

Categorical perception of nonspeech sounds by 2-month-old infants

PETER W. JUSCZYK and BURTON S. ROSNER
University of Pennsylvania, Philadelphia, Pennsylvania 19104

JAMES E. CUTTING
Wesleyan University, Middletown, Connecticut 06457
and
Haskins Laboratories, New Haven, Connecticut 06510

and

CHRISTOPHER F. FOARD and LINDA B. SMITH
University of Pennsylvania, Philadelphia, Pennsylvania 19104

According to recent investigations, adult listeners perceive rise-time differences in both speech and nonspeech stimuli in a categorical manner (Cutting & Rosner, 1974). Adults labeled sawtooth-wave stimuli as either plucked or bowed. The present study uses the high-amplitude sucking technique to explore the 2-month-old infant's perception of rise-time differences for sawtooth stimuli. Infants discriminated rise-time differences which marked off the different nonspeech categories, but did not discriminate equal differences within either category. Thus, the present study shows that infants, like adults, can perceive nonspeech stimuli in a categorical manner.

Considerable evidence indicates that many speech sounds are perceived categorically. With these stimuli, subjects are only slightly better at discriminating sounds that they are at differentially labeling them. This claim is supported by experimental findings from a number of different paradigms including: (a) accuracy (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Mattingly, Liberman, Syrdal, & Halwes, 1971; Pisoni, 1971, 1973), (b) reaction time (Pisoni & Tash, 1974), and (c) average evoked potentials (Dorman, 1974). These results contrast with those observed for a wide variety of nonspeech sounds, varying along such physical continua as frequency, amplitude, and duration, for which the subject's ability to discriminate between stimuli far outstrips his ability to label them differentially (Miller, 1956).

There is also a growing body of evidence which shows that human infants are capable of discriminating speech segments on the basis of minimal phonetic cues. To date, infants have displayed an ability to perceive subtle differences in voicing (Eimas, 1975b;

Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Lasky, Syrdal-Lasky, & Klein, 1975; Streeter, 1976), place of articulation (Eimas, 1974; Morse, 1972), initial burst cues (Miller, Morse, & Dorman, Reference Note 1), and third formant cues for the /ra/—/la/ distinction (Eimas, 1975a). Not only do infants make fine distinctions between speech sounds, but they do so in a categorical manner (i.e., they make inter-phonemic distinctions but not intraphonemic ones). Further, Eimas (1974, 1975b) has shown that infants, like adults (Mattingly et al., 1971), perceive certain acoustic cues categorically in speech contexts but *not* in nonspeech contexts. On the basis of these findings, Eimas (1975b and elsewhere) has suggested that the actual mechanisms which underlie the categorical perception of speech may be part of the biological makeup of the human infant.

Thus, speech appears to be perceived in a quite different fashion from nonlinguistic auditory stimuli. However, several recent developments may require us to reexamine any claim about the distinctive nature of speech perception. Categorical perception has now been observed in a number of instances of nonspeech sounds (Cutting & Rosner, 1974; Cutting, Rosner, & Foard, in press; Locke & Kellar, 1973; Miller, Wier, Pastore, Kelly, & Dooling, 1976). In particular, Cutting and Rosner (1974) have reported categorical perception for nonspeech sounds varying in rise time. They have explored the perception of rise-times in both sawtooth-wave and

This research was supported by NICHD Grants 5T01 HD0037 to the Department of Psychology, University of Pennsylvania, under which the first and fifth authors served as trainees, and Grants HD-01994 and RR-05596 to the Haskins Laboratories. P. W. Jusczyk is now at the Department of Psychology, Dalhousie University, Halifax, Nova Scotia; reprint requests should be sent to him there.

sine-wave stimuli (as well as for affricate-fricatives in speech). Adult listeners usually reported that these nonspeech stimuli sound as though they were produced by a musical stringed instrument. Sounds with rapid rise times (less than 40 msec) were perceived as emanating from a plucked string, whereas sounds with more gradual rise times (greater than 40 msec) were perceived as being produced by a bowed string. The listeners easily identified the stimuli as either "pluck" or "bow." Moreover, the perception of these stimuli was categorical.

In a related study, Cutting, Rosner, and Foard (1976) extended the findings for the sawtooth-wave stimuli by demonstrating selective adaptation effects with them. These effects were similar to those observed with speech stimuli (Eimas & Corbit, 1973) both in direction and degree of shift. Moreover, as in the case of speech stimuli, adaptation shifts for the sawtooth stimuli were greatest when the adapting stimulus shared all dimensions with the test continuum. Finally, a recent report of Blechner (Reference Note 2) indicates that the right-ear advantage often observed for dichotically presented speech sounds (e.g., Shankweiler & Studdert-Kennedy, 1967) also occurs with the sawtooth-wave stimuli.

Although the claim for the distinctive nature of speech perception has been weakened by these lines of research, there has been no indication that infants might exhibit categorical perception for nonspeech sounds. In fact, Eimas (1974, 1975b) has reported that 2- and 3-month-old infants tend *not* to perceive nonspeech cues categorically. However, the cues which Eimas studied were acoustic features which adults do not perceive categorically (Mattingly et al., 1971; Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975). The sawtooth-wave stimuli employed by Cutting and Rosner (1974) would seem to be a better choice for such a test. Not only do adults perceive these sounds categorically, but rise time is also an important acoustic cue in various contexts. Accordingly, the present study explored the perception of rise-time differences in sawtooth-wave stimuli by 2-month-old infants.

METHOD

Procedure

Each infant was tested in a mobile laboratory. The infants were placed in a reclining seat which faced a loudspeaker approximately 2 ft away. Each subject sucked on a blind nipple which one of the experimenters held in place.¹

The experimental procedure was a modification of the high-amplitude sucking technique developed by Sigel and De Lucia (1969). For each infant, the high-amplitude sucking criterion and the baseline rate of high-amplitude nonnutritive sucking were established before presentation of any stimuli. The criterion for high-amplitude sucking was adjusted to produce sucking rates of 10 to 20 sucks/min. After a baseline rate was established, the presentation of stimuli was made contingent upon the rate of sucking. If the time between

each criterion response was 2 sec or more, then each response produced one presentation of the stimulus, which had an average duration of 1,050 msec, followed by 950 msec of silence. If the infant produced a burst of high-amplitude responses within this two second interval, timing apparatus was automatically reset and the 2-sec interval began again.

The criterion for satiation to the first stimulus was a decrement in sucking rate of 25% or more over 2 consecutive minutes compared to the rate in the immediately preceding minute. At this point, the auditory stimulation was changed without interruption by switching channels on the tape recorder. For infants in the experimental conditions, the change resulted in the presentation of a second *acoustically* distinct stimulus. For the infants in the control condition, the channels on the tape recorder were switched but no acoustic change was made. The postshift period lasted for 4 min. The infants' sensitivity to the change in the auditory stimulation was inferred from comparisons of the response rates of subjects in the experimental and control conditions during the postshift period.

Stimuli

The stimuli were sawtooth waves generated on the Moog synthesizer at the Presser Electronic Studio of the University of Pennsylvania. The four stimuli were synthesized at 440 Hz and differed solely in their onset characteristics. The amplitude envelope reached maximum in 0, 30, 60, or 90 msec after onset. By 0 msec rise time, we mean that a stimulus reached maximum amplitude in one-fourth of a period. Previous research by Cutting and Rosner (1974) has indicated that adults easily label the rapid onset (0 and 30 msec) sounds as "plucks." The more gradual onset stimuli (60 and 90 msec) were easily labeled as "bows." The durations of the four nonspeech stimuli were 1,020, 1,050, 1,080, and 1,110 msec, varying according to rise time. The decay period of each stimulus was 1,020 msec.

All the stimuli were prerecorded on three 30-min audio tapes for presentation to the subjects. Tape 1 (pluck-pluck) was composed of 0-msec rise-time stimuli on channel A and of 30-msec rise-time stimuli on channel B. Tape 2 (bow-bow) was composed of 60-msec rise-time stimuli on channel A and of 90-msec rise-time stimuli on channel B. Tape 3 (pluck-bow) was composed of 30-msec stimuli on channel A and of 60-msec stimuli on channel B.

Design

Table 1 shows the within-subjects design for the present experiment. All subjects were seen for two experimental sessions. (Mean interval between sessions was 8 days; range was 5 to 14 days.) In one session, all subjects heard the pluck-bow tape. The other session differentiates the three groups of subjects. Subjects in Group 1 heard the pluck-pluck tape. Subjects in Group 2 heard the bow-bow tape. Subjects in Group 3 were randomly assigned one of the four rise-time stimuli for the entire session (the no-change condition). The order of sessions and the order of stimuli within a session were each counterbalanced.

Apparatus

A blind nipple was connected to a Grass PT5 volumetric pressure transducer, which was coupled, in turn, to a Beckman Type RS Dynograph. An integrator coupler provided a digital output of critical high-amplitude sucking responses. Additional equipment included a 4-track Hitachi tape recorder with speakers, a Hunter digital timer, two relays, and a counter. Each criterion

Table 1
Design

	Session A	Session B
Group 1	Pluck-bow	Pluck-pluck
Group 2	Pluck-bow	Bow-bow
Group 3	Pluck-bow	NO CHANGE

Note— $N = 6$ in each group.

response activated the digital timer for a 2-sec period or restarted the period. Auditory stimulation at a level of 72 ± 2 dB SPL was available to the infant whenever the timer was in an "active" state.

Subjects

The subjects were 18 infants, 9 males and 9 females. Mean age was 8 weeks (range: 5 to 10 weeks). In order to obtain complete data on 18 infants, it was necessary to test 25. Seven infants were dropped from the study for the following reasons: two infants fell asleep prior to shift, three cried excessively prior to shift, and the mothers of two infants were unable to keep the second appointment.

RESULTS

Figure 1 displays the mean number of high-amplitude sucking responses as a function of minutes and experimental groups. For purposes of statistical comparisons, we examined each subject's rate of high-amplitude sucking during five intervals: baseline minute, 3rd minute before shift, average of Minutes 1 and 2 before shift, average of Minutes 1 and 2 after shift, and average of Minutes 3 and 4 after shift. Difference scores were calculated for each subject for the following rate comparisons: (1) acquisition of the sucking response, 3rd minute before shift less baseline; (2) satiation, third minute before shift less average of last 2 min before shift; (3) release from satiation, average of first 2 min after shift less average of first 2 min before shift; (4) late release from satiation, average of 3rd and 4th minutes after shift less average of last 2 min before shift; and (5) satiation to the second stimulus, average of first 2 min after shift less average of 3rd and 4th minutes after shift.

Kruskal-Wallis one-way analyses of variance (Siegel, 1956) were employed to determine if the data for the pluck-bow sessions could be collapsed across the three experimental groups. No significant differences were observed between groups for any of the five comparisons [$\chi^2(2)$ ranged from 0.37 to 4.10]; accordingly, the data for the pluck-bow sessions were collapsed across groups in further analyses.

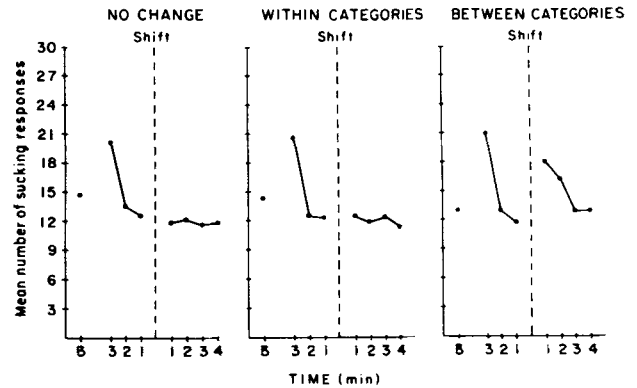


Figure 1. Mean sucking rates as a function of time and experimental session. Time is measured with reference to the moment of the stimulus shift, marked by the vertical dashed line. The baseline rate of sucking is indicated by the letter "B."

Additionally, Kruskal-Wallis tests indicated no differences [$\chi^2(1)$ ranged from 0.03 to 0.92] for the bow-bow and pluck-pluck subgroups, whose data were similarly combined for further treatment.

Wilcoxon matched-pairs signed-ranks tests (Siegel, 1956) were used to analyze performance within each type of session. The results of these analyses, presented in Table 2, indicated that in all sessions subjects acquired the conditioned high-amplitude sucking response and satiated to the first stimulus prior to shift. However, only in the pluck-bow condition did subjects display a reliable increase in sucking after the shift. Moreover, these subjects showed a reliable increase in sucking during the first 2 min after shift followed by a reliable decrease in rate between that period and the next two minutes, thus indicating satiation to the second stimulus. By contrast, subjects in the other three conditions showed no evidence of any increase in sucking after shift. Subsequent analysis of the data for the pluck-pluck, bow-bow, and no-change sessions by Kruskal-Wallis tests indicated no reliable differences in the pattern of responding by subjects in these sessions.

Table 2
t Values for Wilcoxon Matched-Pairs Signed-Ranks Test

Comparison	Experimental Session		
	Pluck-bow (n = 18)	Pluck-pluck or bow-bow (n = 12)	NO CHANGE (n = 6)
Acquisition			
Third minute before shift vs baseline.	0**	0**	0*
Satiation			
Third minute before shift vs average of last 2 min before shift.	0**	0**	0*
Release from Satiation			
First 2 min after shift vs last 2 min before shift.	-1**	-19	0*†
Late Release from Satiation			
Third and 4th min after shift vs last 2 min before shift.	-52.5	23	4
Satiation to Second Stimulus			
First 2 min after shift vs 3rd and 4th min after shift.	-12**	-26	8

** $p < .01$

* $p < .05$

*†Indicates reliable decrease in sucking

Randomization tests on within-subjects data across conditions confirmed these findings.

Butterfield and Cairns (1974) have reported that asymmetrical order effects are sometimes observed for speech stimuli which cross phonetic boundaries (a shift from a voiced to a voiceless stop producing greater dishabituation than from voiceless to voiced). We tested for such asymmetries with the present stimuli. None were discovered, as Kruskal-Wallis tests for the pluck-bow sessions yielded no reliable differences [$\chi^2(1)$ ranged from 0.02 to 1.73] between the two presentation orders.

DISCUSSION

The present data indicate that infants as young as 2 months of age perceive rise-time cues in sawtooth-wave stimuli in a categorical manner, as do adults (Cutting & Rosner, 1974).² This constitutes the first demonstration that infants perceive *acoustic* stimuli other than speech in a categorical fashion (see Bornstein, Kessen, & Weiskopf, 1976, for an indication of categorical perception of hues by infants). Our results are consistent with those observed for speech stimuli (e.g., Eimas, 1974, 1975a,b; Eimas et al., 1971), since infants displayed a reliable increase in sucking only for stimuli chosen from opposite sides of the adult categorical boundary.

How can we explain the 2-month-old's propensity to categorize "plucks" and "bows"? One relevant result (Cutting & Rosner, 1974) is that rise-time is a sufficient cue for the categorical perception of [ʃa] and [tʃa] as in "shop" and "chop." One possible explanation for the present results, then, is that the sawtooth-wave stimuli are perceived categorically just because rise-time is a salient dimension in speech perception. By one interpretation of this linguistic hypothesis, however, every acoustic dimension which is perceived categorically in speech should also be perceived categorically in nonspeech sounds. Yet, Mattingly et al. (1971) reported that second formant transitions which are perceived categorically in speech are not perceived categorically when heard in isolation. These results undercut the strong version of the linguistic hypothesis. An alternative formulation would hold that all dimensions perceived categorically in nonspeech sounds also are perceived categorically in speech sounds. Locke and Kellar's (1973) report of categorical perception of triadic chords seems to contradict this view. Acceptance of this weak version of the hypothesis also leaves open the question of why some dimensions, but not others, are perceived categorically outside of speech.

A second hypothesis can account for our results. This acoustic hypothesis argues that the categorical perception of [ʃa] and [tʃa] is merely a special case of the categorical perception of the acoustic dimension

of risetime. Indeed, many other nonspeech stimuli may also be perceived categorically. According to this view, the categorical perception of speech sounds is a consequence of general properties of the auditory system rather than of a special system devoted entirely to the perception of speech. This is supported by a number of recent findings. For example, Lisker (1975) and Stevens and Klatt (1974) have demonstrated that voice onset time (VOT) is actually composed of several *acoustic* cues. Selective adaptation with the individual acoustic cues from these dimensions produced boundary shifts along the VOT continuum (Cooper, 1974). Similarly, Tartter and Eimas (1975) demonstrated that a number of acoustic cues produced selective adaptation effects for the place-of-articulation continuum as well as for the VOT continuum. Their investigations led Tartter and Eimas to conclude that some selective adaptation effects previously thought explicable only by a phonetic model (e.g., Eimas, Cooper, & Corbit, 1973) can be more simply handled by reference to acoustic features. Thus, these recent studies tend to show that more and more of the presumably unique features of human speech can be explained in terms of the acoustic properties of the sounds. Perhaps the particular combination of information available for auditory analysis determines the activation of higher level analyzers which possibly deal only with phonetic information. Thus, the human's tendency for categorical perception is not limited to speech sounds. The number and variety of nonspeech sounds which are perceived categorically remains to be determined.

REFERENCE NOTES

1. Miller, C. L., Morse, P. A., & Dorman, M. F. *Infant speech perception, memory, and the cardiac orienting response*. Paper presented at the meetings of the Society for Research in Child Development, Denver, Colorado, April 1975.
2. Blechner, M. J. *Right-ear advantage for musical stimuli differing in rise-time*. Manuscript submitted for publication, Haskins Laboratories, New Haven, Connecticut.

REFERENCES

- BORNSTEIN, M. H., KESSEN, W., & WEISKOPF, S. The categories of hue in infancy. *Science*, 1976, **191**, 201-202.
- BUTTERFIELD, E. C., & CAIRNS, G. F. Whether infants perceive linguistically is uncertain, and if they did, its practical importance would be equivocal. In R. L. Schiefelbusch & L. L. Lloyd (Eds.), *Language perspectives: Acquisition, retardation and intervention*. Baltimore: University Park Press, 1974.
- COOPER, W. E. Selective adaptation for acoustic cues of voicing in initial stops. *Journal of Phonetics*, 1974, **2**, 303-313.
- CUTTING, J. E., & ROSNER, B. S. Categories and boundaries in speech and music. *Perception & Psychophysics*, 1974, **16**, 564-571.
- CUTTING, J. E., ROSNER, B. S., & FOARD, C. Perceptual categories for musiclike sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology*, 1976, **28**, 361-378.

- DORMAN, M. F. Auditory evoked potential correlates of speech sound discrimination. *Perception & Psychophysics*, 1974, 15, 215-220.
- EIMAS, P. D. Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, 1974, 16, 513-521.
- EIMAS, P. D. Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics*, 1975, 18, 341-347. (a)
- EIMAS, P. D. Speech perception in early infancy. In L. B. Cohen & P. Salapatek (Eds.), *Infant perception*. New York: Academic Press, 1975. (b)
- EIMAS, P. D., COOPER, W. E., & CORBIT, J. D. Some properties of linguistic feature detectors. *Perception & Psychophysics*, 1973, 13, 247-252.
- EIMAS, P. D., & CORBIT, J. D. Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 1973, 4, 99-109.
- EIMAS, P. D., SIQUELAND, E. R., JUSCZYK, P., & VIGORITO, J. Speech perception in infants. *Science*, 1971, 171, 303-306.
- LASKY, R. E., SYRDAL-LASKY, A., & KLEIN, R. E. VOT discrimination by four and six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, 1975, 20, 215-225.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, 74, 431-461.
- LISKER, L. Is it VOT or a first-formant transition detector. *Journal of the Acoustical Society of America*, 1975, 57, 1547-1551.
- 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, 2, 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., WIER, C. C., PASTORE, R. E., KELLY, W. M., & DOOLING, R. M. Discrimination and labelling 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.
- MIYAWAKI, K., STRANGE, W., VERBRUGGE, R., LIBERMAN, A. M., JENKINS, J. J., & FUJIMURA, O. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, 1975, 18, 331-340.
- MORSE, P. A. The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology*, 1972, 14, 477-492.
- PISONI, D. B. On the nature of categorical perception of speech sounds. (Unpublished PhD dissertation, University of Michigan, 1971) *Dissertation Abstracts International*, 1972, 32, 6693B (University Microfilms No. 72-14,964).
- PISONI, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 1973, 13, 253-260.
- PISONI, D. B., & TASH, J. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 1974, 15, 285-290.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- SIQUELAND, E. R., & DELUCIA, C. A. Visual reinforcement of non-nutritive sucking in human infants. *Science*, 1969, 165, 1144-1146.
- STEVENS, K. N., & KLATT, D. H. Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*, 1974, 55, 653-659.
- STREETER, L. A. Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature*, 1976, 259, 39-41.
- SHANKWEILER, D., & STUDDERT-KENNEDY, M. Identification of consonants and vowels presented to left and right ears. *Quarterly Journal of Experimental Psychology*, 1967, 19, 59-63.
- TARTTER, V. C., & EIMAS, P. D. The role of auditory feature detectors in the perception of speech. *Perception & Psychophysics*, 1975, 18, 293-298.

NOTES

1. Unfortunately, we could not outfit our testing van with the equipment necessary to play earphone-delivered masking noise to the experimenter holding the nipple. However, it should be noted that a number of other infant studies, including Eimas et al. (1971), did not employ this masking procedure either.

2. It has been called to our attention that although each of our stimulus pairs for the within- and between-category sessions had the same rise-time differences (30 msec), the percentage of change was not equivalent for each pair. Thus, the percentage of change for the 60- vs 90-msec pair was less than the percentage of change for the 30 vs 60-msec pair. However, if infants were only discriminating a larger percentage change in the 30- vs 60-msec group, then one would also expect to find that they discriminated the 0- vs 30-msec pair (since the percentage of change is even larger for this pair). In fact, only the 30- to 60-msec change was detected.

Still, one might argue that the 0- to 30-msec change occurs below some threshold that exists for detecting differences in rise times. Were this the case, then one might expect to find categorical perception of "plucked" sounds but continuous perception of "bowed" sounds. Informal listening tasks show that it would still be very difficult to discriminate between rise times of 60 vs 120 msec (the same percentage change as a 30- vs 60-msec pair). Therefore, any difference thresholds, if they exist, within the category "bow" would be quite large.

(Received for publication May 12, 1976;
revision accepted October 27, 1976.)