

Sinusoidal plucks and bows are not categorically perceived, either

STUART ROSEN and PETER HOWELL
University College London, London, England

Cutting and Rosner (*Perception & Psychophysics*, 1974, 16, 564-570) reported that two acoustic nonspeech continua varying in rise time were categorically perceived. We have already shown (Rosen & Howell, *Perception & Psychophysics*, 1981, 30, 156-168) that the reason their sawtooth continuum was perceived in such a way, and in particular why it exhibited a midcontinuum peak in the discrimination function, was entirely due to the stimuli not having the intended rise times. The other nonspeech continuum that varied in rise time and was reported to be categorically perceived used a sinusoidal carrier. Although the labeling functions obtained were not as sharp as those obtained with sawtooth stimuli, the characteristic midcontinuum discrimination peak was found. We generated such a set of sinusoidal stimuli and found no evidence of categorical perception. Just as we have previously found for sawtooth stimuli, discrimination is best at the short rise-time end of the continuum and decreases monotonically with increasing rise time.

Early investigations of the perception of synthetic speech sounds uncovered a set of phenomena that were quite unexpected in view of what was then known about the perception of nonspeech sounds. Three surprising things happened when subjects were asked to label and discriminate sounds from some continua spanning a phonemic contrast, in particular those involving stop consonants. First, the categorization function was sharp, with few stimuli ambiguously labeled. Second, performance in the discrimination task was best for stimuli straddling the phoneme boundary as determined in the categorization experiment. Third, discrimination performance was near chance for stimuli drawn from the same phoneme class. A continuum exhibiting these properties is said to be categorically perceived (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970), since it seems that the ability to discriminate differences between sounds from such a continuum is determined completely by the ability to categorize them differently. At the time, this mode of perception was claimed to be specific to speech, and quite distinct from the continuous perception found with simple nonspeech sounds where listeners can discriminate among many more stimuli than can be accurately labeled (Miller, 1956).

Thinking about this issue was radically altered in the mid-70s by two demonstrations of the categorical

perception of nonspeech continua. Miller, Wier, Pastore, Kelly, and Dooling (1976) presented subjects with a continuum varying the relative onset time between a noise and a buzz (so-called noise onset time, or NOT) that could be considered analogous to a speech voice-onset-time (VOT) continuum. All three criteria for categorical perception were fulfilled. The relevance of these findings for speech perception is a bit more problematical, however, since the category boundary occurs at a different value on the continuum from those found in the analogous speech continua. It is not yet completely clear how to account for this discrepancy, although, at least partly, it may be due to differences in the dynamic aspects of spectral structure at the onset of the "voicing" component. More seriously, Summerfield (1982) claims that, while VOT continua show shifts in phoneme boundary with shifts in first formant frequency, NOT continua do not show similar shifts when their analogous "first formant" frequencies are likewise shifted.

Rather more impressive from the point of view of a correspondence between speech and nonspeech were the findings of Cutting and Rosner (1974). They claimed to demonstrate categorical perception for a continuum of sawtooth waveforms varying only in their rate of onset. Such an acoustic distinction can also signal the phonemic contrast between the voiceless affricate/fricative pair /tʃa/ and /ʃa/. Not only was this speech contrast also categorically perceived, but the category boundaries for the two continua were nearly identical at about 40 msec. This led Cutting and Rosner to theorize that at least some speech continua were categorically perceived because they possessed acoustic cues to which the auditory

This work was supported by the Medical Research Council of Great Britain. S. Rosen's mailing address is: Department of Phonetics and Linguistics, University College London, 4 Stephenson Way, London NW1 2HE, England. P. Howell's is: Department of Psychology, University College London, Gower Street, London WC1E 6BT, England.

system was specifically tuned. Thus, categorical perception was not special to speech. Rather, speech evolved to make use of these natural auditory sensitivities.

Regardless of the merits of this theory, the particular evidence cited by Cutting and Rosner (1974) has not held up in the light of more recent investigations. In Rosen and Howell (1981), we were unable to show categorical perception for a sawtooth continuum varying in rise time as specified by Cutting and Rosner. In particular, the discrimination function did not show a peak in performance at 40 msec, but was always best at the shortest rise times and decreased monotonically with increasing rise time. Furthermore, measurements of Cutting and Rosner's original stimuli showed them not to have the rise times they specified. The spacing between the stimuli was actually greatest at the supposed discrimination peak of 40 msec. By synthesizing a continuum of stimuli with the measured rise times of Cutting and Rosner's stimuli, we were able to replicate their finding of categorical perception. Thus, it is clear that the original claim that a sawtooth rise-time continuum is categorically perceived was solely due to the misconstruction of the stimuli.

Cutting (1982), Hary and Massaro (1982), and Kewley-Port and Pisoni (Note 1) have confirmed our findings. In addition, Cutting reported that a rise-time continuum with logarithmically spaced stimuli was categorically perceived. This strong effect of stimulus spacing has not been discussed in the categorical perception literature before and deserves further investigation. We do not discuss this matter here because, insofar as our interests are primarily in comparing results from similar acoustic contrasts in speech and nonspeech contexts (pluck/bow vs. affricate/fricative), the crucial factor is that the stimulus spacing be identical on all the continua compared.

Cutting and Rosner (1974) also claimed that another nonspeech continuum varying linearly in rise time was categorically perceived, one using sinusoidal waveforms instead of sawtooths. Although the identification function was not as sharp as that obtained with sawtooth waveforms, the discrimination function exhibited the characteristic midcontinuum peak. It seemed unlikely to us that this could be true given the results from the sawtooth continuum, so we attempted to replicate their results.

METHOD

Subjects

Five female undergraduates volunteered to serve as subjects. All were naive as to the purposes of the experiment and were paid for their participation.

Stimuli

A 300-Hz sinusoidal waveform of constant amplitude was generated by a DEC PDP-12 computer. Nine stimuli with linear rises varying in duration from 0 to 80 msec in 10-msec steps were created

by multiplying this sinusoid by the appropriate ramps. Each stimulus started at a negative-going zero crossing and had a linear decay over 1.02 sec from the finish of the rise.

Stimuli were output at run time at 10 kHz via a 10-bit D-A converter and low-pass filtered at 1 kHz (48 dB per octave). Because some other soft noises could be heard from the computer while the stimuli were being played out, a low-level masking noise was mixed with the output of the D-A converters. This noise was bandpass (20 Hz to 2 kHz, with 48 dB per octave roll-offs), and its RMS amplitude was 45 dB down from the peak level of the sinusoidal stimuli.

Finally, the output of the mixer was attenuated and fed to Sennheiser HD414X headphones. Stimuli were binaurally presented at a comfortable level, constant for all subjects.

Oscillographic traces of the voltage across the headphones showed both the rise and decay of the stimuli to be linear and within 1 msec of the desired value except for the 0-msec rise. In this stimulus, the shape of the rise is determined by the 300-Hz sinusoid, and so is not linear. In addition, the rise is complete only after about 1.2 msec, slightly longer than the duration of a quarter-cycle of the sinusoid, .833 msec. This is presumably due to the limited rate of rise possible through a 1-kHz low-pass filter.

These stimuli are nearly identical to those described in Experiment II of Cutting and Rosner (1974), differing only in frequency and in the detailed shape of the rises and falls. (Cutting and Rosner used two sets of sinusoidal stimuli, 294 and 440 Hz, and did not specify the shapes of the rises and falls). They are completely comparable to the stimuli used in Experiment 3 of Rosen and Howell (1981), except that the carrier is a single sinusoid of 300 Hz rather than the first 13 harmonics of a 300-Hz sawtooth waveform.

Procedure

The computer handled all details of stimulus presentation, response handling, and data analysis. Subjects sat in a quiet room and responded by pressing one of two buttons on a response box.

All subjects participated in a categorization experiment first. Each of the nine stimuli was presented 10 times, in a random order constrained such that all stimuli were presented before any were repeated. The subjects were asked to label the stimuli as "pluck" or "bow." At the beginning of the session, the subjects heard the 0- and 80-msec stimuli three times alternately as exemplars while the buttons on the response box, marked "pluck" and "bow," lit appropriately.

At the end of this task, subjects' responses to the two endpoint stimuli were inspected. If less than 95% of these two stimuli were correctly classified, the subjects ran another categorization session, undergoing the identical procedure. When responses were correct to 95% of the endpoint stimuli, the subjects moved on to the discrimination task. This was done to ensure that subjects could reliably categorize the continuum. Three of the subjects met this criterion in their first session, and the other two in their second.

The discrimination task was two-step ABX with approximately 2.2 sec between the onsets of the triplet. Two runs of 28 trials (7 comparisons \times 4 orders) in a random order, again subject to the condition that all be presented before any were repeated, gave eight observations per stimulus pair per subject.

RESULTS

Figure 1 shows the mean categorization and discrimination curves, averaged over the five subjects, as a function of rise time. Only the last session performed by each subject is included in the means of the categorization function. The labeling ability exhibited by these subjects in judging sinusoids is very similar to that found by Rosen and Howell (1981) in their Experiment 3 for naive subjects classifying

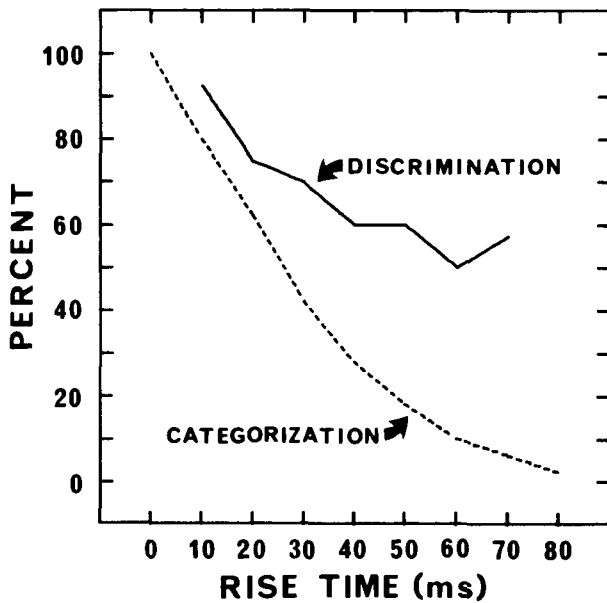


Figure 1. Mean categorization and discrimination functions from five naive subjects for a sinusoidal rise-time continuum. For the categorization curve, ordinate values give the percentage of "pluck" responses for each stimulus. For the discrimination curve, the values give percent correct. The discrimination was two-step ABX with the results for a particular stimulus pair plotted at the midpoint of the pair. Each point in the categorization curve represents 50 observations, while those in the discrimination curve represent 40.

equivalent sawtooth waveforms. Maximum-likelihood fitting of a cumulative normal function (Bock & Jones, 1968) shows no significant differences in either the slopes or the category boundaries.

More importantly, discrimination is best for the shortest rise times and decreases gradually as rises are lengthened. This is the same pattern that was found for sawtooth waveforms by Rosen and Howell. Results from individual subjects tell the same story. All subjects showed best discrimination for the stimulus pair with the shortest rise times.

It might be argued that the discrimination function is not peaked at the category boundary because the subjects are not categorizing the stimuli into clearly dichotomous classes. Figure 2 shows a pair of categorization and discrimination functions obtained from two rather more experienced subjects (the authors), who followed the same procedure as the other subjects. Both of them labeled the two extreme stimuli correctly 100% of the time, and so participated in only one session of categorization. Their mean labeling function is considerably sharper than that shown in Figure 1. Furthermore, maximum-likelihood fitting of cumulative normal curves to the individual results shows that both of the experienced subjects have a steeper categorization function than any of the naive subjects, so this difference in slopes

is not a result of data averaging. Even though the labeling function from the experienced subjects is considerably steeper than that produced by the naive subjects, the discrimination functions are quite similar, and in neither case is a midcontinuum peak found.

DISCUSSION

As suspected, no evidence was found for the categorical perception of sinusoidal stimuli varying in rise time. This is consistent not only with our previous results for sawtooth stimuli (Rosen & Howell, 1981), but also, more importantly (since the stimuli are more comparable), with the findings of van Heuven and van den Broecke (1979). They measured the variability of settings in a rise-time reproduction task using sinusoidal stimuli and showed that the standard deviation of the adjustments was a linear increasing function of rise time. Since standard deviations in an adjustment task are usually linearly related to difference limens in a two-interval two-alternative forced-choice (2I-2AFC) task (Cardozo, 1965; Wier, Jesteadt, & Green, 1976) and the transformation of percent correct in a 2I-2AFC task to percent correct in an ABX task is monotonically increasing (Macmillan, Kaplan, & Creelman, 1977), the constant difference between stimuli in the ABX task should lead to best performance for the shortest rise times, with performance decreasing monotonically with increasing rise time. The data were not definitive because the reproduction task is not reliable in the short rise time end of the continuum.

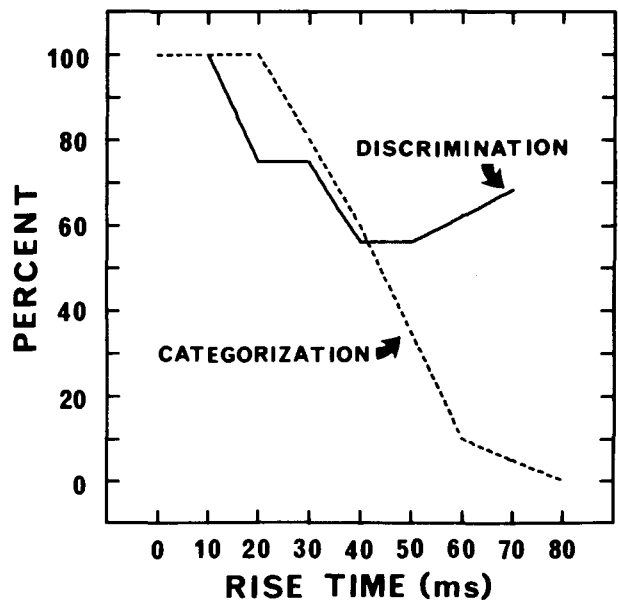


Figure 2. As for Figure 1, but from two experienced subjects (the authors). Each stimulus was categorized 20 times; each ABX pair was observed 16 times.

The finding that nonspeech rise-time continua are not categorically perceived might be thought to provide support for the notion that "speech is special," since the rise-time-cued speech continuum investigated by Cutting and Rosner (1974) was reported to be categorically perceived. In fact, this result has not held up either. Rosner (Note 2) constructed a voiceless affricate/fricative continuum by varying the rise time of the initial frication noise from 5 to 65 msec in 10-msec steps. Performance in a standard ABX discrimination task was best at the short-rise-time end of the continuum, decreasing with increasing rise time, as has been found for the nonspeech rise-time continua.

Furthermore, some acoustic measurements of our own (Howell & Rosen, 1983) show that natural voiceless affricates and fricatives do not have the rise times reported by Gerstman (1957) which form the basis for the range of rise times employed by Cutting and Rosner (1974) and Rosner (Note 2) in their synthetic continua. Measured rise times are considerably longer than those previously reported and vary systematically with the material spoken. For CVs in isolation, the mean rise times (averaged over four subjects and four vowels) were 54 msec for affricates and 120 msec for fricatives. Thus, the 40-msec "natural" boundary claimed by Cutting and Rosner, even if it did exist, would not serve to distinguish affricates from fricatives in real speech. In any case, we have shown for a voiceless affricate/fricative continuum covering a natural range of rise times (40-120 msec), as well as one covering the previously used unnatural range (0-80 msec), that discrimination performance is always best at the short-rise-time end of the continuum (Howell & Rosen, in press). Thus, rise time is not categorically perceived in speech or nonspeech.

It is not clear why Cutting and Rosner (1974) obtained the results they did with the sinusoidal stimuli. We can only surmise that their rise times, like those of the sawtooth stimuli, are different from those specified. The two continua of Cutting and Rosner were generated using a similar technique (Cutting, 1982, Note 3); a large number of stimuli were recorded from a Moog synthesizer and digitized, and those with appropriate rise times selected for the continuum. This method presumably failed for the sinusoidal sounds for the same reasons that it failed for the sawtooths.

REFERENCE NOTES

1. Kewley-Port, D., & Pisoni, D. B. *Identification and discrimination of rise time: Is it categorical or noncategorical?* (Research

on Speech Perception Progress Report No. 7). Bloomington: Indiana University, Department of Psychology, 1981.

2. Rosner, B. S. *Rise time and the perception of transient sounds*. Final Progress Report for Grant NS-13545 at the University of Pennsylvania, 1 December 1977 to 30 April 1981.

3. Cutting, J. E. Personal communication, October 1980.

REFERENCES

- BOCK, R. D., & JONES, L. V. *The measurement and prediction of judgment and choice*. San Francisco: Holden-Day, 1968.
- CARDOZO, B. L. Adjusting the method of adjustment: SD vs DL. *Journal of the Acoustical Society of America*, 1965, 37, 786-792.
- CUTTING, J. E. Plucks and bows are categorically perceived, sometimes. *Perception & Psychophysics*, 1982, 31, 462-476.
- CUTTING, J. E., & ROSNER, B. S. Categories and boundaries in speech and music. *Perception & Psychophysics*, 1974, 16, 564-570.
- GERSTMAN, L. J. *Perceptual dimensions for the friction portions of certain speech sounds*. Unpublished doctoral dissertation, New York University, 1957.
- HARY, J. M., & MASSARO, D. W. Categorical results do not imply categorical perception. *Perception & Psychophysics*, 1982, 32, 409-418.
- HOWELL, P., & ROSEN, S. Production and perception of rise time in the voiceless affricate/fricative distinction. *Journal of the Acoustical Society of America*, 1983, 73, 976-984.
- HOWELL, P., & ROSEN, S. Natural auditory sensitivities as universal determiners of phonemic contrasts. In B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Explanations of linguistic universals*. The Hague: Mouton, in press.
- MACMILLAN, N. A., KAPLAN, H. L., & CREELMAN, C. D. The psychophysics of categorical perception. *Psychological Review*, 1977, 84, 452-471.
- MILLER, G. A. The magical number seven plus or minus two, or, some limits on our capacity for processing information. *Psychological Review*, 1956, 63, 81-96.
- MILLER, J. D., WIER, C. C., PASTORE, R. E., KELLY, W. J., & DOOLING, R. J. Discrimination and labeling 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.
- ROSEN, S., & HOWELL, P. Plucks and bows are not categorically perceived. *Perception & Psychophysics*, 1981, 30, 156-168.
- STUDDERT-KENNEDY, M., LIBERMAN, A. M., HARRIS, K. S., & COOPER, F. S. Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 1970, 77, 234-249.
- SUMMERFIELD, A. Q. Differences between spectral dependencies in auditory and phonetic temporal processing: Relevance to the perception of voicing in initial stops. *Journal of the Acoustical Society of America*, 1982, 72, 51-61.
- VAN HEUVEN, V. J. J. P., & VAN DEN BROECKE, M. P. R. Auditory discrimination of rise and decay times in tone and noise bursts. *Journal of the Acoustical Society of America*, 1979, 66, 1308-1315.
- WIER, C. C., JESTEADT, W., & GREEN, D. M. A comparison of method-of-adjustment and forced-choice procedures in frequency discrimination. *Perception & Psychophysics*, 1976, 19, 75-79.

(Manuscript received January 24, 1983;
revision accepted for publication May 9, 1983.)