*Perception & Psychophysics*
*1974, Vol. 16 (3), 564-570*

# Categories and boundaries in speech and music*

JAMES E. CUTTING†
*Yale University and Haskins Laboratories, New Haven, Connecticut 06510*

and

BURTON S. ROSNER
*University of Pennsylvania, Philadelphia, Pennsylvania 19104*

Perception & Psychophysics
1974, Vol. 16 (3), 564-570

# Categories and boundaries in speech and music*

JAMES E. CUTTING†
*Yale University and Haskins Laboratories, New Haven, Connecticut 06510*

and

BURTON S. ROSNER
*University of Pennsylvania, Philadelphia, Pennsylvania 19104*

Perceptual categories and boundaries arise when Ss respond to continuous variation on a physical dimension in a discontinuous fashion. It is more difficult to discriminate between members of the same category than to discriminate between members of different categories, even though the amount of physical difference between both pairs is the same. Speech stimuli have been the sole class of auditory signals to yield such perception; for example, each different consonant phoneme serves as a category label. Experiment I demonstrates that categories and boundaries occur for both speech *and* nonspeech stimuli differing in rise time. Experiment II shows that rise time cues categorical differences in both complex and simple nonspeech waveforms. Taken together, these results suggest that certain aspects of speech perception are intimately related to processes and mechanisms exploited in other domains. The many categories in speech may be based on categories that occur elsewhere in auditory perception.

Speech is replete with perceptual categories, as so much research at the Haskins Laboratories has shown (see Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman, Harris, Hoffman, & Griffith 1957; Lisker & Abramson, 1964; Pisoni, 1971, 1973; among many others). Different phonemes, particularly the sop consonants, serve as category labels when synthetic speech stimuli are varied along a particular dimension and the listener must identify each item in the array. Typically, continuous physical change yields markedly discontinuous percepts. For example, when the slope of the second-formant transition in certain speech patterns is varied in equal steps by adjusting its starting frequency, the resulting stimuli are perceived as [ba] or [da] or [ga] but never as anything else and rarely as anything in between. Not only are these identifications quantal, but listeners often discriminate poorly between two acoustically different but phonemically identical items. In contrast, Ss can more easily discriminate items separated by the same amount of acoustic difference when accompanied by a phonetic difference. This curious nonlinearity is known as categorical perception. The discontinuity in the discrimination function points to the locus of a perceptual boundary.

Auditory signals carrying no linguistic information (so-called "nonspeech" sounds) might also have categories and boundaries. Clearly, however, many nonspeech sounds do not. No stable boundaries occur for sine waves of different frequencies (Sawusch & Pisoni, 1974). None have been found in ABX discrimination tasks for spectral inversions of speech stimuli (Liberman, Harris, Kinney, & Lane, 1961), for "chirps" and "bleats," which are brief segments of speech stimuli normally carrying important phonetic information in a speech context (Mattingly, Liberman, Syrdal, & Halwes, 1971), or for speech-like sounds with phonemically irrelevant formant transitions (Cutting, in press).

Lane (1965) suggested that categorical-like discrimination functions could be induced for nonspeech patterns through a moderate amount of S training. However, Studdert-Kennedy, Liberman, Harris, and Cooper (1970) pointed out that there are three necessary criteria for categorical perception: (a) "peaks," regions of high discriminability in the discrimination function, (b) "troughs," regions where discrimination performance is near chance, and (c) a correspondence between the peaks and troughs and the shape of the identification functions, with peaks occurring at the identification boundaries and troughs within each category. Lane's data did not meet the last two criteria. More recently, Miller, Pastore, Wier, Kelly, and Dooling (1974) have met all three criteria with buzz and noise sounds varying in temporal relationship. The discrimination functions for their stimuli, which simulate aspects of voiced and voiceless stop consonants, have peaks and troughs that correspond to identification functions. However, unlike speech sounds, their stimuli cannot be

---

*This research was supported by NICHD Grant HD-01994 to the Haskins Laboratories. Requests for reprints should be sent to the first author at Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510.

†Now at Wesleyan University, Middletown, Connecticut, and at Haskins.

labelled, except in an arbitrary fashion.

Can normally occurring stimuli meet the three criteria for categorical perception? A logical candidate here is musical sounds. Locke and Kellar (1973), using musicians and nonmusicians as Ss, varied the middle components of triadic chords in a search for categorical perception. Their results indicate some categorization by musicians, but considerably less by nonmusicians. More importantly, neither group yielded results as compelling as those usually found for speech. In the present studies, we investigated the identification and discrimination of selected music-like sounds, and, for comparison, of speech syllables. We chose to vary rapidity of stimulus onset, called attack or rise time.

## EXPERIMENT I

### Method

**Stimuli.** Two classes of stimuli were synthesized for identification: 18 speech stimuli and 18 nonspeech stimuli. Nonspeech stimuli were sawtooth waves generated on the Moog synthesizer at the Presser Electronic Studio at the University of Pennsylvania. (The "sawtooth" on this instrument suffers from some low-pass filtering.) Two nine-item arrays consisted of stimuli which differed solely in their onset characteristics. One array was synthesized at 440 Hz, and the other at 294 Hz. Amplitude envelopes reached maximum intensity in 0, 10, 20, 30, 40, 50, 60, 70, or 80 msec after onset. By 0-msec rise time, we mean that a stimulus reached maximum amplitude in one-fourth of a period. Rise times were measured by digitizing the waveforms and displaying them with high resolution on a computer-controlled oscilloscope. The rapid-onset stimuli sounded like the plucking of a stringed instrument, whereas the slower onset stimuli sounded like the playing of the same instrument with a bow. The durations of the nonspeech stimuli were between 1,020 and 1,100 msec, varying according to rise time. Oscillograms of stimuli with 10- and 70-msec rise times are shown in Fig. 1. The sawtooth stimuli had some low-frequency amplitude modulation due to the Moog oscillator. Note that the decay of the Moog items lasts about 1 sec. As decays become more rapid, or as the tails are trimmed, the items are less convincing as notes which could originate from stringed instruments.

Speech stimuli were generated on the Haskins Laboratories' parallel-resonance synthesizer. Like the nonspeech stimuli, they formed two nine-item arrays, with members of each array differing in rise time by 10-msec increments, from 0 to 80 msec. In one array, items were identifiable as either [tʃa] as in *chop* or [ʃa] as in *shop*, and in the other array as either [tʃae] as in *chad* or [ʃae] as in *shad*. All speech stimuli shared the same pitch contour and were between 410 and 490 msec in duration, differing again according to rise time. Oscillograms of speech syllables with 10- and 70-msec rise times appear in Fig. 1. Sine waves with 10- and 70-msec rise times are shown for comparison purposes.

All stimuli were recorded on audio tape, then digitized, edited, and stored on a disk file using the PCM system at Haskins (Cooper & Mattingly, 1969). Stimuli were reconverted to analog form at the time test tapes were recorded for each task.

**Tapes.** Two identification and two discrimination tapes were recorded, one of each for the two classes of stimuli. Identification tapes consisted of a random sequence of 144 items: (2 arrays) by (9 items per array) by (8 observations per items). Onset-to-onset time was 4 sec, with a 7-sec pause between blocks of 12 items.

Discrimination tapes consisted of ABX stimulus triads with 2.1 sec between onsets of items within a triad. Since the Moog items average 1,060 msec in duration, the 2.1-sec value was selected to make the average interstimulus interval (ISI) 1.04 sec, very close to the 1-sec ISI used by Pisoni (1971). Since the crucial information for discrimination in both musical and speech stimuli occurs at or near stimulus onset, we felt that the onset-to-onset interval of 2.1 sec should be used for the affricate/fricative stimulus triads as well. For these triads, the average ISI was 1.55 sec. The interval between all triads was 5 sec. For both sets of stimuli the four permutations of each ABX comparison were represented: ABA, ABB, BAB, and BAA. All comparisons were two-step comparisons; that is, Stimulus A and Stimulus B were members of the same array and differed in rise time by 20 msec. Thus, 0- and 20-msec items were compared, 10 and 30, 20 and 40, 30 and 50, 40 and 60, 50 and 70, and 60 and 80, yielding seven possible comparisons for each array. Discrimination tapes consisted of a random sequence of 56 triads: (2 arrays per stimulus class) by (7 comparisons per array) by (4 ABX permutations).

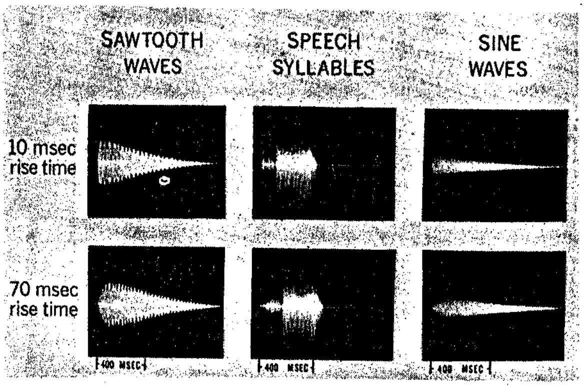


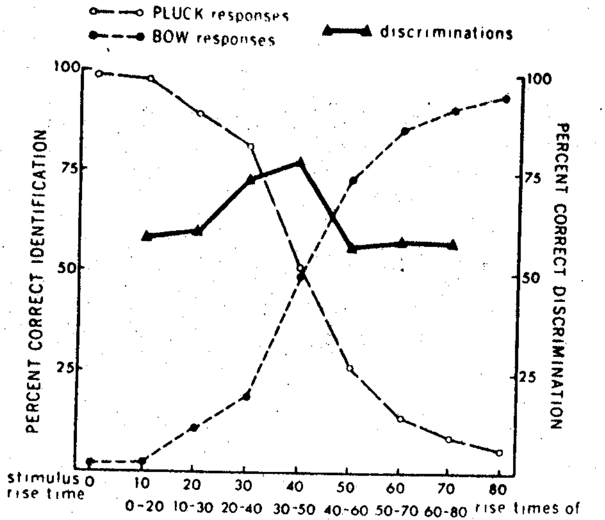
Fig. 1. Oscillograms of sample stimuli used in Experiments I and II.

**Subjects, Apparatus, and Procedure.** Twenty Yale University undergraduates participated in two tasks on each of 2 testing days as part of a course requirement. They were not selected according to musical ability. Audio tapes were played on an Ampex AG500 tape recorder and broadcast over an Ampex 620 loudspeaker in a partially sound-attenuating room. On the first day, Ss listened to the musical sounds. For preliminary training, the end-point stimuli were played in an alternating sequence five times each at both frequencies. Ss were told to regard the 0-msec stimuli as a *plucked* string of a musical instrument (like that of a guitar) and to regard the 80-msec stimuli as a *bowed* string of a musical instrument (like that of a violin). Ss readily agreed that these labels were easy to use. During the identification task, they checked off their response to each stimulus, *pluck* or *bow*, on a prepared response sheet. During the discrimination task, they listened to each triad and wrote A or B, indicating which of the first two items in the triad they felt was identical to the third item.

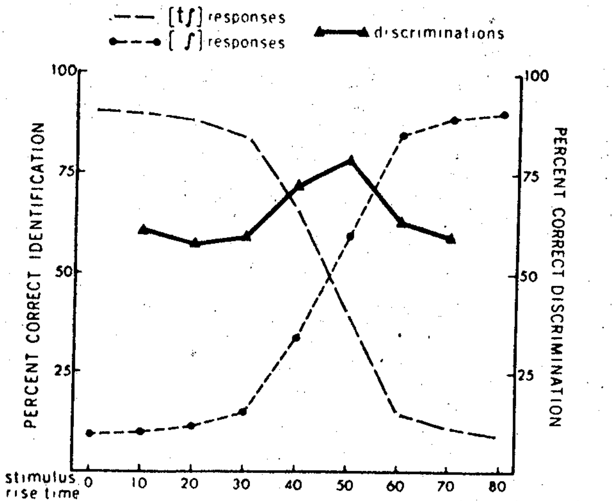
On the second day, Ss listened to the speech sounds. The tasks and basic instructions were the same as the first day, except that during the identification task Ss checked off their response, CH or SH, for each item.

### Results and Discussion

**Sawtooth Waves.** The top panel of Fig. 2 shows that identifications of the musical stimuli were quite categorical. It combines results for the 294- and 440-Hz arrays, which did not differ in their effects. The stimuli with rise times of 0, 10, 20, and 30 msec were identified as a plucked sound on 92% of all trials; the 50-, 60-, 70-, and 80-msec stimuli were identified as bowed sounds on 87% of all trials; only



SAWTOOTH WAVES



SPEECH SYLLABLES

**Fig. 2. Mean identification and discrimination functions for sawtooth waves and speech syllables.**

the 40-msec stimulus was ambiguous. The ABX discrimination function is overlaid on the identification results and shows a pronounced peak at the rise-time comparisons of 20-40 msec and 30-50 msec. Each point is significantly greater by a sign test than its adjacent within-category comparison, 10-30 msec ($z = 2.23$, $p < .02$) and 40-60 msec ($z = 3.65$, $p < .001$). There were no significant differences among the other discriminations. We checked whether these results could reflect clicks induced at short rise times in the loudspeaker. We found such artifacts only at 0 and 10 msec, but not at 20 msec and longer.

Obtained discrimination values for each comparison and two predicted values appear in Table 1. The first set of predicted values was calculated by the formula used by researchers at the Haskins Laboratories (Liberman et al, 1957; Pisoni, 1971); the second set of values stems from a formula of Fujisaki and Kawashima (1968, 1970; Pisoni, 1971).[1] As usual, the Haskins prediction underestimates discriminability, because it fails to account for contributions of information from echoic memory. The Fujisaki prediction utilizes the asymptotic, or "trough," value as an estimate of such auditory contribution and bases the estimates for the other points on this value along with identification scores. Not surprisingly, the Fujisaki prediction fits the data more closely than the Haskins prediction. Goodness-of-fit measures were calculated from individual obtained and Haskins-predicted scores (see Pisoni, 1971, p. 20). No S's obtained discrimination function differed significantly from his predicted function.[2]

**Speech Syllables.** The identification and discrimination functions for all of the speech stimuli appear in the lower panel of Fig. 2. Stimuli with rise times of 0, 10, 20, and 30 msec were identified as beginning with [tʃ] on 88% of all trials; stimuli with 60-, 70-, and 80-msec rise times were identified as beginning with [ʃ] on 80% of all trials; the 40- and 50-msec stimuli were ambiguous. The discrimination function peaks at the rise-time comparisons of 30-50 and 40-60 msec. Each point is significantly greater than its adjacent within-category comparison, 20-40 ($z = 2.36$, $p < .01$) and 50-70 ($z = 2.56$, $p < .01$). There were no significant differences among the other comparisons or, of course, between the [tʃa]-[ʃa] and the [tʃae]-[ʃae] arrays.

Table 1 shows the obtained, Haskins-predicted, and Fujisaki-predicted discrimination values for the speech stimuli. The pattern here follows that for the sawtooth stimuli. Again, no S's obtained discrimination function differed significantly from his function predicted by the Haskins formula.

Previous accounts of categorical perception in speech have dealt almost exclusively with the perception of stop consonants (Liberman et al, 1967; Studdert-Kennedy et al, 1970; Pisoni, 1971, 1973). The results of the present study indicate that categorical perception is nearly as marked for the affricate/fricative distinction as it is for distinctions among stops.

**A Comparison of Speech and Music.** The results of the two classes of stimuli are remarkably parallel: speech and musical sounds both yielded reasonably quantal identification functions together with discrimination functions whose peaks and troughs correspond to those boundaries and categories. The reason that the results are not even more impressive probably reflects two causes: stimuli are played over a loudspeaker rather than through earphones, and the

**Table 1**
**Obtained and Predicted Discrimination Values for Sawtooth and Speech Stimuli in Terms of Percentage Correct**

| Stimuli | Comparison (msec) | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0-20 | 10-30 | 20-40 | 30-50 | 40-60 | 50-70 | 60-80 |
| Sawtooth Waves | | | | | | | |
| Obtained | 61 | 64 | 72 | 78 | 58 | 60 | 59 |
| Haskins Predicted | 54 | 58 | 70 | 78 | 64 | 58 | 56 |
| Obtained Minus Predicted | 7 | 6 | 2 | 0 | −6 | 2 | 3 |
| Fujisaki Predicted | 61 | 63 | 73 | 80 | 67 | 62 | 60 |
| Obtained Minus Predicted | 0 | 1 | −1 | −2 | −9 | −2 | −1 |
| Speech Stimuli | | | | | | | |
| Obtained | 61 | 58 | 59 | 70 | 76 | 61 | 58 |
| Haskins Predicted | 54 | 55 | 60 | 70 | 74 | 64 | 55 |
| Obtained Minus Predicted | 7 | 3 | −1 | 0 | 2 | −3 | 3 |
| Fujisaki Predicted | 60 | 61 | 63 | 71 | 76 | 65 | 59 |
| Obtained Minus Predicted | 1 | −3 | −4 | 1 | 0 | −4 | −1 |

points displayed in Fig. 2 are group data. Individual data tend to show higher discrimination peaks and sharper category boundaries; averaging softens these extremes. For a comparison with other speech results in identification and discrimination, see Pisoni (1971).

Category boundaries for the two classes of stimuli were at somewhat different points, between 30 and 40 msec for the music stimuli and between 40 and 50 msec for the speech stimuli, but this does not impair the overall similarity of the results. Their similarity, in fact, is considerably greater than might have been expected. The stimulus classes differ radically in the aspect of the signal that has been varied: the nonspeech stimuli are periodic throughout their duration, whereas the speech stimuli are aperiodic during the portion of the stimulus that contains the variation. Furthermore, rise time is essentially the only variable that separates the musical items. In the ABX task, stimuli within a triad differed in duration by only 20 msec. This is well below the difference limen for sounds of approximately 1 sec duration (Fraisse, 1963). In contrast, rise time *and* duration are cues for the affricate/fricative segment of the speech syllables. The durational differences for our stimuli equal or exceed the difference limen. Gerstman (1957) noted that either rise time or duration can cue this distinction, but neither cue by itself is as potent as the two together. Thus, the choice of stimuli in the present experiment should have differentially favored categorical perception of the speech stimuli, since two important acoustic cues were varied instead of one. The results, however, do not reflect this advantage.

We must now consider why plucked and bowed notes demonstrated a form of categorical perception comparable to that found for affricates and fricatives. Consider first the labelling hypothesis advanced by Lane (1965): the nonspeech discrimination functions shown in Fig. 2 may be attributable to the verbal labels *pluck* and *bow*. Since the identification task preceded the discrimination task, Ss practiced as much as 12 min at labelling the sounds. Perhaps the effects of this practice carried over into the ABX task, and discriminations were mediated by labels. An obvious way to eliminate this flaw is to reverse the order of the tasks and have Ss perform the ABX task before the identifications. This was done in Experiment II.

Categories and boundaries may occur for sawtooth stimuli because their waveforms have complex spectra or because they sound similar to modes of playing stringed instruments that occur in our environment. Categorical perception may not occur for simpler sounds. In other words, does this boundary between pluck and bow occur for all sounds which vary in rise time, or just for certain sounds? Experiment II also addressed this question.

## EXPERIMENT II

### Method

The sawtooth identification and discrimination tapes used in Experiment I were employed here as well. Another set of 18 stimuli was generated; this consisted of sine waves varied in exactly the same manner as the sawtooth stimuli (at both 294 and 440 Hz). They were arranged in the same random orders as the sawtooth series for the two tasks, and corresponding identification and discrimination tapes were recorded. Sine-wave stimuli with 10- and 70-msec rise times are shown in Fig. 1. In an effort to sharpen up the identification and discrimination functions, signals were presented through matched Telephonics earphones (Model TDH-30). Twelve Yale University undergraduates with no special musical training served as paid volunteers who listened to both discrimination and both identification tapes. ABX tasks preceded identification tasks. Nested within tasks, half the Ss listened first to the sawtooth stimuli and then to the sine waves, while the others listened in reverse order. The labels, "pluck" and "bow," were never mentioned until after the two discrimination tasks were completed. Otherwise, instructions were identical to those of Experiment I.

### Results and Discussion

**Sawtooth Waves.** The top panel of Fig. 3 shows that the identification function of the tawtooth stimuli

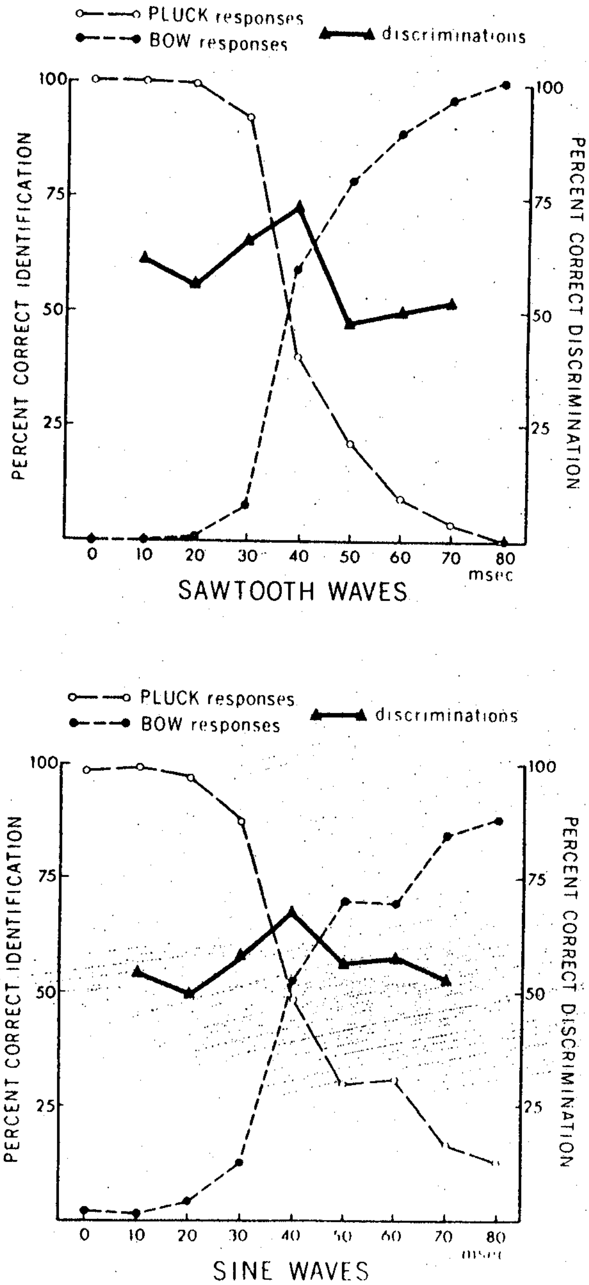


Fig. 3. Mean identification and discrimination functions for sawtooth and sine waves.

is very quantal and the discrimination function quite peaked. Reversing the task order obviously had little, if any, effect; playing the stimuli over earphones did sharpen up both functions. The 0-, 10-, 20-, and 30-msec stimuli were identified as plucked on 98% of all trials; the 50-, 60-, 70-, and 80-msec stimuli were identified as bowed on 92% of all trials; the 40-msec stimulus was ambiguous. The discriminability peak at the 30-50 comparison is significantly different from that at the 40-60 comparison ($z = 2.74$, $p < .005$); while not significantly different from the 20-40 comparison, it is significantly different from 10-30 ($z = 2.16$, $p < .05$). Obtained and Haskins-predicted scores are compared in Table 2, which reveals a pattern like that in Table 1 for the same stimuli. No S's obtained function differed significantly from his predicted scores. No Fujisaki predictions were made, since no stable estimate of asymptote (trough) is apparent in the data.

**Sine Waves.** In the lower panel of Fig. 3 are the identification and discrimination functions for the sine-wave stimuli. The identification function differs considerably from that of the sawtooth waves. While the 0-, 10-, 20-, and 30-msec stimuli were identified as plucked on 94% of all trials, the 50-, 60-, 70-, and 80-msec stimuli were identified as bowed on only 77% of all trials. The discrimination function, however, is quite peaked. The 30-50 comparison is significantly different from the same two comparisons as the sawtooth waves, 40-60 ($z = 2.16$, $p < .05$) and 10-30 ($z = 2.45$, $p < .01$). Obtained and Haskins-predicted scores for the sine waves appear in Table 2. They resemble those for the sawtooth stimuli. Again, no S's obtained and predicted functions significantly differed from one another.

**A Comparison of the Musical Stimuli.** The discrimination results show that categorical perception clearly occurs for both types of stimuli. Moreover, the category boundary as determined by the peak in the discrimination function is in the same place—about 40 msec. Although performance levels were slightly lower for the sine waves on the ABX task, there was no statistically significant difference between the nonspeech sounds.

One could hardly ask for a more striking identification function from the sawtooth stimuli. The sine wave function, however, is less quantal, particularly at longer rise times. The identification curves for the sine waves do not significantly differ from those for the sawtooth stimuli at rise times of 0 through 50 msec. Beyond 50 msec, however, the results for the two waveforms diverge. Indeed, "bow" is not very impressive as a category for sine waves. Sine waves of long rise time sound more like a flute played legato style than like a stringed instrument. Thus, the category "bowed" may be inappropriate for sine waves. A better opposition might have been "staccato" for relatively short rise times vs "legato" for longer ones. The role of verbal labels in revealing categories may need careful attention.

## Categories and Boundaries in Audition

Categorical perception presents sharp contrast to our more usual continuous perception. Most simple, psychophysical continua are perceived "continuously": one can discriminate many more items than one can identify (Pollack, 1952; Miller, 1956). In categorical perception, on the other hand, one may be

Table 2
Obtained and Predicted Discrimination Values for Sawtooth and Sine Wave Stimuli in Terms of Percentage Correct

| Stimuli | Comparison (msec) | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0-20 | 10-30 | 20-40 | 30-50 | 40-60 | 50-70 | 60-80 |
| Sawtooth Waves | | | | | | | |
| Obtained | 61 | 55 | 66 | 72 | 47 | 50 | 53 |
| Haskins Predicted | 51 | 53 | 68 | 79 | 63 | 59 | 53 |
| Obtained Minus Predicted | 10 | 2 | −2 | −7 | −16 | −9 | 0 |
| Sine Waves | | | | | | | |
| Obtained | 54 | 49 | 56 | 68 | 56 | 58 | 53 |
| Haskins Predicted | 51 | 55 | 67 | 77 | 59 | 58 | 59 |
| Obtained Minus Predicted | 3 | −6 | −11 | −9 | −3 | 0 | −6 |

able to discriminate only as well as one can identify the different items of the continuum. This unusual situation has been thought to be a unique characteristic of speech perception (Liberman et al, 1967). We have found it to occur for nonspeech sounds, obviously demonstrating that it is not unique to speech. Thus, adjustments must be made in the general account of perceptual categories in audition. Nevertheless, our finding is not a disproof of the importance of categorical perception in speech. Instead, speech may contain the most prominent examples of perceptual categories, even though such categories also occur elsewhere in nonspeech.

To develop this perspective, it is necessary to consider the nature of the ABX discrimination task. It reveals categories and boundaries, in part, because the ABX task forces the listener into echoic memory overload. Stimuli A and B must be remembered so that subsequent comparisons can be made with Stimulus X. Because echoic memory is simply insufficient to store all three stimuli, some form of coding must take place. Since Stimulus B arrives shortly after Stimulus A, A must be coded into some memorable form before B clobbers its echo; and B, in turn, must be coded before X clobbers it. By the time Stimulus X is in the system, Stimulus A may have been stripped of its acoustic form and Stimulus B may have been similarly coded. (However, unlike A, B may also retain an echo, as suggested by Fujisaki and Kawashima, 1968, 1970.) Crucial within-category information may have been lost during coding of A and B simply because such differences did not remain long enough in short-term memory. Thus, within-category ABX judgments are often reduced to guesswork. The quick loss of within-category information is the crux of categorical perception. Obviously, the ABX task is not the best possible discrimination task because of this memory and coding feature (Pisoni, 1971, 1973), but it may be most sensitive to perceptual categories in audition.

Coding processes do not always involve this quick loss of acoustic information. In fact, our experiments are the first to demonstrate such loss for common nonlinguistic sounds. For stop consonants, the acoustic nature of the speech sound is difficult to perceive: "the sound escapes us and we perceive the event, almost immediately, as phonetic [Studdert-Kennedy, in press]." It would seem that plucked and bowed notes must share this nearly instantaneous coding in view of the results they yield in the ABX task. However, a problem arises in comparing these musical sounds to speech: plucked-note and bowed-note encoding cannot be phonetic. This fact, coupled with the result of the first experiment, which demonstrated that rise time can cue perceptual categories in both speech and music, suggests that certain aspects of phonetic coding may be intimately related to the coding of naturally occurring nonlinguistic sounds.

Liberman, Mattingly, and Turvey (1972) have noted that the speech code is an efficient code. Transforming an echo of a speech sound into a phonetic representation of that sound is roughly equivalent to transforming a 40,000-bit/sec signal into a 40-bit/sec signal. This enormous saving entails the cost of losing such "unneeded" auditory information as within-category differences. The categorization process transforms massive amounts of information in the auditory signal into a "unitary neural event [p. 320]." A unitary neural representation would obviously be easier to store and to use for subsequent analyses and comparisons than would a degraded echo.

The problem then is to explain why it is advantageous to code (categorize) plucks and bows. It seems exactly backwards to suppose that phonetic processes were "tricked" into analyzing our musical sounds as if they were speech. An evolutionary view (Lieberman, 1973) is more reasonable. It assumes that speech perception developed around existing properties of the auditory system, one of which may be the detectability of different categories along the dimension of rise time. Many distinct categories exist in speech—the different consonant phonemes served as labels for each—and they are cued by acoustic variation in a number of dimensions, including rise time. Thus far, however, rise time is the only dimension found to cue perceptual categories in common nonlinguistic sounds according to the criteria suggested by Studdert-Kennedy et al (1970). Thus, categories and boundaries should be seen primarily as a well-developed characteristic of speech perception. Such nonlinearities of perception occur in nonspeech as well, but that is as it should be. The fabric of speech perception and the mechanisms behind it could not have been woven wholly out of new cloth. Remnants of underlying auditory, nonlinguistic processes should and do show through. The categorical perception of musical sounds varying in rise time is apparently one of these threads.

## REFERENCES

Cooper, F. S., & Mattingly, I. G. A computer-controlled PCM system for the investigation of dichotic speech perception. *Journal of the Acoustical Society of America*, 1969, **46**, 115(A).

Cutting, J. E. Different speech-processing mechanisms can be reflected in the results of discrimination and dichotic listening tasks. *Brain & Language*, in press.

Fraisse, P. *The psychology of time*. New York: Harper & Row, 1963.

Fujisaki, H., & Kawashima, T. The influence of various factors on the identification and discrimination of synthetic speech sounds. *Reports of the Sixth International Congress on Acoustics*, Tokyo, 1968.

Fujisaki, H., & Kawashima, T. Some experiments on speech perception and a model for the perceptual mechanisms. *Annual Report of the Engineering Research Institute*, Tokyo, 1970, *29*, 207-214.

Gerstman, L. J. Perceptual dimensions for the friction portion of certain speech sounds. Unpublished PhD dissertation, New York University (Psychology), 1957.

Lane H. Motor theory of speech perception: A critical review. *Psychological Review*, 1965, **72**, 275-309.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 431-461.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, **54**, 358-368.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. The discrimination of relative onset time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, 1961, **61**, 379-388.

Liberman, A. M., Mattingly, I. G., & Turvey, M. T. Language codes and memory codes. In A. W. Melton and E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C: Winston, 1972.

Lieberman, P. On the evolution of language: A unified view. *Cognition*, 1973, **2**, 59-94.

Lisker, L., & Abramson, A. S. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 1964, **20**, 384-422.

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

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. Discrimination in speech and nonspeech modes. *Cognitive Psychology*, 1971, **1**, 131-157.

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., Pastore, R. E., Wier, C. C., Kelly, W. M., & Dooling, T. M. Discrimination and labeling of noise-buzz sequences with varying noise-lead times. *Journal of the Acoustical Society of America*, 1974, **55**, 390(A).

Pisoni, D. B. On the nature of categorical perception of speech sounds. Unpublished PhD dissertation, University of Michigan (Psycholinguistics), 1971. Also published as a supplement to the Haskins Laboratories Status Report on Speech Research.

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

Pollack, I. The information of elementary multidimensional auditory displays. *Journal of the Acoustical Society of America*, 1952, **26**, 155-158.

Sawusch, J. R., & Pisoni, D. P. Category boundaries for speech and nonspeech sounds. *Journal of the Acoustical Society of America*, 1974, **55**, 536(A).

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Copper, F. S. Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 1970, **77**, 234-249.

Studdert-Kennedy, M. Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. Springfield, Ill: Thomas, in press.

## NOTES

1. Predicted functions were first obtained for each individual, then averaged to yield the values in the tables. Since there were so few observations per comparison per subject, the asymptotic value used in the Fujisaki and Kawashima predictions for each individual was, in fact, the mean asymptote for all Ss.

2. This result should not be taken as indicative of a "good fit" for predicted and obtained scores. Because of the small number of observations, an average difference between predicted and obtained scores of at least 34% at each comparison was necessary to reach significance.

(Received for publication May 16, 1974;
revision received June 25, 1974.)