

## Auditory and linguistic processing of cues for place of articulation by infants\*

PETER D. EIMAS

*W. S. Hunter Laboratory of Psychology, Brown University, Providence, Rhode Island 02912*

Two- and 3-month-old infants were found to discriminate the acoustic cues for the phonetic feature of place of articulation in a categorical manner; that is, evidence for the discriminability of two synthetic speech patterns was present only when the stimuli signaled a change in the phonetic feature of place. No evidence of discriminability was found when two stimuli, separated by the same acoustic difference, signaled acoustic variations of the same phonetic feature. Discrimination of the same acoustic cues in a nonspeech context was found, in contrast, to be noncategorical or continuous. The results were discussed in terms of infants' ability to process acoustic events in either an auditory or a linguistic mode.

A major conclusion of the research on the perception of speech over the past two decades is that human listeners perceive speech quite differently from the way in which they perceive nonspeech sounds. Speech perception is said to occur in a linguistic or speech mode as opposed to an auditory mode (Eimas, in press; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman, 1970; Studdert-Kennedy, 1974). Evidence for this conclusion comes from studies in which listeners identified and discriminated series of synthetic speech sounds that varied continuously along a single acoustic dimension. Listeners were typically quite consistent in assigning phonetic labels to the various stimuli. Moreover, and most importantly, the ability to discriminate pairs of stimuli was strongly determined by the phonetic assignments. Thus, two stimuli that were acoustic variations of the same phonetic category or feature were discriminated only slightly better than would be expected by chance, whereas two stimuli that were separated by the same acoustic difference but members of different phonetic categories were highly discriminable. This form of perception has been termed categorical and is particularly apparent for those acoustic dimensions that distinguish the stop consonants (Liberman, 1957; Liberman, Harris, Hoffman, & Griffith, 1957; Eimas, 1963; Mattingly, Liberman, Syrdal, & Halwes, 1971; Lisker & Abramson, 1970; Abramson & Lisker, 1970; Pisoni, 1973). Categorical-like perception has also been found with other consonantal distinctions (e.g., Miyawaki, Liberman, Fujimura, Strange, & Jenkins,

1972) and, to a considerably lesser extent, with vowel stimuli under some conditions (Pisoni, 1973; Stevens, Liberman, Studdert-Kennedy, & Ohman, 1969; Fujisaki & Kawashima, 1969).

Although these findings are of considerable importance in and of themselves, they are particularly revealing when the discriminability functions for the synthetic speech stimuli are compared with the discriminability functions for the same acoustic dimensions when presented in a nonspeech context. Mattingly et al (1971) showed that the discontinuities in the discriminability functions, so apparent when the synthetic sounds were perceived as speech, were absent when the stimuli were not perceived as speech. More specifically, they showed that variations in the starting frequency and direction of the second-formant transition (which are sufficient for the perceived distinctions among the voiced stop consonants [b, d, g] and their voiceless counterparts [p, t, k]) were perceived categorically and continuously in speech and nonspeech contexts, respectively. The latter condition was arranged by presenting only the second-formant transitions which are heard as bird-like chirps. Similar effects have also been obtained by Miyawaki et al (1972). In addition, Eimas, Cooper, and Corbit (1973) have shown the acoustic information signaling the onset of voicing is perceived differently when presented in a speech as opposed to a nonspeech context.

Although it is certainly the case that all sounds, speech or nonspeech, must undergo some common auditory analysis, the sounds of speech would appear to undergo some additional, specialized processing that permits the extraction of distinctive phonetic features. It is the process of feature extraction that is categorical in that it reduces the continuous variation in acoustic-auditory information to a set of discrete feature values. The actual phonetic experience, that is, the perception of a particular phone involves additional higher processes, including most likely

\*The studies reported here were supported by Grant HD 05331 from the National Institute of Child Health and Human Development, United States Public Health Service. I wish to thank Jane Carlstrom, Jean Fechheimer, and Sheila Polofsky for their capable assistance in the collection and analysis of the data. I also wish to acknowledge my indebtedness to Franklin S. Cooper for his generosity in making the facilities of the Haskins Laboratories available for the preparation of stimulus materials.

matching a set of features with its appropriate label.

Additional evidence favoring a special speech processor comes from studies of dichotic listening. A number of researchers have shown that speech signals are better perceived by the left hemisphere (right ear) whereas nonspeech signals are better processed by the right hemisphere (left ear) (Kimura, 1961, 1964; Studdert-Kennedy & Shankweiler, 1970). Electrophysiological studies of neural activity have also supported this conclusion in adults (Wood, Goff, & Day, 1971).

Of particular interest have been the recent electrophysiological (Molfese, 1972) and behavioral findings (Eimas, Siqueland, Jusczyk, & Vigorito, 1971) that the specialized speech processes may be operative as early as the first few weeks of life. Eimas et al (1971) investigated the infant's ability to discriminate small differences in voice onset time. Voice onset time, which is defined by Lisker and Abramson (1964) in articulatory terms as the time between the release burst and the onset of laryngeal pulsing or voicing, is a sufficient cue for distinguishing between the voiced and voiceless stop consonants [b] vs [p], [d] vs [t], and [g] vs [k]. They found that infants 1 and 4 months of age were better able to discriminate a 20-msec difference in voice onset time when the two stimuli to be discriminated were from different adult phonetic categories, [b] and [p], than when they were from the same adult phonetic category, [b] or [p]. As was true for adult listeners (Abramson & Lisker, 1970), acoustic variations of the same phonetic feature were not discriminable. In later studies, Eimas (in press) was able to replicate these results with the apical stops, [d, t], and to find some evidence for the categorical perception by young infants of a voicing distinction not found in English.

Additional research has shown that infants between the ages of 2 and 6 months are sensitive to differences in place of articulation, acoustically represented by variations in the second- and third-formant transitions (Moffitt, 1971; Morse, 1972). Although Morse has contended that infants process place distinctions in a linguistically relevant manner, the supporting evidence was not compelling. It is toward a resolution of this issue, that is, the manner in which place information is perceived by infants, that the present studies were directed. Another purpose was to determine whether the mechanisms that are necessary to decode information signaling place of articulation are, as is apparently the case with adult listeners, the unique property of the speech processor.

## EXPERIMENT I

The purpose of Experiment I was to determine whether the perception of place distinctions by young infants was categorical. The stimuli were synthetic

speech sounds that varied in the starting frequency and direction of the second- and third-formant transitions. These variations correspond to the variations in articulatory movements necessary for the production of bilabial stops [b, p] as opposed to apical [d, t] or velar [g, k] stops. Discrimination of a particular acoustic difference was measured under two conditions: (1) when the acoustic variation signifies a change in place of articulation as measured by adult identification functions, and (2) when the variation represents acoustic variants of the same place of articulation. Evidence of greater discriminability in the first condition permits the inference that infants are capable of perceiving information related to place distinctions in a categorical fashion and hence in a linguistically relevant manner.

## Method

**Procedure.** The procedure was a modification of the methodology developed by Siqueland and DeLucia (1969). Each infant was tested individually in a small sound-shielded room, moderately illuminated by the rear-projected image of a colorful object. The visual image, in addition to providing illumination, tended to maintain the infant's orientation to the speaker, which was situated just above the screen, about 45 cm from the infant's head. The infant was placed in a reclining seat and presented with a blind nipple. The nipple was held gently by one of the Es, who listened to music over a set of headphones. The second E monitored the recording apparatus and controlled the presentation of the speech patterns. During the first 2 or 3 min, the high-amplitude criterion and the baseline rate of high-amplitude sucking were established. The amplitude criterion was defined as the level which yielded a baseline sucking rate of approximately 20 or 30 responses/min. By permitting the amplitude criterion to vary from infant to infant, it was possible to reduce the variability associated with baseline rates of sucking as well as to establish a baseline rate for each infant such that changes in either direction could occur without serious contamination by either floor or ceiling effects. Immediately after obtaining the baseline rate, the first speech sound was made contingent upon high-amplitude sucking. If the time between each high-amplitude sucking response was at least 1 sec, then each such response produced one presentation of the stimulus pattern 300 msec in duration plus 700 msec of silence. However, if the infant produced a burst of sucking responses with interresponse times less than 1 sec, as was typical, then each response did not produce one presentation of the stimulus. Rather, each response recycled the timing apparatus and the 1-sec on period began again. This limitation in the presentation of auditory feedback was imposed to prevent the occurrence of the reinforcing sound longer than 1 sec after the last response.

The presentation of an auditory stimulus in this manner typically results in an increase in the infant's rate of high-amplitude sucking, compared with the baseline rate. After several minutes of this contingency (the time varies from infant to infant from about 4 or 5 min to over 15 min), the infant usually shows a decrement in performance, presumably as a result of a diminution in the reinforcing quality of the once novel speech stimulus. When the rate of sucking diminished by 20% or more for 2 consecutive minutes compared with the minute immediately preceding the first minute of decrement, the feedback stimulus was changed without interruption by switching the channel selector on the tape deck. The second synthetic speech pattern was presented, likewise contingent upon high-amplitude sucking, for 4 min, after which the testing session was terminated.

One group of infants, Group D, received two stimuli, one of which signaled the adult phonetic category [d] and the other the

phonetic category [g]. The second group of infants, Group S, likewise received two stimuli, but these sound patterns were acoustic variations of the same adult phonetic category [d]. The order in which the two stimuli were presented was counterbalanced across infants in both groups. The infants in the control condition, Group C, were randomly assigned *one* of the four stimuli administered to Groups D and S. At the point at which a change in stimulation would have occurred for the infants of Groups S and D, the records of the control infants were marked, the channel selector control was switched, and the session continued for 4 min. Thus, the control records were completely comparable with those of the experimental infants. Given that infants are highly responsive to novel stimuli, the presentation of a new, discriminable stimulus would be expected to result in a different rate of sucking when compared with the sucking rates for infants who did not receive a change in stimulation. Thus, either an increase in response rate associated with a change in stimulation greater than that shown by the control infants or a decrease less than that of the controls is taken as inferential evidence that the infants perceived the two stimuli as different.

**Stimuli.** The stimuli were four synthetic speech sounds prepared by means of the parallel resonance synthesizer at the Haskins Laboratories by Pisoni (1971). Each stimulus was 300 msec in duration, with the initial 40 msec simulating a period of closure voicing by means of a low-amplitude first formant centered at 150 Hz. The next 40 msec was a period of transition during which all three formants moved from their starting frequencies to their terminal steady-state values of 743, 1,620, and 2,862 Hz for the first, second, and third formant, respectively. These steady-state formant values are appropriate for the American English vowel [ae]. The steady-state portion of each pattern was 220 msec in duration. The starting values for the second- and third-formants are given in Table 1 for each of the two stimuli presented to Groups D and S, along with their phonetic identification as determined by adult listeners (Pisoni, 1973). All four stimuli had the same first-formant starting values, 150 Hz. As may be seen in Table 1, the only differences among the stimuli were in starting frequencies and direction of the second- and third-formant transitions (and, of course, in one instance, in the phonetic identification or place of articulation value). Of importance to note is the fact that the acoustic differences between the two stimuli used for Group D were very nearly identical to the differences between the stimuli used for Group S. Thus, any differences in discriminability between the two groups cannot be attributed to an inequality in the acoustic differences between the stimuli. For additional details concerning the construction of these stimuli, the reader is referred to Pisoni (1971).

The stimuli were recorded on high-quality magnetic tape from which continuous loops were made, with each 300-msec speech pattern separated by 700 msec of silence. There were several copies of each tape loop. The loops used for Groups D and S had one stimulus on Channel 1 and the second stimulus on Channel 2. The control tapes contained the same stimulus on both channels. This arrangement of the stimuli permitted all infants to be treated in a like manner as well as the nearly instantaneous switching between stimuli.

**Apparatus.** Part of the apparatus was a blind nipple on which the infant sucked. The positive pressure generated by the infant's sucking was transduced to provide a record of all sucking responses as well as a digital record of criterional high-amplitude sucking responses by means of a HP 7202B polygraph and a Hunter digital timer. Additional equipment included a two-channel tape deck (Sony Model TC 850), a VM Model 33-1 speaker, power supply, HP 8805A preamplifier, and Lafayette 5710 event timer, arranged to provide auditory feedback when the power supply was activated by a criterion response. Each sucking response of criterion amplitude activated a power supply for 1 sec (or restarted a period of 1-sec activation), which resulted in a rapid increase in the intensity of auditory feedback stimulation from an inaudible level to one about 13 to 15 dB above the background noise level of 63 dB.

**Table 1**  
**Starting Frequencies (in Hz) of the Second- and Third-Formant Transitions: Experiment I**

		Stimulus 1	Stimulus 2	Difference
Group D	F-2	1845	1996	151
	F-3	2862	2525	337
Group S	F-2	1541	1695	154
	F-3	2862	3195	333

*Note—Adult listeners consistently identified the stimuli for Group S and Stimulus 1 for Group D as [d]. Stimulus 2 for Group D was identified as [g].*

Intensity readings were made with a General Radio Type 1551-C sound-level meter, B scale. The background noise was produced almost entirely by the room exhaust system and the slide projector fan.

**Subjects.** The Ss were 24 2-month-old and 24 3-month-old infants from the Greater Providence, Rhode Island, area. Half of the infants at each age level were males and females. In order to obtain complete data on 48 infants, it was necessary to test 115 infants. The success rate of approximately 40% is typical of the satiation/release-from-satiation procedure that we have used to assess the speech processing capacities of young infants. There were no reliable differences in failures due to experimental condition, age, or sex. The reasons for failure to complete testing were many and included such factors as falling asleep (33%), crying (25%), initial failure to suck on the nipple (20%), ceasing to suck during the course of the experiment (7%), and a group of factors consisting of failure to show satiation, an extremely erratic pattern of sucking, equipment failure, and E error (15%). The majority of the infants (78%) who failed to complete the session were eliminated prior to the shift in stimulation. There were no reliable differences in the infants who were eliminated after the shift as a function of sex or experimental treatment, and 80% of these infants were eliminated for crying or falling asleep. The criterion for the former was loud crying and a failure to take the nipple, whereas the criterion for the latter was the characteristic sucking pattern associated with sleep, that is, brief bursts of very low-amplitude sucking. Most of the postshift eliminations occurred during the first or second postshift minute.

As often as possible, the infants were assigned randomly to the three conditions. However, given the high attrition rate and the restrictions that the three groups (each with 16 infants) be counterbalanced with respect to age, sex, and order of stimulus presentation, random assignments could not always be made.

## Results and Discussion

The mean number of sucking responses is displayed in Fig. 1 as a function of minutes and treatments. An analysis of the minute-by-minute sucking rates for the 5 min immediately prior to the shift in stimulation (or at that point in time when a shift would have occurred in the case of the control infants) revealed nonreliable differences between groups. As would be expected from the nature of the experiment, the mean response rate for the 2 min before shift (the satiation period) was significantly lower than the third minute before stimulus change ( $p < .01$ ).

Inasmuch as the measure of discriminability is based on a change in response rate correlated with a change in stimulation, and given that there were individual differences in response rates during the satiation period, the analyses of postshift performance

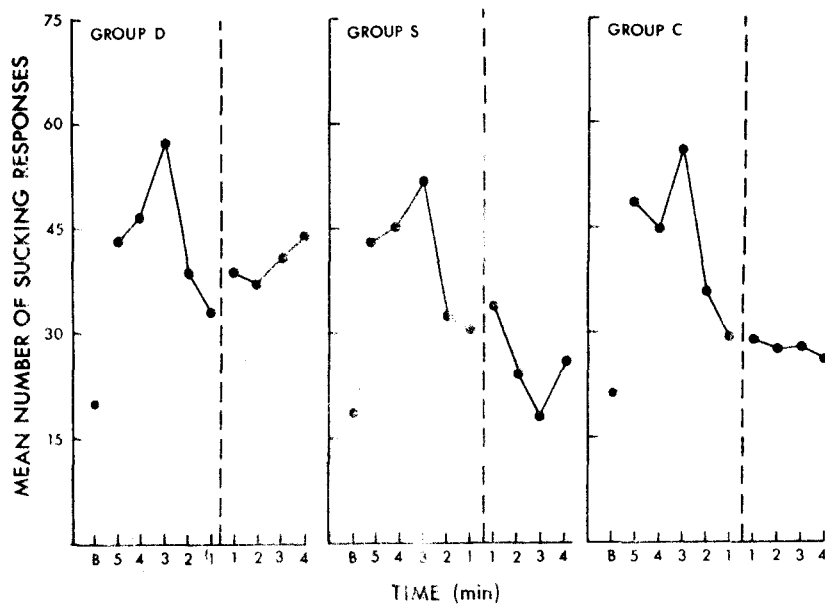


Fig. 1. Mean number of sucking responses as a function of time and experimental conditions. Time is measured with reference to the moment of stimulus shift, which is indicated by the dashed line. The baseline rate of sucking is indicated by the letter "B."

were conducted on difference scores. For each infant, the mean response rate for the 2 min immediately preceding the stimulus shift was subtracted from each of the four 1-min measures of postshift performance. An analysis of variance, Groups by Minutes, performed on these difference scores, revealed a significant groups effect,  $F(2,45) = 4.5$ ,  $p < .025$ , and a significant Groups by Minutes interaction,  $F(6,135) = 8.4$ ,  $p < .001$ . Group D, which showed a mean increment of 4.1 responses/min for the 4-min postshift period, differed reliably from Groups S and C ( $p < .025$  in each instance). Groups S and C, which showed mean decrements of 6.1 and 4.8 responses/min, respectively, did not differ significantly. The significant interaction, as can be seen in Fig. 1, was primarily a function of the increase in rate of responding shown by Group D over the 4-min postshift period, whereas Groups S and C tended to show a decreasing rate of response with time.<sup>1</sup>

The cues for place of articulation, like the cues for voice onset time, are discriminated in a categorical-like manner by young infants; that is to say, the infant's ability to discriminate place cues is determined largely by whether or not the acoustic stimuli signal the same or different values along the phonetic feature, place of articulation. Before considering the perceptual mechanisms that might be responsible for processing of speech in a linguistic mode, we wish to consider whether these mechanisms are part of the more general auditory processing system or whether they are part of the specialized speech processor.

## EXPERIMENT II

Experiment II compared the infant's ability to discriminate acoustic cues for place when these cues

are presented in a speech context, as in Experiment I, and when they are presented in a nonspeech context. The speech stimuli for this experiment were two-formant synthetic speech patterns, which varied in the starting frequency and direction of the second-formant transition only. All other characteristics of the stimuli did not vary from one pattern to another. Variations in the second-formant transition are by themselves sufficient cues to signal variations in place of articulation, and indeed, as Liberman et al (1967) have noted: "...[the second-formant transition] is probably the single most important carrier of linguistic information in the speech signal [p. 434]." Based on identification functions from adult listeners obtained by Mattingly et al (1971), pairs of synthetic speech patterns, separated by a constant acoustic difference, were selected such that the two stimuli were perceived either as different phones, [b] and [d], or as the same phone, [b]. The two different types of stimulus pairs corresponded to the Group D and Group S stimuli of Experiment I. In order to present the identical relevant acoustic information in a nonspeech context, additional patterns were synthesized that eliminated all acoustic features, except the second-formant transitions of the speech stimuli. These patterns, 40 msec in duration, are perceived as bird-like chirps by adult listeners (Mattingly et al, 1971). Thus, both sets of stimuli vary in precisely the same manner, but yet are perceived in radically different ways by adult listeners.

Should both sets of stimuli be perceived in the same categorical manner, then it is possible to conclude that, at least during infancy, the more general auditory processing system is differentially sensitive to variations in rapidly changing frequency values. Furthermore, it would seem that the points of greater discriminability were selected to mark phonetic

boundaries (cf. Stevens, 1972; Stevens et al, 1969). However, should categorical-like perception be characteristic of only the speech stimuli (as in adult listeners), then categorical decisions regarding phonetic feature values cannot be ascribed to the functioning of the general auditory processing system, but rather must be considered to be a result of some form of additional linguistic processing to which only the sounds of speech are subjected.

## Method

**Stimuli.** The speech stimuli were six synthetic speech patterns constructed by means of a parallel resonance synthesizer at the Haskins Laboratories by Mattingly et al (1971). Each stimulus was 245 msec in duration. The initial 15 msec simulated closure voicing by means of a low-amplitude first formant at 150 Hz. The following 40 msec was a transitional period, during which the first and second formants moved from their starting values to their terminal steady-state values of 743 and 1,620 Hz, respectively. The steady-state portions were 190 msec in duration. These stimuli were perceived as the voiced stops [b] or [d] plus the vowel [ae]. The nonspeech stimuli, 40 msec in duration, were likewise synthesized by means of the parallel resonance synthesizer. Each nonspeech pattern corresponded exactly to one of the second-formant transitions of the speech stimuli. Table 2 gives the starting frequencies for the six second-formant transitions. The difference in the second-formant starting frequency between the two stimuli of the two pairs used for Group D and the single pair used for Group S was approximately the same. This equality held, of course, for both the speech and nonspeech stimuli. The stimuli were recorded and made into continuous loops, as in Experiment I. The silent interval between successive stimuli was 755 msec for the speech sounds and 960 msec for the nonspeech sounds.

**Apparatus.** The apparatus was the same as that used in Experiment I.

**Procedure.** The general procedural details were the same as those used in Experiment I. For each type of stimulus, speech or nonspeech, there were three groups of 16 infants, D, S, and C. The infants of Group D (both speech and nonspeech) received one of two pairs of stimuli, half of the infants receiving one pair and the remaining infants the other pair. When the stimuli were speech sounds, one stimulus of each pair signaled the phonetic category [b], while the second stimulus signaled the phonetic category [d]. Two different pairs were used to be sure that, in at least one instance, the two stimuli lay on opposite sides of the phonetic boundary. The infants in Group S received a single pair of stimuli, which, when speech sounds, were both representatives of the category [b]. The control infants (Group C) received one of the six speech or six nonspeech stimuli, randomly selected. All stimuli were presented at approximately 15 dB above a background noise level of 63 dB (B scale, Type 1551-C General Radio sound-level meter).

**Subjects.** The Ss were 96 infants from the Greater Providence, Rhode Island, area. Approximately half of the infants were males and half females. Likewise, approximately half of the sample were 2 months of age and the remaining infants were 3 months old. Forty-four percent of the infants tested completed the experiment, and there were no reliable differences in the rate of failure due to age, sex, type of stimulus, or experimental treatment. Of the 122 infants who failed to complete the experiment, 34% cried, 27% fell asleep, 12% failed to suck on the nipple at all, 8% stopped sucking at some time during the course of the experiment, and 18% were eliminated for a variety of reasons, including failure to achieve the satiation criterion, very erratic responding, and equipment failure. The majority (86%) of the infants who failed to complete the experiment were eliminated before the stimulus change, and of those who were eliminated after the change, 83% either fell asleep or cried. As in the first experiment, the infants who were eliminated

**Table 2**  
Starting Frequencies (in Hz) of the Second-Formant Transitions: Experiment II

			Stimulus		Difference
			1	2	
Group D	Pair 1	F-2	1312	1541	229
	Pair 2	F-2	1386	1620	234
Group S		F-2	1232	1465	233

*Note—Adult listeners consistently identified the stimuli for Group S and Stimulus 1 for both Group D pairs as [b]. Stimulus 2 for both Group D pairs was identified as [d].*

after the stimulus change did not differ reliably due to sex or experimental treatment and most of the infants were eliminated before the third postshift minute. When possible, infants were randomly assigned to the six conditions. But, again, the high attrition rate and the restrictions that the six groups be balanced as closely as possible for age, sex, order of stimulus presentation, and, in the case of the infants in Group D, for stimulus pair, completely random assignments were simply not possible.

## Results and Discussion

An analysis of the response rates for the 5 min immediately preceding the stimulus change revealed no reliable differences due to experimental conditions or type of stimulus, speech or nonspeech. All six groups showed a reliable increment in sucking rate at the third minute before shift compared with the baseline rate ( $p < .01$ ) and a reliable decrement during the final 2 min before shift, as would be expected ( $p < .01$ ).

Difference scores were again used in all analyses of postshift performance. The mean changes in response rates over the final 4 min are shown in Table 3. Preliminary analyses of the postshift data revealed no significant differences as a function of age, sex, order of stimulus presentation, or, in the case of the infants assigned to Group D, the stimulus pair. An analysis of variance, Type of Stimulus (speech vs nonspeech) by Groups (D vs S vs C) by Minutes yielded only one significant effect, namely, the Groups by Type of Stimulus interaction,  $F(2,90) = 4.1$ ,  $p < .025$ . Individual comparisons revealed the following: (1) When the stimuli were synthetic speech patterns, there was evidence for categorical perception in that Group D responded at a reliably higher rate than did Groups S and C, which did not differ from each other. (2) When the stimuli were nonspeech sounds, there

**Table 3**  
Mean Recovery in Response Rate (Responses Per Minute) for the Entire 4-Min Postshift Period: Experiment II

Stimuli	Experimental Condition		
	Group D	Group S	Group C
Speech	+8.5	-4.6	-1.2
Nonspeech	+4.0*	+7.9	+6

*\*In a replication study with nine infants, the mean increment in response rate was +7.3 responses/min.*

was no evidence for categorical perception; Groups D and S did not differ reliably. (3) Although Group S did differ from Group C, the infants of Group D did not differ significantly from the control infants when the stimuli were nonspeech sounds.

It is, of course, possible that the auditory processing system finds some pairs of nonspeech sounds more discriminable than other pairs (cf. Mattingly et al., 1971). On the other hand, it is likewise possible that the slightly depressed performance of the nonspeech Group D (compared with Group S) was due to sampling error. Indeed, the difference between the two groups could be attributed to the relatively poor performance of only four infants in Group D. To test these alternatives, an additional nine infants were tested on their ability to discriminate the two pairs of nonspeech chirps previously used for Group D. The results indicate that these stimuli can be reliably discriminated and that the original estimate of discriminability may well have been underestimated due to sampling error: the mean increase in response rate was 7.3 responses ( $p < .05$ ), very near the performance level of Group S.

Regardless of the true level of discriminability for the stimuli of the nonspeech Group D, it is the case that infants discriminate variations in the second-formant transition quite differently when they are presented in a speech context as opposed to a nonspeech context. Further confirmation of this conclusion comes from comparisons of Groups D and S, which received speech stimuli, with the same groups which received nonspeech stimuli. A 2 by 2 analysis of variance revealed a highly significant Groups by Type of Stimulus interaction,  $F(1,60) = 20.5$ ,  $p < .001$ : Group S, which received speech stimuli, responded less frequently than did the remaining three groups, which, in turn, did not differ significantly from one another.

In summary, the results of these two experiments indicate that infants are capable of categorizing continuous variations in the acoustic dimension that signal adult phonetic distinctions corresponding to place of articulation. Moreover, the manner in which this categorization takes place agrees perfectly with the three major and possibly universal phonetic distinctions based on place of articulation.<sup>2</sup> Finally, the processing system that subserves perception of the phonetic feature of place differs from the system that is capable of processing the same acoustic information in a nonspeech context.

## GENERAL DISCUSSION

It remains now to consider (1) the perceptual mechanisms that could be responsible for the categorical perception of place cues in speech contexts, (2) how acoustic differences that are

discriminable in one context (nonspeech) might not be discriminable in another context (speech), and (3) what functions perception in a speech mode might serve in the infant during the earliest stages of language acquisition.

The major theoretical positions pertaining to the perception of speech are complex models, in that perception is assumed to be more directly linked to internal mediating events than to external acoustic events or their immediate auditory transforms (see, for example, Liberman, 1957, 1970; Liberman et al., 1967; Stevens & House, 1972). Thus, in the motor theory of speech perception of Liberman and his associates, decoding of the incoming acoustic signal is made by reference to invariant motoric commands to the articulatory system, undoubtedly at the level of neural activity. Exactly how the acoustic information makes contact with these central articulatory commands has not as yet been fully explicated. In a similar manner, the analysis-by-synthesis model of Stevens and his colleagues makes the assumption that perception occurs when the temporarily stored speech event, in auditory form, is matched by a second, internally generated or synthesized representation of the speech event in question. This process of generation must necessarily involve considerable knowledge of production processes, more specifically, of the rules by which distinctive phonetic features are converted into articulatory commands and of the rules by which auditory patterns are associated with these commands.

Although motor theories of speech perception are possible models of adult speech perception, the application of this class of model to the speech-processing capacities of the infant raises certain difficulties. Most prominent of these is the need to explain how the young, virtually inarticulate infant has come to possess the complex knowledge required for the conversion of phonetic features to articulatory commands, and how he has come to associate particular auditory patterns with articulatory commands. Without ascribing this knowledge to the biological endowment of the infant, there would appear to be no way in which an infant could have acquired this information by traditional means during the first few weeks or months of life. However, to attribute this knowledge to the infant's biological endowment would seem to extend considerably the cognitive competencies that we are willing to impute to genetically determined factors.

A number of researchers have begun to consider alternate models of speech perception that are based on feature detectors and that do not require knowledge of production routines (Abbs & Sussman, 1971; Cole & Scott, 1972; Cooper, in press; Eimas & Corbit, 1973; Eimas, Cooper, & Corbit, 1973; Lieberman, 1970; Stevens, 1972, 1973).<sup>3</sup> Although

feature detectors which permit phonetic decisions could be either auditory or linguistic in nature, the weight of evidence favors linguistic feature detectors (but see Stevens, 1973, for an account of place discrimination based on auditory property detectors).

Recently, Eimas and Corbit (1973) and Eimas, Cooper, and Corbit (1973) have presented evidence consistent with the hypothesis that there exist phonetic feature detectors for the acoustic consequences of the two modes of voicing found in English and numerous other languages (Lisker & Abramson, 1964). These detectors were assumed to be sensitive to relatively restricted ranges of complex acoustic energy (voice onset time), part of the central speech processing system, and finally the direct mediators of the categorization of voicing information. Similarly, Cooper (in press) and Cooper and Blumstein (1974) have found evidence for the existence of phonetic feature detectors underlying the perception of the acoustic consequences of the three major place distinctions. If these phonetic feature detectors exist and are operative early in the lives of infants, then it is possible to accommodate the categorical perception of infants without recourse to motoric theories of speech perception (Eimas, in press; Cutting & Eimas, in press). Essentially, it has been argued that, given the passive nature of a feature detector system of analysis, the presentation of an acoustic signal with sufficient linguistic information to activate the speech processor (cf. Molfese, 1972) will likewise activate those phonetic feature detectors for which there is an adequate stimulus. Continued presentation of the same speech sound eventuates in the adaptation of the stimulated detectors. This adaptation may be related to the diminution of the reinforcing properties of novel speech stimuli as they become increasingly familiar and the subsequent reduction in effort to obtain the stimulus. The introduction of an acoustically different, but phonetically identical, speech pattern will not be experienced as novel by the infant, whereas an acoustically *and* phonetically different speech sound will be perceived as novel. From a phonetic feature analysis and the infant's strong attraction to novelty, it is easily predicted that the postshift period of Group D should be marked by a substantial increment in response rate to obtain the new stimulus. On the other hand, the postshift period of Group S (and, of course, of Group C) should reveal a continued decrease in the rate of responding. As for accommodating the continuous discrimination of second-formant transition variations when presented in a nonspeech context, we need only assume that there was not sufficient linguistic information in the signal to activate the speech processor. Hence, the analysis of these sounds is restricted to the auditory processing system, the discriminative capacity of

which is more closely tied to physical (acoustic) differences *per se* than is that of the speech analyzers. (For evidence pertaining to the insensitivity of the speech processor to variations in the acoustic difference between two phonetically distinct sounds, the reader is referred to Eimas, in press.) In essence, then, we have tried to argue that the discriminability of speech in infants is based primarily on the detection of different values of one or more phonetic features, such as voice onset time or place of articulation.<sup>4</sup>

The ability of infants (and adults) to discriminate certain variations of second-formant transitions in a nonspeech context and yet be unable to discriminate the same acoustic information when it is presented in a speech context is a seemingly paradoxical finding. As mentioned above, it must be the case that acoustic events, regardless of their nature, initially undergo a considerable degree of common auditory analysis, but as the results indicate, some of the auditory information from this common processing is apparently lost or not available when the stimuli are speech signals. Perhaps the information arising from the additional linguistic analysis that only the sounds of speech undergo takes precedence in some manner over the output information of the auditory analyzers. Of equal likelihood, however, is the hypothesis that linguistic processing reduces the amount of auditory information that can be held in a temporary, probably preperceptual store (cf. Massaro, 1972), as a consequence of the additional time and/or processing capacity required by the linguistic analysis. Thus, after linguistic processing, the relatively brief, but complex, auditory information associated with phonetic features becomes unavailable, for all intents and purposes, to response decision units. Although at this time the available evidence does not permit a resolution of this problem, it is of interest to note that a number of researchers have considered the categorical nature of speech perception to be due, at least in part, to memorial processes that affect the stored auditory analogs of the acoustic input (cf. Fujisaki & Kawashima, 1969; Pisoni, 1973).

The capacity of infants to perceive speech in a linguistic mode through the activation of a special speech processor with phonetic feature detectors serves a number of functions. In the first place, this form of analysis of speech provides the infant with an automatic means for the recognition of speech. As a consequence, speech signals will undergo special processing, without conscious decisions on the part of the listener (infant or adult), provided only that there is sufficient linguistic information in the signal to activate the speech processor (cf. Eimas, Cooper, & Corbit, 1973; Mattingly et al, 1971). Secondly, the speech processing system, with its phonetic feature analyzers, not only provides information concerning the phonetic feature structure of the speech sounds

but also automatically acts to segment the continuous stream of speech into discrete elements—a necessary step for the analysis of speech and language. Finally, this form of processing, in not requiring that the infant actively learn to recognize speech or to segment it into phonetic features must certainly hasten, and perhaps even make possible, the acquisition of human languages.

## REFERENCES

- ABBS, J. H., & SUSSMAN, H. M. Neurophysiological feature detectors and speech perception: A discussion of theoretical implications. *Journal of Speech & Hearing Research*, 1971, **14**, 23-36.
- ABRAMSON, A. S., & LISKER, L. Discriminability along with voicing continuum: Cross-language tests. In *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967*. Prague: Academia, 1970. Pp. 569-573.
- COLE, R. A., & SCOTT, B. Phoneme feature detectors. Paper presented at the meetings of the Eastern Psychological Association, Boston, Massachusetts, April 1972.
- COOPER, W. E. Adaptation of phonetic feature analyzers for place of articulation. *Journal of the Acoustical Society of America*, in press.
- COOPER, W. E., & BLUMSTEIN, S. E. A "labial" feature analyzer in speech perception. *Perception & Psychophysics*, 1974, **16**, 591-600.
- CUTTING, J. E., & EIMAS, P. D. Phonetic feature analyzers in the processing of speech by infants. In J. Kavanagh and J. E. Cutting (Eds.), *The role of speech in language*. Cambridge, Mass: M.I.T. Press, in press.
- EIMAS, P. D. The relation between identification and discrimination along speech and nonspeech continua. *Language & Speech*, 1963, **6**, 206-217.
- EIMAS, P. D. Speech perception in early infancy. In L. B. Cohen and P. Salapatek (Eds.), *Infant perception*. New York: Academic Press, in press.
- 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.
- FUJISAKI, H., & KAWASHIMA, T. On the modes and mechanisms of speech perception. Annual Report of the Engineering Research Institute, Vol. 28, Faculty of Engineering, University of Tokyo, Tokyo, 1969. Pp. 67-73.
- KIMURA, D. Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology*, 1961, **15**, 166-171.
- KIMURA, D. Left-right differences in the perception of melodies. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 355-358.
- LIBERMAN, A. M. Some results of research on speech perception. *Journal of the Acoustical Society of America*, 1957, **29**, 117-123.
- LIBERMAN, A. M. The grammars of speech and language. *Cognitive Psychology*, 1970, **1**, 301-323.
- 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.
- LIEBERMAN, P. Towards a unified phonetic theory. *Linguistic Inquiry*, 1970, **1**, 307-322.
- LISKER, L., & ABRAMSON, A. S. A cross-language study of voicing of initial stops: Acoustical measurements. *Word*, 1964, **20**, 384-422.
- LISKER, L., & ABRAMSON, A. S. The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967*. Prague: Academia, 1970. Pp. 563-567.
- MASSARO, D. W. Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, 1972, **79**, 124-145.
- MATTINGLY, I. G., LIBERMAN, A. M., SYRDAL, A. K., & HALWES, T. Discrimination in speech and nonspeech modes. *Cognitive Psychology*, 1971, **2**, 131-157.
- MIYAWAKI, K., LIBERMAN, A. M., FUJIMURA, O., STRANGE, W., & JENKINS, J. J. Cross-language study of the perception of the F3 cue for [r] versus [l] in speech and nonspeech patterns. Paper presented at the XXth International Congress of Psychology, Tokyo, 1972. (Also in Status Report on Speech Research, Haskins Laboratories, New Haven, Connecticut, 1973, SR-33. Pp. 67-76.)
- MOFFITT, A. R. Consonant cue perception by twenty- to twenty-four-week-old infants. *Child Development*, 1971, **42**, 717-731.
- MOLFESE, D. L. Cerebral asymmetry in infants, children and adults: Auditory evoked responses to speech and noise stimuli. Unpublished PhD dissertation, The Pennsylvania State University, 1972.
- MORSE, P. A. The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology*, 1972, **14**, 477-492.
- PISONI, D. P. On the nature of categorical perception of speech sounds. Unpublished PhD dissertation, University of Michigan, 1971. (Also in Supplement to Status Report on Speech Research, Haskins Laboratories, New Haven, Connecticut, November 1971)
- PISONI, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 1973, **13**, 253-260.
- SIQUELAND, E. R., & DELUCIA, C. A. Visual reinforcement of non-nutritive sucking in human infants. *Science*, 1969, **165**, 1144-1146.
- STEVENS, K. N. The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David, Jr., and P. B. Denes (Eds.), *Human communication: A unified view*. New York: McGraw-Hill, 1972, pp. 51-66.
- STEVENS, K. N. The potential role of property detectors in the perception of consonants. Paper presented at Symposium on Auditory Analysis and Perception of Speech, Leningrad, USSR, August 1973.
- STEVENS, K. N., & HOUSE, A. S. Speech perception. In J. Tobias (Ed.), *Foundations of modern auditory theory*. Vol. II. New York: Academic Press, 1972. Pp. 3-62.
- STEVENS, K. N., LIBERMAN, A. M., STUDDERT-KENNEDY, M., & OHMAN, S. E. G. Cross-language study of vowel perception. *Language & Speech*, 1969, **12**, 1-23.
- STUDDERT-KENNEDY, M. The perception of speech. In T. A. Sebeok (Ed.), *Current trends in linguistics*. Vol. XII. The Hague: Mouton, 1974.
- STUDDERT-KENNEDY, M., & SHANKWEILER, D. P. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 1970, **48**, 579-594.
- WOOD, C. C., GOFF, W. P., & DAY, R. S. Auditory evoked potentials during speech perception. *Science*, 1971, **173**, 1248-1251.

## NOTES

1. In one respect, the present findings differ from those obtained with the dimension of voice onset time: the recovery function for Group D showed a rising function, whereas for voice onset time, the Group D recovery function attained its maximum response rate



during the initial 2 min. This effect appears to be real in that the same rising function for Group D was found in a replication study using similar three-formant synthetic speech stimuli. The recovery function for the comparable Group D in Experiment II, showed a less pronounced increment over time, with the peak response rate occurring at the third postshift minute. The reason for the delay in the postshift response peak with place cues is not clear (but see Cutting and Eimas, in press, for some possibly analogous findings in the adult speech perception literature).

2. Comparison of the pattern of results of the first experiment with that obtained by Pisoni (1971, 1973) indicates that infants may be bound by phonetic feature values in discriminating speech signals even more than adult listeners. Pisoni, using a variety of psychophysical procedures, found that the level of discriminability for the stimuli of Group S was consistently above chance, albeit only 7%-10%. It will be recalled that the recovery function for Group S did not differ from that of the control infants. Moreover, in Experiment II, an even larger acoustic difference than used by Mattingly et al (1971) failed to produce any evidence of within-phonetic category discrimination. Whether infants are actually more constrained than adult listeners by the phonetic feature structure or whether the apparent difference is one attributable to procedural differences remains to be determined. In any event, there appears to be ample evidence that the discrimination of speech signals by infants is controlled to a large

extent by the phonetic composition of the acoustic event.

3. We recognize that there exist, as yet, a number of unresolved problems in hypothesizing a model of speech perception based on the extraction of phonetic information by feature detectors. These problems include, for example, delineating the phonetic features and the invariant stimulus information necessary for the activation of the various detectors. The latter issue is particularly difficult, given the extent to which a number of cues sufficient for signaling phonetic feature distinctions are context-conditioned, and vary greatly from speaker to speaker (cf. Liberman et al, 1967). Whether these issues can be resolved is to a large extent an empirical problem. What we have tried to do in this paper and others is to demonstrate that this form of analysis can accommodate at least a portion of the speech data, whether from infant or adult listeners.

4. It should be borne in mind that the feature detector model is used in the present paper to accommodate the discriminability data only. It is obviously the case that additional, high-level processes of integration, as well as some form of matching sets of distinctive features with appropriate labels, are necessary before an identification response can be made.

(Received for publication March 1, 1974;  
revision received July 6, 1974.)