked and bowed sounds is by no means assured.

The ambiguity of the situation prompted me to try to resolve it with more data. The three experiments reported here are an attempt to do just that.

EXPERIMENT 1: LINEAR INCREMENTS OF RISE TIME FOR STIMULI WITH LINEAR ONSET RAMPS

My first attempt was to replicate the results of Rosen and Howell (1981). In particular, they found that when rise times were measured carefully, discrimination performance could be accounted for by a Weber fraction for rise time (Rosen & Howell, 1981, Figure 5b). A Weber fraction model, of course, assumes one perceives logarithmically.[1]

Method

One array of stimuli was generated on computer: a linear increment series of sawtooth stimuli with rise times of 10, 15, 20, 25, 30, 35, 40, 45, and 50 msec. All stimuli were 1 sec in duration. Like the stimuli of Rosen and Howell, these had linear onset ramps and differed in linear increments of rise time, but with

5 msec between adjacent members of the array rather than their 10 msec. From maximum peak amplitude, all stimuli decayed linearly to zero amplitude. Stimuli were generated on a Hewlett-Packard 1000L computer, stored as digital waveforms, and sent through a Hewlett-Packard (Model 59303A) digital-to-analog converter to an AP circuit (Model 255-5) variable-frequency filter, and finally to an Ampex (Model AA-620) amplifier and loudspeaker. The sampling rate for the stimuli was 3,500 Hz, and the high-frequency cutoff on the filter was set to 1,600 Hz. The fundamental frequency of the waveforms was set to 500 Hz. With seven samples per cycle, and the maximum and minimum amplitudes set to ±1.00, one cycle of waveform consisted of values 1.00, .6667, .3333, .0, −.3333, −.6667, and −1.00, which repeated for 500 cycles. This pattern was then multiplied by the desired amplitude envelope for each stimulus to attain the digital versions of the stimuli, and multiplied further to use the full 10-bit capacity of the digital-to-analog converter. Functionally, these stimuli have all of their energy at 500, 1,000, and 1,500 Hz. Note that these stimuli were not recorded on audio tape, in part because van Heuven and van den Broecke (1979) reported that tape recorders have filter characteristics that may distort waveforms. Stimuli were measured on a Tektronix 5000 series storage oscilloscope subsequent to digital-to-analog conversion, and found not to vary from linear increments of rise time.

Two Cornell University graduate students, naive as to the purposes of the study, participated in 12 listening sessions. Each session lasted about 45 min. Listeners sat in front of the loudspeaker at a distance of 1 m in a relatively quiet room. At that distance, peak amplitude of the stimuli was about 75 dBA, as measured on a General Radio Company sound-level meter (Type 1551-C) for free-field situations. Each of the 12 sessions consisted of an identification sequence of 90 items (9 stimuli × 10 observations per stimulus), followed by an ABX discrimination sequence of 84 triads (7 ABX comparisons × 12 observations per comparison). The ABX comparisons consisted of all two-step stimulus comparisons: 10-20, 15-25, 20-30, 25-35, 30-40, 35-45, 40-50-msec pairs, occurring equally in ABA, ABB, BAB, and BAA configurations. In this manner, over the 12 sessions, there were 120 observations per stimulus per listener for identification of the stimuli, and 144 observations per listener for each discrimination pair of stimuli. Identification sequences had 3.4 sec between items, and discrimination sequences had 1.6-sec intervals within a triad and 4.2 sec between triads. Four different random-order identification sequences and four different random-order discrimination sequences were employed. Each listener listened to each order three times, rotated across the 12 sessions. Participation in one session a day was modal, and no more than two test sessions were run on any given day. Each listener was paid $40 upon completing the 12th session.

Three Predicted Functions

The categorical perception literature is necessarily chock full of comparisons between predicted and obtained discrimination functions. In essence, since discriminability is supposed to be no better than identifiability, one ought to be able to predict the discrimination function from the identification function. The first attempts at such prediction occurred before signal detection theory matured. Liberman, Harris, Hoffman, and Griffith (1957) offered the formulation which, when corrected for error in the original paper (see Cutting & Rosner, 1976, and Pollack & Pisoni, 1971), is given in Equation 1:

$$P(c) = .5[1 + (p_1 - p_2)^2],\qquad(1)$$

where P(c) is the probability of correct discrimination, p_1 is the probability of assigning Stimulus A to one of the categories, and p_2 is the probability of assigning Stimulus B to that same category. This prediction, based on low-threshold psychophysics (Macmillan, Kaplan, & Creelman, 1977), has come to be called the Haskins prediction.

A second prediction derives from the work of Macmillan et al. (1977) and is based on signal detection theory. An identification distance between compared stimuli is computed by subtracting the z-transformed probabilities of assigning them to the same category.[2] This difference can be considered a d'-like prediction. From here, I part with Macmillan et al. in the strategy with which they display their analyses, but I do not stray from their logic. These d'-like predictions can be converted into d'_s scores, which are more appropriate for ABX discrimination tasks. The notation is due to Pastore and Scheirer (1974); the method of conversion is that described by Macmillan et al. (1977); and the tables for interconversion are those given by Kaplan, Macmillan, and Creelman (1978, pp. 811-813). Assuming H to stand for hits and FA for false alarms, the following relation holds:

$$d'_s = |z(H)| + |z(FA)|. \qquad (2)$$

Since the proportion of hits and the proportion of false alarms sum to unity in this paradigm, their standardized absolute values are the same. Then, by detransforming the z scores of the predicted function, one may obtain a predicted discrimination function in terms of probability correct. Such a function is easily compared with the Haskins prediction, and I will call it the Macmillan et al. prediction.

The third prediction is a modification of that used by Rosen and Howell (1981). They, along with van Heuven and van den Broecke (1979), suggest that a Weber fraction can account for the discrimination data. Of course, there are an infinite number of Weber fraction curves; thus, a predicted curve must be anchored. Rosen and Howell (1981) anchor their prediction with difference limen from discrimination performance. This is fine for their purposes since they set out to see if a Weber fraction model fit the data. But, with regard to the spirit of predicted and obtained functions in the categorical literature, this procedure is inadequate: to predict from the obtained results is simultaneously to eat and have one's cake. A more principled prediction must come from the identification functions. Here I employ again some of the ideas of Macmillan et al. (1977). If one takes the mean identification distance of all two-step pairs of stimuli in the array, one has a good estimate of the mean predicted d' score for all comparisons. This is then converted into d'_s, as in the Macmillan et al. prediction. These d'_s predictions are then

weighted by the log difference in rise times of the compared stimuli, and then converted into a percent correct prediction. Let me be more concrete. The natural logarithms of rise times of the nine stimuli in this experiment are: 2.303, 2.708, 2.996, 3.218, 3.401, 3.555, 3.689, 3.807, and 3.912, respectively, for stimuli with 10- through 50-msec rise time. The difference in logarithms for the seven two-step comparisons is: .693, .510, .405, .337, .288, .252, and .223, for pairs 1-3 through 7-9, respectively. The mean of these differences is .387. Dividing each of these differences by the mean yields an array of seven weights: 1.79, 1.32, 1.05, .87, .74, .65, and .58. Then, having determined the mean predicted d'_s, that value is then multiplied by each of the seven weights, and then reconverted into a percent correct prediction as before, using Formula 2 and a z table.[3]

Results and Discussion

The results for both subjects are shown in Figure 1. Four panels each are devoted to each subject. In the two panels at the top display are shown the complementary identification functions for plucked and bowed responses. Both listeners demonstrate reasonably discrete identification categories. Listener
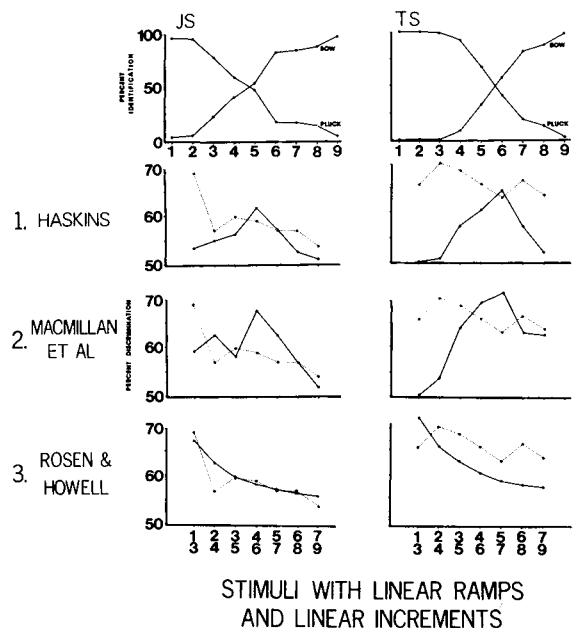
Figure 1. Identification and discrimination performance for sawtooth stimuli generated with equal linear increments in rise time and with linear onset ramps. Complementary identification functions for two listeners, J.S. and T.S., are shown in the top panels. The relationship between the observed discrimination function (the dotted lines) and each of the three predicted functions is shown in the panels beneath them. The Haskins prediction is based on low-threshold psychophysics, the Macmillan et al. prediction on signal-detection theory, and the Rosen and Howell prediction on the assumption that discrimination is a logarithmic function of differences in rise time.

Table 1
Observed and Predicted Scores (of 144) for Listeners in Experiments 1 and 2

		Stimulus Comparison							
		1-3	2-4	3-5	4-6	5-7	6-8	7-9	χ^2 (7)
					Experiment 1				
(1) J.S.	Observed	99	82	86	85	82	82	74	
	Haskins	76.8	78.5	80.7	88.4	82.6	75.7	73.8	15.07*
	Macmillan et al.	85.7	89.8	83.7	97.7	90.1	81.7	77.5	13.06
	Rosen & Howell	97.0	90.6	86.8	84.5	82.8	81.2	80.6	3.08
(2) T.S.	Observed	95	101	99	95	91	96	92	
	Haskins	72.1	73.6	82.6	87.2	93.3	82.8	75.1	54.88†
	Macmillan et al.	72.2	77.6	92.2	99.9	103.5	91.2	90.9	35.84†
	Rosen & Howell	103.1	95.4	91.2	87.8	85.7	83.9	82.8	12.27
					Experiment 2				
(1) J.S.	Observed	88	91	94	103	83	83	81	
	Haskins	75.6	80.6	88.1	91.1	85.6	75.3	72.3	14.88*
	Macmillan et al.	80.6	91.6	95.3	100.8	101.5	81.9	74.2	13.38
	Rosen & Howell	87.9	87.9	87.9	87.9	87.9	87.9	87.9	9.68
(2) T.S.	Observed	73	81	80	90	93	95	85	
	Haskins	72.1	72.9	75.7	92.4	96.2	83.4	76.4	7.68
	Macmillan et al.	72.9	79.0	88.1	104.2	105.1	98.3	90.2	13.95
	Rosen & Howell	91.1	91.1	91.1	91.1	91.1	91.1	91.1	16.74*
(3) M.C.	Observed	74	81	93	76	68	71	72	
	Haskins	73.7	75.1	84.4	96.3	84.3	78.9	74.3	23.82**
	Macmillan et al.	76.9	85.5	99.7	107.3	92.2	88.8	84.5	66.27†
	Rosen & Howell	90.7	90.7	90.7	90.7	90.7	90.7	90.7	51.85†
(4) J.E.C.	Observed	97	104	120	102	106	83	77	
	Haskins	72.6	89.3	109.4	99.4	78.5	71.1	72.0	49.85†
	Macmillan et al.	77.0	105.1	121.5	114.1	92.0	72.6	72.0	25.20†
	Rosen & Howell	93.5	93.5	93.5	93.5	93.5	93.5	93.5	41.19†

*$p < .05$.   **$p < .005$.   †$p < .001$.

J.S. has a boundary between plucked and bowed categories at slightly less than 30 msec, Listener T.S., at almost 35 msec. These values are generally consistent with those of both Rosen and Howell (1981) and Cutting and Rosner (1974).

The relationship between the observed discrimination functions and each of the predicted discrimination functions is shown in the lower six panels, Listener J.S. on the left and Listener T.S. on the right. They are also shown in Table 1. The observed functions are shown as dotted lines and are repeated in each of the panels of the figure. The discrimination functions were obtained by averaging the predictions for each of the 12 sessions.[4] Two measures of comparison between obtained and predicted functions will be entertained. The first, and most common in the literature, is a chi-square goodness-of-fit measure. This metric is particularly good at measuring the *proximity* of the two functions. When the chi-square value is low and nonsignificant, the prediction is a good one. A second measure, designed to better capture the relative *shape* of the two functions, is an analysis-of-variance trend test, using the values of the predicted functions (normalized around zero) as weights.[5] When this F ratio is relatively high and statistically significant, the prediction is a good one.

Consider first the data of Listener J.S. The difference between the Haskins-predicted and the obtained

functions is marked and reliable [$\chi^2(7) = 15.07$, $p < .025$].[6] Also, there is no significant trend in the observed data that corresponds to the Haskins prediction [$F(1,66) = .19$, n.s.]. Thus, neither the proximity nor the shape of the Haskins prediction reflects performance. Roughly the same is true of the Macmillan et al. prediction: the proximity measure of that function is somewhat better [$\chi^2(7) = 13.06$, $p < .05$], but its shape measure is not [$F(1,66) = .66$, n.s.]. In this manner, neither prediction from the assumptions of categorical perception fits the observed data. The Rosen and Howell prediction fares much better: its proximity is good [$\chi^2(7) = 3.08$, n.s.], and its shape is, too [$F(1,66) = 5.72$, $p < .05$]. Thus, the data of J.S. replicate the results of Rosen and Howell (1981) and not those of Cutting and Rosner (1974).

The data of Listener T.S. reveal almost the same pattern. The Haskins prediction is widely discrepant in its proximity to the observed data [$\chi^2(7) = 54.9$, $p < .001$] and also in its shape [$F(1,66) = .46$, n.s.]. The same is true for the Macmillan et al. prediction [$\chi^2(7) = 35.8$, $p < .001$, and $F(1,66) = .11$, n.s.]. Thus, it is clear that the data of T.S. are not categorical. However, the Rosen and Howell prediction does only a little better: the proximity of the predicted function to the obtained is only fair [$\chi^2(7) = 12.3$, $p < .10$], and the shape of the predicted function does not reveal the trend in the data [$F(1,66) = .56$, n.s.].

Taken together, the data of the two subjects strongly support one claim of Rosen and Howell (1981): an array of sawtooth wave trains that have linear increments in rise time is not perceived categorically. Their notion that a Weber fraction governs discriminability of these stimuli, however, receives less support: Listener J.S. satisfactorily shows a logarithmic trend, but Listener T.S. does not. To investigate the Rosen and Howell prediction further, a second experiment was performed. If a Weber-fraction view is correct, stimuli generated with equal logarithmic increments of rise time ought to yield a flat discrimination function. Thus, log stimuli roughly within the range of rise times used in Experiment 1 should not be found to yield categorical perception.

EXPERIMENT 2: LOGARITHMIC INCREMENTS OF RISE TIME FOR STIMULI WITH LINEAR ONSET RAMPS

Method

Exactly the same procedure was used as in Experiment 1. The two listeners, J.S. and T.S., were brought back again. They were joined by Listener M.C., a summer research assistant also naive as to the purposes of the study. Each of the three performed in 12 sessions for $40. Again, modal participation was one session per day, but as many as three or four were occasionally run in a 24-h period. For comparison purposes, I also ran myself in 12 sessions. As I am probably the most experienced pluck-and-bow listener around, my data can be used as a backdrop of expertise against the other three participants. Again, 120 observations per stimulus comparison per listener were obtained in the identification task, and 144 observations per stimulus comparison per listener, in the discrimination task.

The stimuli used in this experiment were exactly like those in Experiment 1 except for rise time. Ten stimuli were employed, with linear onset ramps 15.0, 18.0, 21.6, 25.9, 31.1, 37.3, 44.8, 53.7, 64.5, and 77.3 msec in duration. These stimuli differ in natural log steps of .183. These values and this log increment were chosen because, together, they conveniently allow saturation of the rise-time domain of interest. Again, all stimuli were measured on oscilloscope subsequent to digital-to-analog conversion, and found not discrepant in rise times from logarithmically equal steps. The first nine of these stimuli are shown in Figure 2. Listeners J.S., M.C., and J.E.C. heard identification and discrimination sequences using these nine stimuli; Listener T.S., because of a relatively longer rise-time boundary in the previous experiment, heard the last nine stimuli.

Results and Discussion

The results for the three naive listeners are shown in Figure 3. Again, four panels are devoted to each. The complementary identification functions are shown in the top panels for each listener, and again they demonstrate relatively quantal functions with boundaries very near where they occurred in Experiment 1. For J.S. this boundary is at about 30 msec, for T.S. it is at about 41 msec, and for M.C., 32 msec. The boundary for J.E.C. is 28 msec.

Again, consider first the data of J.S. The Haskins prediction, as shown in Table 1 and in the left-hand panel of Figure 3 beside the numeral 1, is not particularly close to the obtained discrimination func-
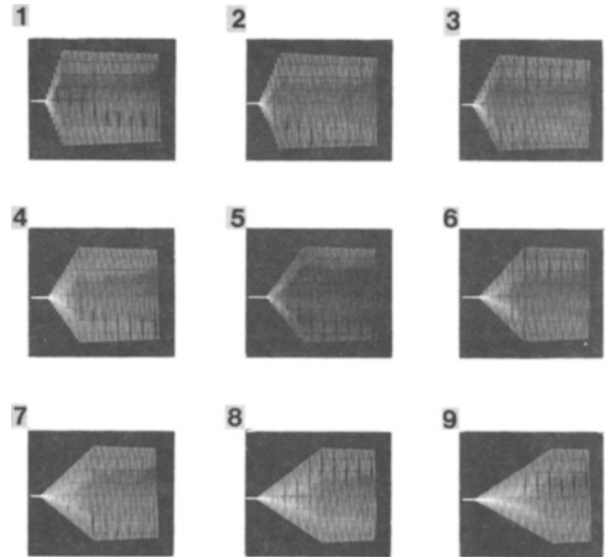

Figure 2. The first 100 msec of the first nine stimuli used in Experiment 2, in which stimuli were generated in equal logarithmic rise-time steps. Since the fundamental frequency of these stimuli is 500 Hz, each pulse width is 2 msec. These stimuli were photographed from a storage oscilloscope after digital-to-analog conversion and filtering.

tion [$\chi^2(7) = 14.9$, $p < .05$], but the shape of that function is quite a good predictor of the trend in the observed data [$F(1,66) = 7.93$, $p < .01$]. The same is generally true for the Macmillan et al. prediction, also in the table and in the left-hand panel of the same figure beside the numeral 2: its proximity is somewhat better [$\chi^2(7) = 13.4$, $p < .10$], and as a trend it predicts the shape of the obtained data about as well [$F(1,66) = 4.25$, $p < .05$]. This is in relative contrast to the Rosen and Howell prediction in the table and in the figure beside numeral 3: it is relatively close to the obtained data [$\chi^2(7) = 9.7$, $p < .20$], but its shape is nothing like that observed (no F ratio can be determined since a flat function is, literally, no trend). A post hoc inspection of the data reveals that performance on comparisons between Stimuli 4 and 6 was better than on those between Stimuli 1 and 3 [$t(11) = 2.45$, $p < .05$] and also better than on those between Stimuli 5 and 7 [$t(11) = 2.49$, $p < .05$]. The Rosen and Howell prediction can account for neither of these differences, suggesting further that the categorical predictions are superior for these data. Thus, the data of Listener J.S. bear out two facts: first, these results are not particularly consistent with a Weber-fraction model for the discriminability of these stimuli, and second, they are quite close to a traditional view of categorical perception.

Consider, secondly, the data of T.S., shown in the middle panels of Figure 3 and also in Table 1. The Haskins prediction is both fairly close to the obtained discrimination function [$\chi^2(7) = 7.7$, n.s.] and a good predicter of shape [$F(1,66) = 5.42$, $p < .05$]. The
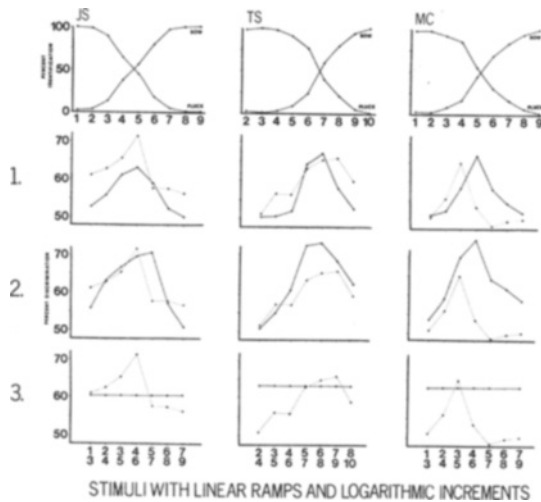
Figure 3. Identification and discrimination performance for sawtooth stimuli generated with equal logarithmic increments in rise time and with linear onset ramps. Again, the complementary identification functions are shown in the top panels for three listeners, J.S., T.S., and M.C. Beneath these are the relations between the observed discrimination function (the dotted lines) and each of the three predicted functions. Number 1 is the Haskins prediction, No. 2, the Macmillan et al. prediction, and No. 3, the Rosen and Howell prediction. J.S. and M.C. heard Stimuli 1 through 9, and T.S. heard Stimuli 2 through 10.

same pattern is essentially found for the Macmillan et al. prediction: the proximity to the obtained data is fairly good [$\chi^2(7) = 13.9$, $p < .10$] and the shape is quite a good predicter [$F(1,66) = 12.36$, $p < .01$]. The Rosen and Howell prediction, on the other hand, is not very close to the obtained data [$\chi^2(7) = 16.7$, $p < .025$], and again no trend can be predicted. A post hoc analysis of Comparisons 2-4 and 7-9, for example, reveals that these are reliably different [$t(11) = 2.65$, $p < .02$], again suggesting that the Weber-fraction model of the discrimination of these sounds is not correct. Note, however, that there is no significant trough at the longer rise-time end of the continuum. In particular, Comparison 7-9, for which performance was highest, is not significantly greater than Comparison 8-10 [$t(11) = 1.14$, $p < .20$]. This fact will be dealt with in more detail in the general discussion.

Consider next the data of Listener M.C. As shown in the table and figure, none of the three predictions are proximal to the obtained discrimination function [Haskins, $\chi^2(7) = 23.8$, $p < .005$; Macmillan et al., $\chi^2(7) = 66.3$, $p < .001$; and Rosen and Howell, $\chi^2(7) = 51.8$, $p < .001$]. In terms of shape, the pattern is little different: the Haskins prediction [$F(1,66) = .20$, n.s.] and the Macmillan et al. prediction [$F(1,66) = 1.96$, n.s.] do not account for the shape of the obtained discrimination function, in large part because the peaks in the predicted and obtained functions are at different comparisons. Again, there is no shape

comparison that can be made with the Rosen and Howell prediction. However, post hoc inspection of the results reveals that performance on the 3-5 comparison is significantly better than that on the 5-7 comparisons [$t(11) = 3.23$, $p < .005$] and marginally better than that on the 1-3 comparison [$t(11) = 1.86$, $p < .08$]. Such results suggest that the Rosen and Howell prediction of no differences is not a correct prediction, and that the data are more categorical than they are uniform. Again, this overall pattern will be treated in more detail in the general discussion.

Finally, for comparison purposes, consider my own data, shown in Table 1. None of the three predictions does well in terms of proximity to the obtained function, but, of the three, the Macmillan et al. prediction is best. With regard to prediction of the shape of the function, the Haskins prediction does very well [$F(1,66) = 25.11$, $p < .001$], as does the Macmillan et al. prediction [$F(1,66) = 22.16$, $p < .001$]. The Rosen and Howell prediction does not fare well, since there is a significant main effect of stimulus comparison [$F(6,66) = 6.52$, $p < .001$], indicating reliable differences in performance among the seven ABX pairings.

In summary, the results of Experiment 2 are fairly consistent with the view that plucked and bowed music-like sounds are perceived categorically when generated with rise times in logarithmically equal steps. One may quibble, however, with the use of so few listeners, all of whom by the end of this experiment were exceedingly well-practiced. Thus, a third experiment was conducted. Its aim was to replicate the results of both Experiment 1 and Experiment 2, but with a new set of stimuli and with a new set of naive listeners.

EXPERIMENT 3: LINEAR AND LOGARITHMIC INCREMENTS OF RISE TIME FOR STIMULI WITH CURVILINEAR ONSET RAMPS

Rosen and Howell (1981) made the assumption in their study that stimuli with linear onset ramps were the proper comparison for the stimuli of Cutting and Rosner (1974). Yet, the stimuli of Cutting and Rosner more clearly have nonlinear onset ramps that approximate the first 90 deg of a sine function than they do of a linear ramp function (see Rosen & Howell, 1981, Figure 4). Thus, to make the comparisons with the original study more closely, a set of stimuli was generated with curvilinear onset ramps.

Method

Two sets of stimuli were generated on computer: an array with linear increments that included eight stimuli with rise times of 8, 16, 24, 32, 40, 48, 56, and 64 msec and an array with logarithmic increments that included eight stimuli with rise times of 18, 21.6, 25.9, 31.1, 37.3, 44.8, 53.7, and 64.5 msec. Thus, the linear-array

stimuli differed in rise times of 8 msec between successive items and the logarithmic-array stimuli by natural log steps of .183. All stimuli had sinusoidal onset-ramp envelopes; that is, the rise time portion of each stimulus had the profile of the first 90 deg of a sine function. Otherwise, these stimuli were identical to those used in Experiments 1 and 2. Again, stimuli were inspected on a high-resolution oscilloscope, and rise times were found to be very close to their intended values.

For generality's sake, the procedure was altered slightly. Unlike in the previous studies, stimuli here were recorded on audio tape and played back on a Revox (Model A77) tape recorder, through Telephonics (Model TDH39) headphones at 70 dB. Twelve members of the Cornell Summer School community participated individually in one 75-min session. They were paid $4 for their participation. These listeners included J.S., T.S., and M.C., but all three participated in this study before participating in Experiments 1 and 2. Participants listened to four test sequences: an identification and a discrimination sequence for both linear and logarithmic stimuli. Half the listeners heard the linear stimuli first, half the logarithmic stimuli first. All subjects listened first to the identification sequence within each stimulus type, then to the discrimination sequence. Identification sequences consisted of a random presentation of 80 items: 8 stimuli in each array × 10 observations per stimulus. The discrimination sequences consisted of 96 ABX triads: 6 two-step comparisons × 16 observations per comparisons. Intervals between stimuli in both tasks were the same as in Experiments 1 and 2.

Results and Discussion

Grouped results and predictions are shown in Figure 4, and the individual observed and predicted discriminations are shown in Tables 2 and 3. The upper two panels of Figure 4 generally reveal identification functions like those found earlier. Moreover, the boundaries between categories are at about 35 msec for each array.

Consider first the linear array and the predicted and obtained discrimination functions in the lower three panels on the left-hand side of Figure 4. As in Experiment 1, the left-hand third of the obtained function is ill-fitting with respect to the Haskins prediction. As shown in Table 1, 3 of the 12 listeners (T.S., J.B., and T.C.) have predicted functions that are reliably different in proximity to the observed function. As a group, the mean of the individual predicted functions does not predict the shape of the obtained function in any reliable manner [$F(1,55) = 2.54$, n.s.]. The Macmillan et al. prediction fares no better: four listeners (T.S., M.C., E.C., and T.C.) have predicted functions reliably different in proximity to the observed function, and as a group the mean of the individual functions does not predict the shape of the obtained function [$F(1,55) = .15$, n.s.]. Interestingly, the Rosen and Howell prediction is only a little better: whereas only 1 (M.C.) of 12 listeners has a predicted function that deviates in proximity from the obtained function, the mean of the individual trends is not a reliable predictor [$F(1,55) = 2.68$, n.s.]. This may seem odd given the apparent proximity of the two functions in Figure 4, but remember that the Rosen and Howell prediction is a function that is concave upward and the obtained
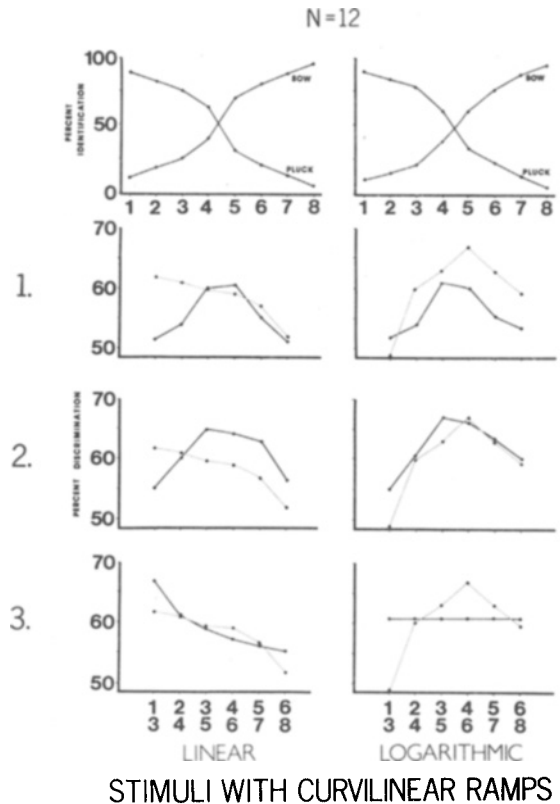

STIMULI WITH CURVILINEAR RAMPS

Figure 4. Identification and discrimination performance for sawtooth stimuli generated with both equal linear and equal logarithmic increments in rise time, and with curvilinear onset ramps. The mean complementary identification functions for 12 listeners are shown in the top panels and the relation between mean observed (dotted lines) and mean predicted discrimination in the panels beneath them. Again, No. 1 is the Haskins prediction, No. 2, the Macmillan et al. prediction, and No. 3, the Rosen and Howell prediction.

function is concave downward. In short, these results replicate those of Experiment 1: categorical perception was clearly not found for these stimuli, but the discrimination function is not clearly one that is predictable from a Weber-fraction view.

Consider next the results of logarithmic stimuli in the right-hand panels of Figure 4 and in Table 3. The Haskins prediction fares very well: only 1 (E.C.) of 12 listeners has a reliably different predicted function from that obtained, and, as a group, the mean of predictions across listeners serves as a reliable predictor of the shape of the observed function [$F(1,55) = 7.51$, $p < .01$]. The Macmillan et al. prediction presents more mixed results: 4 (J.B., G.R., H.A., and E.C.) of the 12 listeners have predicted functions reliably different in proximity from their obtained functions, but the group trend predicts the shape of the obtained function very well [$F(1,55) = 27.06$, $p < .001$]. In general, these results are consistent with the view that these logarithmic stimuli are perceived categor-

Table 2
Observed and Three Predicted Scores (of 16) for the Linear Array in Experiment 3

		Stimulus Comparison						
		1-3	2-4	3-5	4-6	5-7	6-8	$\chi^2(6)$
(1) J.S.	Observed	6	9	10	8	8	8	
	Haskins	8.08	8.32	8.32	8.32	8.32	8.32	.98
	Macmillan et al.	8.11	8.46	8.46	8.56	8.98	9.97	1.55
	Rosen & Howell	9.39	8.89	8.69	8.50	8.43	8.37	2.32
(2) T.S.	Observed	11	14	15	10	11	9	
	Haskins	8.00	8.32	9.28	13.12	9.28	8.00	18.62**
	Macmillan et al.	8.00	10.00	10.10	14.35	9.87	8.00	20.21**
	Rosen & Howell	11.22	10.13	9.57	9.26	9.01	8.89	9.95
(3) M.C.	Observed	5	7	8	6	7	8	
	Haskins	8.08	10.00	9.28	8.72	9.28	8.08	5.54
	Macmillan et al.	8.38	11.28	9.87	9.66	11.71	8.56	15.12*
	Rosen & Howell	11.40	10.20	9.68	9.30	9.05	8.94	15.63*
(4) G.N.	Observed	11	8	9	9	7	7	
	Haskins	8.00	8.00	8.32	8.32	8.32	8.00	1.81
	Macmillan et al.	8.00	8.00	8.98	8.98	8.98	8.00	2.18
	Rosen & Howell	8.56	8.37	8.24	8.20	8.17	8.14	1.20
(5) J.B.	Observed	15	12	12	6	12	9	
	Haskins	8.32	10.88	10.88	8.32	8.32	8.32	13.15*
	Macmillan et al.	9.97	13.18	11.60	8.56	9.97	9.97	7.41
	Rosen & Howell	12.56	11.06	10.30	9.82	9.50	9.33	5.79
(6) F.R.	Observed	12	10	9	12	12	11	
	Haskins	8.72	8.32	8.08	10.88	10.88	8.00	4.12
	Macmillan et al.	10.91	8.98	8.13	12.02	13.18	8.00	1.97
	Rosen & Howell	11.45	10.25	9.70	9.33	9.07	8.84	3.43
(7) E.K.	Observed	10	11	9	10	9	8	
	Haskins	8.72	8.00	8.32	8.32	8.72	8.72	2.09
	Macmillan et al.	9.01	8.00	8.46	8.46	9.01	10.96	3.65
	Rosen & Howell	9.87	9.18	8.88	8.69	8.57	8.50	.61
(8) G.R.	Observed	10	8	9	14	12	10	
	Haskins	8.08	8.32	10.88	10.88	8.32	8.08	6.06
	Macmillan et al.	8.76	9.96	12.05	11.60	8.98	8.38	5.97
	Rosen & Howell	11.45	10.24	9.70	9.33	9.07	8.95	7.16
(9) M.D.	Observed	7	7	7	9	9	4	
	Haskins	10.00	8.00	8.08	8.72	8.08	8.08	5.06
	Macmillan et al.	10.59	8.00	8.11	8.98	8.11	8.19	6.27
	Rosen & Howell	8.88	8.48	8.38	8.32	8.27	8.22	4.40
(10) H.A.	Observed	9	7	12	8	7	11	
	Haskins	8.08	8.08	10.00	8.32	9.28	8.32	2.72
	Macmillan et al.	8.11	8.11	10.56	8.48	9.87	8.57	2.80
	Rosen & Howell	9.50	8.94	8.69	8.56	8.43	8.37	3.87
(11) E.C.	Observed	13	10	8	10	8	5	
	Haskins	8.00	8.32	10.88	13.12	9.28	8.00	9.87
	Macmillan et al.	8.00	10.10	13.25	14.95	11.92	8.00	32.23†
	Rosen & Howell	12.56	11.06	10.30	9.82	9.50	9.33	5.21
(12) T.C.	Observed	10	14	6	10	8	9	
	Haskins	8.00	9.28	13.12	8.72	8.32	8.72	23.84†
	Macmillan et al.	8.00	11.74	14.35	9.01	9.97	10.91	44.40†
	Rosen & Howell	12.56	11.06	10.30	9.82	9.50	9.33	7.47

*p < .05.   **p < .01.   †p < .001.

ically, and these findings replicate those of Experiment 2. The Rosen and Howell prediction did well as a predicter of proximity—none of the 12 listeners deviated in predictions significantly from what was obtained—but there is, of course, no trend to predict the shape of the discrimination function. The fact that the main effect of stimulus comparison was statistically significant $[F(5,55) = 4.21, p < .01]$ means that systematic differences did occur and that the Rosen and Howell prediction of no differences was not borne out.

GENERAL DISCUSSION

Categorical Perception and Plucks and Bows

There are two general interpretations of categorical perception, and it is useful to separate them. Following Macmillan et al. (1977) and Wood (1975),

Table 3
Observed and Three Predicted Scores (of 16) for the Logarithmic Array in Experiment 3

		Stimulus Comparison						
		1-3	2-4	3-5	4-6	5-7	6-8	$\chi^2(6)$
(1) J.S.	Observed	6	9	10	8	9	9	
	Haskins	8.08	8.72	8.32	8.72	8.08	8.08	1.07
	Macmillan et al.	8.11	8.98	9.56	9.66	8.38	8.38	1.23
	Rosen & Howell	9.22	9.22	9.22	9.22	9.22	9.22	2.05
(2) T.S.	Observed	11	10	11	12	13	10	
	Haskins	8.00	8.00	9.28	9.28	9.28	10.00	6.43
	Macmillan et al.	8.00	8.00	11.71	10.54	9.87	12.48	5.65
	Rosen & Howell	9.64	9.64	9.64	9.64	9.64	9.64	3.42
(3) M.C.	Observed	6	9	10	12	10	10	
	Haskins	8.32	8.72	8.08	8.00	8.32	9.28	4.76
	Macmillan et al.	8.96	9.14	8.11	8.00	8.48	12.48	6.77
	Rosen & Howell	8.89	8.89	8.89	8.89	8.89	8.89	3.45
(4) G.N.	Observed	11	9	10	12	10	7	
	Haskins	8.00	8.72	8.72	9.28	9.28	8.00	3.06
	Macmillan et al.	8.00	11.71	10.54	11.28	12.48	8.00	4.62
	Rosen & Howell	9.82	9.82	9.82	9.82	9.82	9.82	2.31
(5) J.B.	Observed	7	10	10	13	12	13	
	Haskins	8.00	8.32	13.12	10.88	8.32	8.08	11.49
	Macmillan et al.	8.00	10.91	14.34	12.03	9.97	8.76	14.25*
	Rosen & Howell	10.56	10.56	10.56	10.56	10.56	10.56	4.95
(6) F.R.	Observed	9	11	10	14	13	10	
	Haskins	8.00	8.00	8.32	10.88	11.92	8.32	4.41
	Macmillan et al.	8.00	8.00	8.98	11.60	13.74	9.97	2.86
	Rosen & Howell	9.64	9.64	9.64	9.64	9.64	9.64	6.22
(7) E.K.	Observed	8	8	10	6	6	9	
	Haskins	8.72	8.08	8.32	8.32	8.72	10.00	2.50
	Macmillan et al.	9.66	8.11	8.48	8.48	9.66	11.28	5.16
	Rosen & Howell	9.08	9.08	9.08	9.08	9.08	9.08	3.61
(8) G.R.	Observed	8	9	8	9	10	11	
	Haskins	8.08	8.08	11.92	10.00	8.08	8.32	5.65
	Macmillan et al.	8.98	8.19	13.74	10.59	8.19	9.97	15.07*
	Rosen & Howell	9.76	9.76	9.76	9.76	9.76	9.76	1.01
(9) M.D.	Observed	7	10	9	11	6	6	
	Haskins	9.28	8.08	8.32	8.72	8.32	8.32	3.78
	Macmillan et al.	9.76	8.11	8.48	9.01	8.48	9.01	4.97
	Rosen & Howell	8.76	8.76	8.76	8.76	8.76	8.76	3.88
(10) H.A.	Observed	6	9	11	12	10	10	
	Haskins	9.28	10.00	8.72	10.00	8.32	8.00	4.36
	Macmillan et al.	11.71	12.48	8.11	12.48	8.98	8.00	13.94*
	Rosen & Howell	9.89	9.89	9.89	9.89	9.89	9.89	3.87
(11) E.C.	Observed	6	11	10	14	14	10	
	Haskins	8.00	8.32	13.12	11.92	8.32	8.08	12.70*
	Macmillan et al.	8.00	10.01	14.93	12.92	10.01	8.76	23.81†
	Rosen & Howell	10.56	10.56	10.56	10.56	10.56	10.56	9.41
(12) T.C.	Observed	8	10	11	7	10	8	
	Haskins	8.00	10.00	10.88	9.28	9.28	8.08	.83
	Macmillan et al.	8.00	12.51	13.18	10.48	11.72	8.76	5.64
	Rosen & Howell	10.56	10.56	10.56	10.56	10.56	10.56	4.97

*$p < .05$.   †$p < .001$.

these are (1) the category-boundary effect and (2) the equivalence of identification and discrimination. The *category boundary effect* for a discrimination task (as opposed to, say, a reaction time task) consists of relatively good discrimination performance between categories and relatively poor performance within categories. Notice that, for this effect to occur, no particular commitment is made as to exact performance level for between- and within-category comparisons or as to the exact location of the discrimination peak. Commitment is made only to the fact that the discrimination peak is generally between categories and that the performance level at the peak is greater than it is in the troughs. Translated into the terms of data analyses in this paper, the trend of discrimination performance ought to be predictable using the various predictive measures, but the performance level may not exactly correspond to those

predicted. In other words, generally speaking, the analysis-of-variance trend tests ought to be significant, but the chi-square analysis need not demonstrate very close correspondence between the two functions.

The *equivalence of identification and discrimination*, on the other hand, makes a stronger commitment to the proximity of the predicted and obtained functions, and relatively less to the exact shape of the curves. Translated into the terms of data analyses in this paper, the overall proximity of discrimination performance ought to be predictable using the various measures, but the trends need not be precisely the same. Thus, generally speaking, the chi-square analysis should demonstrate no difference in the two functions, but the analysis-of-variance trend tests need not be reliable.

That these two notions of categorical perception can be separated may be seen in the data of E.K. for the linear array in Experiment 3 (Table 2). These data show no peak in either obtained or predicted discrimination curves, but neither the Haskins nor the Macmillan et al. prediction is discrepant from the obtained scores. Thus, for E.K. we have an instance of general identification-discrimination equivalence without a category-boundary effect. The data of J.E.C. in Experiment 2 show exactly the opposite effect: there is a strong trend in the discrimination function predicted from both the Haskins and Macmillan et al. models, but the predicted and obtained functions are far apart. Thus, there is a boundary effect but not equivalence.

The point is that both of these patterns of results, the category-boundary effect and identification-discrimination equivalence, can be called categorical perception. It is probably safest, however, to suggest that categorical perception occurs only when both effects are generally found, and when only one of the effects occurs, categorical perception is only suggested. When neither occurs, of course, categorical perception has not occurred. With these ideas in mind, reconsider the data presented in the three experiments.

Linear increments. The results of J.S. and T.S. in Experiment 1 and the pooled results of 12 listeners in Experiment 3 strongly suggest that categorical perception does not occur for sawtooth wavetrain stimuli that differ in rise time along linearly spaced increments. Neither the category-boundary effect nor the identification-discrimination equivalence occurs for either the Haskins or the Macmillan et al. predictions in the three data sets.

Logarithmic increments. The data of J.S. in Experiment 2 show a category-boundary effect from the points of view of both the Haskins and Macmillan et al. predictions. Identification-discrimination equivalence, on the other hand, is indicated only for the

Macmillan et al. prediction. Thus, three of four measures for J.S. suggest categorical perception.

The data of T.S. are somewhat more puzzling. From the point of view of both the category-boundary effect (defined by the trend test) and identification-discrimination equivalence (defined by the goodness-of-fit test), T.S.'s data are categorical. Yet, on grounds of the general shape of the obtained discrimination function, one is loathe to declare outright that these data are really categorically from either the Haskins prediction or the Macmillan et al. prediction. Note that the identification boundary is quite discrepant from the other listeners, and even from T.S.'s own data in Experiment 1. The only rationale I can suggest for this is that the extended boundary may have something to do with the fact that T.S. suffered a small rupture of one eardrum between the time of Experiments 1 and 2, but was declared clinically sound within 10 days of the incident. Regardless of the reason for the effect, it is probably safest to state simply that identification-discrimination equivalence was found, but that, because the continuum was not sampled far enough into the long rise-time domain, the category-boundary effect is less assured. Thus, two of the two equivalence measures were met for T.S., with the two boundary measures not readily assessable.

The data of M.C. are also somewhat puzzling. In this case, neither the boundary effects nor the equivalence effects occurred for either the Haskins prediction or the Macmillan et al. prediction. Yet, there is clearly a peak in the discrimination function, although somewhat misplaced. Perhaps the best that can be said is that none of the four measures are satisfied completely, but that, because the data show a displaced boundary effect, categorical perception according to that criterion is not completely ruled out. Thus, the two equivalence measures were not met, and the two boundary measures are, again, not readily assessed as having failed.

The data of J.E.C. are more clear-cut. Both boundary measures are satisfied, those for the Haskins and the Macmillan et al. predictions, but neither of the equivalence measures is met.

The data of the 12 listeners of logarithmic array in Experiment 3 are also more straightforward. Although some deviation occurs for the individual equivalence measures in the Macmillan et al. prediction, as a group the listeners' data generally meet both equivalence criteria and clearly meet both boundary criteria.

Thus, if one considers the four listeners in Experiment 2 and the one group in Experiment 3 as five separate occasions in which both boundary and equivalence measures of categorical perception can be taken, then the following overall assessment can be made: 6 of 6 boundary measures (determined by

analysis-of-variance trend tests between the two predicted and one obtained function) are satisfactorily met, with 4 others indeterminate (those of T.S. and M.C. in Experiment 2, for different reasons), and 5 of 10 equivalence measures are satisfactorily met. In general, then, I suggest that categorical perception occurred for sawtooth wavetrains when items along the continuum were generated in equal logarithmic steps.

These data, like those of Rosen and Howell (1981), suggest substantial individual differences in identification and discrimination performance. Discussion of such differences, however, is beyond the intended scope of this paper. The similarities among the data, to my eye, outweigh any differences that are revealed.

Performance Level and Stimulus Density

Scrutiny of Figures 3 and 4 reveals that the peaks that occur in both the predicted and obtained discrimination functions for the logarithmic arrays are not outstandingly high. Liberman et al. (1957), for example, found peak predicted and obtained performances for two-step comparisons of 85% and above, whereas those values in the present study are seldom much above 65%. Indeed, one might think that neither predicted nor obtained peaks are above chance. This, however, is false. The peaks of the Haskins predictions, the lowest of the categorical predictions in Experiments 2 and 3, are all reliably above chance [ts(11) = 4.66, 6.69, 5.34, and 3.11, for J.S., T.S., M.C. in Experiment 2 and the group in Experiment 3, respectively, ps < .01]. And the peaks of the obtained functions are also reliably above chance [ts(11) = 7.18, 4.56, 3.37, and 3.47, for J.S., T.S., M.C., and the group, respectively, ps < .01].

The relatively low peaks in the obtained and predicted functions appear to be due to the relatively higher density of stimuli along the continuum. In Experiment 1, for example, my two-step comparisons are closer to the one-step comparisons used by others, such as Pisoni and Lazarus (1974). That is, since the stimuli were generated in 5-msec increments, the two-step comparison stimuli are 10 msec apart; in Pisoni and Lazarus (1974), they were generated in 10-msec increments, so that two-step comparisons are 20 msec apart. Since the log stimuli used in Experiments 2 and 3 are roughly as dense in the 15- to 60-msec region as are the linear stimuli, I would argue that they too should be thought of in these terms: that the results are perhaps better compared (in terms of peaks and troughs) with the one-step comparisons of others. Pisoni and Lazarus (1974, Figure 2), for example, show predicted and obtained peaks for /ba/-/pa/ stimuli in the 60%-70% range. Thus, I would contend that the phenomenon demonstrated here is about as robust as many speech results for categorical perception.

A Reassessment of the Import of Plucks and Bows

In the early and middle 1970s, there were several paradigmatic results thought to be peculiar to speech perception. Several were outlined by Wood (1975), and they were also summarized by Cutting (1978) and Schouten (1980). Four results seemed particularly salient: (1) categorical perception in both adults and infants, (2) right-ear/left-hemisphere advantages, (3) asymmetric integrality with redundancy gain, and (4) selective adaption. All of these had been found for speech sounds, had not yet been found for nonspeech sounds, and were thought to converge on the auditory-phonetic distinction. Plucked and bowed sounds were the first (and possibly the only) set of nonspeech sounds to yield all four effects: categorical perception in adults (Cutting & Rosner, 1974, 1976; Cutting, Rosner, & Foard, 1976) and infants (Jusczyk, Rosner, Cutting, Foard, & Smith, 1977), right-ear/left-hemisphere advantages (Blechner, Note 3), asymmetric integrality with redundancy gain (Blechner, Day, & Cutting, 1976), and selective adaptation (Cutting et al., 1976; Remez, Studdert-Kennedy, & Cutting, 1980). It is important to point out that this list was, at best, a concatenative set. It was thought that the whole pattern revealed the anatomy of phonetic perception, not simply any one particular result. The reason for this is that many of the separate results had been (or have now been) obtained with stimuli other than speech. For example, categorical-like perception had been found for the perception of musical intervals (Locke & Kellar, 1973; Siegel & Siegel, 1977), for onsets of buzz-tone (Miller, Wier, Pastore, Kelly, & Dooling, 1976) and two-tone stimuli (Pisoni, 1977), for critical flicker fusion (Pastore, 1976), and for several others. Yet none of these stimuli had been shown to yield left-hemisphere effects, or asymmetric integrality. Likewise, some stimuli had been shown to reveal left-hemisphere effects (Bever & Chiarello, 1974; Halperin, Nachshon, & Harmon, 1973), but neither categorical perception nor asymmetric integrality. Moreover, certain stimuli have revealed patterns of asymmetric integrality (Pomerantz & Sager, 1975) but have not been shown to yield categorical perception or left-hemisphere effects. Given that Rosen and Howell (1981) have demonstrated that the original plucked and bowed sounds are not perceived categorically, we now have hard evidence from Cutting, Rosner, and Foard (1976) that effects of selective adaptation are independent of categorical perception as well, and, given the extensive literature in vision stemming from the work of Blakemore and Campbell (1969), this is what we should expect.

To be sure, categorical perception was the cornerstone of the apparent edifice for peculiarities of speech at the time of the Cutting and Rosner (1974) study, and this is certainly why those results, along with

those of Miller et al. (1976), received particular attention. But, even if categorical perception did not occur for plucked and bowed sounds, those other results are not impugned, for none of them are dependent on the prior results of categorical perception. Since the results of experiments reported here demonstrate categorical perception for logarithmically arrayed plucked and bowed sounds, nearly the whole edifice is still in place. Only the results of Jusczyk et al. (1977) still suffer from the inappropriateness of the original stimuli, and the possibility of finding categorical perception for these sounds in infants remains unfulfilled.

It happens, however, that researchers in speech perception do not use this particular conflation of results as hallmarks of speech perception (Liberman & Studdert-Kennedy, 1977), and fully admit that these results occur throughout perception. In fact, they do not always occur for speech sounds. For example, there are many ways to get *non*categorical perception of stop consonants (see, e.g., Barclay, 1972, Carney, Widin, & Viemeister, 1977, Pisoni & Lazarus, 1974, and Sachs & Grant, Note 4), and I will have more to say about this in the next section. Thus, the force of the research program of investigating parallels between the perception of speech and music-like sounds is considerably diminished. It seems of little import to speech perception today whether or not plucked and bowed sounds are perceived categorically. Where it does matter, as in the work of Stevens (in press), categorical perception of non-linguistic sounds could still play a role—contrary to the conclusion of Rosen and Howell (1981).

Where categorical perception of plucks and bows matters to a greater degree is in the study of the psychophysical relationship between identification and discrimination tasks. For Macmillan et al. (1977), for example, it is largely irrelevant whether the continuum of stimuli is generated in linear or logarithmic increments; it matters only that discrimination performance can be predicted on the basis of identification performance. But it appears that this relationship holds for plucked and bowed sounds generated logarithmically, but not linearly. This discrepancy poses a problem for analyses such as those of Macmillan et al. (1977). Additional assumptions such as those from range-frequency analysis (Parducci & Perritt, 1971) may be needed. Moreover, that a linear-increment continuum does not yield a categorical discrimination function, a finding of both Rosen and Howell (1981) and this paper, is inconsistent with the results of Macmillan et al. (Note 1). More research is needed before these differences can be reconciled.

Categorical Perception and Stimulus Uncertainty

The original demonstration of categorical perception in speech (Liberman et al., 1957) used an ABX discrimination procedure. This procedure is something of an oddity, since ABB comparisons are typically much easier to judge than are ABA comparisons. These differences, and several kinds of empirical results, eventually gave rise to much discussion of auditory and phonetic memories for speech (see, e.g., Pisoni, 1973). The general success of the ABX paradigm led to its continued use. In fact, it was generally taken that no perceptual results could be determined to be indicative of categorical perception unless the ABX paradigm was used, or at least one similar to it. More recently, same-different tasks and others more amenable to signal detection analyses have also been used, but since the work of Macmillan et al. (1977; see also Creelman & Macmillan, 1979) interconversion and comparisons of results from different tasks have seemed to become easier.

The most important aspect of the ABX task, however, may be a methodological attribute orthogonal to task form. The typical ABX task used in most speech perception research, and in those nonspeech studies intended as one or another kind of control condition, has been a task of *high stimulus uncertainty*. That is, the listener knows only that Stimuli A and B are different; he or she does not know the assignment of members of the stimulus array to Stimuli A and B. To be more concrete, using an example from the experiments in this paper, the listener may know that Stimuli A and B differ by 10 msec rise time, but he or she does not know on a particular trial that Stimulus A will be the item with 25-msec rise time and Stimulus B, that with 35-msec rise time. If the listener is given such knowledge on a trial-to-trial basis, the task becomes one of *low stimulus uncertainty*. What is interesting is that we now know that, under such conditions, categorical perception is not found for stop consonants (Carney et al., 1977; Sachs & Grant, Note 4). Given that stop consonants do not yield categorical perception under such conditions, one would not expect plucked and bowed sounds to do so either. Since van Heuven and van den Broecke (1979) used a method of adjustment task, one of low stimulus uncertainty, in their investigation of the perception of onset times, it comes as no surprise that they did not find those sounds to be perceived categorically.

The major conclusion about stimulus uncertainty in a discrimination task seems to be as follows: If one is interested in exploring the capabilities of sensory and perceptual systems working at their limit, then a low stimulus uncertainty task would seem most appropriate. If, on the other hand, one is interested more in what the sensory and perceptual systems may be doing under conditions closer to real life, then a high stimulus uncertainty task may be more appropriate. The rationale for this latter conclusion should be transparent: As we move about in our environment, interacting with our conspecifics

and our surrounds, we are under conditions of considerably more stimulus uncertainty than those in the usual psychophysical situation. To capture something of that state, a high stimulus uncertainty task may be more successful at determining what the organism does do, rather than what it can do.

CONCLUSION

Four conclusions can be drawn from the results presented here coupled with those of previous papers. The first is that categorical perception *can* be found for the perception of certain arrays of plucked and bowed sounds. Thus, the claim of Cutting and Rosner (1974) that such sounds can be perceived categorically is upheld; but the data on which that claim was made are no longer valid (Rosen & Howell, 1981).

The second is that it is no longer very clear exactly what the conditions are for categorical perception of plucked and bowed sounds. Remez (1978) has found such results for stimuli covarying in rise time and amplitude at onset. Macmillan (Note 1) has found such results for linearly arrayed stimuli, but Rosen and Howell (1981) and I have not. I have now found such results for logarithmically arrayed stimuli, and there have yet to be attempts at replication. Thus far, all positive results share the method if using high stimulus uncertainty, but, given the different results for different scale increments found in this paper, stimulus uncertainty is only part of the answer. The other parts have yet to be determined.

The third conclusion is that the general fickleness of the phenomenon of categorical perception in any stimulus domain makes it unworthy of being held in too high esteem as a touchstone for a particular kind of perceptual process. The fact that stimulus uncertainty and range-frequency effects may cause the phenomenon to appear and disappear render categorical perception more of a curiosity than a fundament of a particular kind of perception. It is no longer the snappy result it once appeared to be.

And the final conclusion is that the curiosity that categorical perception once presented to psychophysics appears now to be doubly curious. Categorical perception, construed as general identification-discrimination equivalence, is fairly unusual in psychophysical situations. The much more common circumstance, of course, is that ability to discriminate stimuli considerably outstrips one's ability to identify them. It is the lack of difference between the two abilities that drove Macmillan et al. (1977) to be interested in the phenomenon. Now that the phenomenon appears subject to factors of stimulus range and frequency, in addition to those of stimulus uncertainty (Creelman & Macmillan, 1979), there is yet another factor to be accounted for.

REFERENCE NOTES

1. Macmillan, N. A. *Categorical perception of musical sounds: The psychophysics of plucks and bows.* Paper presented at the 20th annual meeting of the Psychonomic Society, Phoenix, November, 1979. (Abstract reprinted in *Bulletin of the Psychonomic Society*, 1979, 11, 241.)
2. Macmillan, N. A. Personal communication, November 18, 1980.
3. Blechner, M. J. *Right-ear advantage for musical stimuli differing in rise time* (Status Report on Speech Research SR-47). New Haven, Conn: Haskins Laboratories, 1976.
4. Sachs, R. M., & Grant, K. W. *Stimulus correlates in the perception of voice onset time (VOT): II. Discrimination of speech with high and low uncertainty.* Paper presented at the 92nd meeting of the Acoustical Society of American, San Diego, November 1976. (Abstract reprinted in *Journal of the Acoustical Society of America*, 1976, 60, S91.)
5. Nye, P. W. Letter to Burton S. Rosner, February 27, 1981.

REFERENCES

BARCLAY, J. R. Noncategorical perception of a voiced stop: A replication. *Perception & Psychophysics*, 1972, 11, 269-273.
BEVER, T. G., & CHIARELLO, R. J. Cerebral dominance in musicians and nonmusicians. *Science*, 1974, 195, 537-539.
BLAKEMORE, C., & CAMPBELL, F. W. On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology*, 1969, 203, 237-260.
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.
CARNEY, A. E., WIDIN, G. P., & VIEMEISTER, N. F. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 1977, 62, 961-970.
CREELMAN, C. D., & MACMILLAN, N. A. Auditory phase and frequency discrimination: A comparison of nine procedures. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, 5, 146-156.
CUTTING, J. E. There may be nothing peculiar to perceiving in the speech mode. In J. Requin (Ed.), *Attention and performance VII.* Hillsdale, N.J: Erlbaum, 1978.
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. Discrimination functions predicted from categories in speech and music. *Perception & Psychophysics*, 1976, 20, 87-88.
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.
HALPERIN, Y., NACHSHON, I., & CARMON, A. Shift in ear superiority in dichotic listening to temporal pattern nonverbal stimuli. *Journal of the Acoustical Society of America*, 1973, 53, 46-50.
HAYS, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.
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.
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.

LIBERMAN, A. M., & STUDDERT-KENNEDY, M. Phonetic perception. In R. Held, H. Leibowitz, & H. L. Teuber (Eds.), *Handbook of sensory physiology* (Vol. 8). Heidelberg: Springer-Verlag, 1977.

LOCKE, S., & KELLAR, L. Categorical perception in a nonlinguistic mode. *Cortex*, 1973, **9**, 355-369.

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

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**, 410-417.

PARDUCCI, A., & PERRETT, L. Category rating scales: Effects of relative spacing and frequency of stimulus values. *Journal of Experimental Psychology Monographs*, 1971, **89**, 427-452.

PASTORE, R. E. Categorical perception: A critical re-evaluation. In S. K. Hirsh, D. H. Eldredge, I. J. Hirsch, & S. R. Silverman (Eds.), *Hearing and Davis*. St. Louis: Washington University Press, 1976.

PASTORE, R. E., & SCHEIRER, C. J. Signal detection theory: Considerations for general application. *Psychological Bulletin*, 1974, **81**, 945-958.

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

PISONI, D. B. Identification and discrimination of the relative onset of two component tones: Implications for voicing performance 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**, 328-333.

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

POMERANTZ, J. R., & SAGER, L. C. Asymmetric integrality with dimensions of visual patterns. *Perception & Psychophysics*, 1975, **18**, 460-466.

REMEZ, R. E. An hypothesis of event-sensitivity in the perception of speech and bass violins (Doctoral dissertation, University of Connecticut, 1978). *Dissertation Abstracts International*, 1978, **39**, 5618-B. (University Microfilms No. 7911404)

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

ROSEN, S. M., & HOWELL, P. Plucks and bows are not categorically perceived. *Perception & Psychophysics*, 1981, **30**, 156-168.

SCHOUTEN, M. E. H. The case against a speech mode of perception. *Acta Psychologica*, 1980, **44**, 71-98.

SIEGEL, J. A., & SIEGEL, W. Categorical perception of tonal intervals: Musicians can't tell sharp from flat. *Perception & Psychophysics*, 1977, **21**, 399-407.

STEVENS, K. N. Constraints imposed by the auditory system on 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.

VAN HEUVEN, V. J. J. P., & VAN DEN BROECKE, J. P. R. Auditory discrimination of rise and decay time in tone and noise bursts. *Journal of the Acoustical Society of America*, 1979, **66**, 1308-1315.

WINER, B. J. *Statistical principles in experimental design* (1st ed.). New York: McGraw-Hill, 1962.

WOOD, C. C. Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analyses. *Journal of Experimental Psychology: Human Perception & Performance*, 1975, **1**, 3-20.

NOTES

1. Rosen and Howell (1981) also entertain a variable Weber-fraction model, which allows considerable modification to a perfectly logarithmically perceived continuum. However, their data in Experiment 1, though not that in Experiment 3, are better fit by a pure Weber fraction. Moreover, it is not clear what the justification for a variable Weber-fraction model would be.

2. For computational purposes, Macmillan et al. (1977, p. 465) interpreted 0% and 100% in their reanalyses of Cutting and Rosner's (1974) data as 1% and 99%, respectively. This allows the maximum value of the z-transformed performance to be ±2.33. These values struck me as perhaps too extreme, overestimating small differences in performance at the limit. Thus, the calculations used here interpret 0% and 100% performance as not different from, and hence equal to, 5% and 95%, yielding maximal z scores of ±1.96.

3. There is, of course, no guarantee that the mean identification distance should provide a proper estimate of general discriminability, but this is the best metric available that uses all the data. Also, there is a fourth predicted function, one that is used by Fujisaki and Kawashima (see Pisoni, 1975), which is very much like the Haskins prediction except that it assumes an additional factor of short-term memory to account for the typically greater-than-chance performance within categories found for some classes of speech stimuli, such as vowels. Since it involves an additional parameter and seemed not particularly appropriate to this research effort, it was not used here.

4. An alternative way to calculate the predicted function would be to take the z-transformed identification distances for the sum of the responses over the 12 sessions and calculate the prediction that way. This would conform to a more usual psychophysical approach. This was not done because I felt that it might ignore some very real differences in the listener across sessions. Moreover, summing across these sessions tends to depress values toward chance performance. In essence, I judged that the mean of 12 moderately good estimates of listener performance would be better than 1 very good estimate. In practice, this proved to be true for all listeners in Experiments 1 and 2.

5. Consider the following example, taken from the results of J.S. in Experiment 2. If the predicted performance (out of 144) on each of the seven comparisons is 80.6, 91.5, 95.3, 100.8, 101.5, 81.9, and 74.2, respectively, then the mean prediction across all comparisons is 89.4. Subtracting this value from each total yields scores of -8.8, 2.1, 5.9, 11.4, 12.1, -7.5, and -15.2. Then, following Winer (1962, pp. 132-134), these, instead of orthogonal polynomials, are used as trend weights for each comparison.

6. All chi-square values were determined in the following manner: A chi-square test was performed on the 2×2 table of observed and predicted scores for each comparison. Using Yates' correction (Hays, 1963, pp. 585-586), the formula is:

$$\chi^2 = (|O-E|-.5)^2/E + (|O-E|-.5)^2/(N-E), \qquad (3)$$

where O is the observed discrimination performance, E is the expected and N the number of observations per discrimination comparison. Where O and E are within .5 of one another, chi-square is set to zero. Since there are seven comparisons in Experiments 1 and 2, each with one degree of freedom, a chi-square score with seven degrees of freedom is achieved by adding up the seven chi-squares. A chi-square with six degrees of freedom was used in Experiment 3, since there were only six comparisons.

APPENDIX

The original stimuli used by Cutting and Rosner (1974) were recorded at the University of Pennsylvania from a Moog synthesizer onto an Ampex AG500 tape recorder. Many hundreds of tokens were recorded, dozens each of
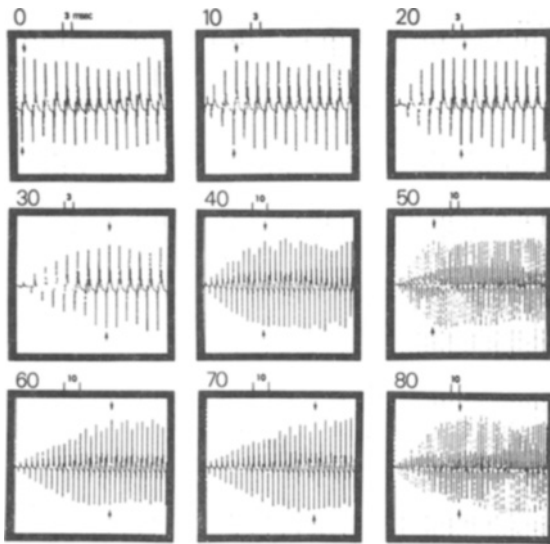
Figure A1. High-resolution digital oscillograms of the original 294-Hz Cutting and Rosner (1974) stimuli. Nominal rise time values are given in the upper left of each of the nine panels; millisecond increments are given at the top of each panel to determine the resolution of the oscillogram; and arrows are placed at the negative and positive peaks in the oscillogram, marking maximum amplitude. Rise times for these digital representations are given in Table A1.

different frequencies and waveforms at many different rise-time settings. The rise times were approximated at that point by reading off the potentiometer responsible for onset variation. The audiotape was taken to the Haskins Laboratories, and the entire recording session was digitized on magnetic tape. Next, the digitized waveforms were inspected and stimuli closest to the nominal rise times—0 through 80 msec—were stored in disk files for both 294- and 440-Hz stimuli. At the time the stimuli were stored, hard-copy high-resolution oscillograms were made of all stimuli. Over the years, I retained only those for the 294-Hz stimuli, and these are shown in Figure A1.

These were measured for appropriate rise time by eye at the time the stimuli were stored. Recently, I measured them again using a modification of the technique described by Rosen and Howell (1981). Those rise times for the digital waveforms are shown in Table A1, along with the nominal values reported by Cutting and Rosner (1974) and the analog values reported by Rosen and Howell (1981). Arrows in Figure A1 indicate the maximum negative-to-positive peak difference used to determine the digital rise time values. These values are not nearly as discrepant from the nominal values as those reported by Rosen and Howell, but they are slightly off.

How these digital stimuli got to be so different when converted into analog form is incompletely known, but the best account I can offer is given below, and the details come from Nye (Note 5). The Haskins system, as it existed in the Spring of 1974 when the original tapes were made, band-passed the acoustic signal between 200 and 3,900 Hz with a sampling rate of 10 kHz. Since the system was built to deal almost exclusively with speech signals, it was tailored to the characteristics of the speech waveform. Upon digitization, the input signal's high frequencies were enhanced by 6 dB per octave above 500 Hz. Upon digital-to-analog conversion, the upper frequencies were deemphasized by the same amount. The waveforms shown in Figure A1 reveal this high-frequency amplification by the sharp peaks at each negative- and positive-going pulse. Notice that these do not exist in the analog versions, as shown by Rosen and Howell (1981, Figure 4). Note further that it was these emphasized peaks that were used to determine the nominal rise-time values. The sawtooth waveforms used by Cutting and Rosner (1974) had much more high-frequency energy than the usual speech signal. The high-frequency deemphasis filtering of the system apparently could not handle all this energy, and it was converted into low frequency and added to the output spectrum. In engineering, this effect is known as aliasing. In this manner, the rise-time values for the digital waveforms and the analog waveforms can be made to come into better register.

In addition to the discrepancies in rise times, there is also the differences in quality of some of the stimuli. Rosen and Howell (1981) report "ringing"-like sounds and other artifacts. These are no doubt present. The original stimuli were chosen, not because of their absolute identity along all dimensions and rise time, but because of their relative identity coupled with their general naturalness of sound. Moog synthesizers, after all, are musical instruments. The DEC PDP-12 of Rosen and Howell and the Hewlett-Packard 1000L of the current study are not musical instruments. Moreover, whereas these artifacts, which Rosen and Howell report, may account for most of the variance in the data of Cutting and Rosner (1974), they seem to do less well for the data of Cutting et al. (1976), who used exactly the same stimuli.

Table A1
Rise-Time Values (in Milliseconds) Reported by Cutting and Rosner (1974) and by Rosen and Howell (1981) and as They Occurred for Stimuli in Digital Format

	Stimulus								
	1	2	3	4	5	6	7	8	9
Nominal Values Reported by Cutting and Rosner (1974)	0	10	20	30	40	50	60	70	80
Analog Values Reported by Rosen and Howell (1981)	4	6	15	18	35	43	56	67	72
Digital Values From High-Resolution Digital Oscillograms	2	12	21	28	39	46	61	71	82

Note—All values reported are for 294-Hz stimuli only.