Perception & Psychophysics
1975, Vol. 18 (5), 341-347

Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants

PETER D. EIMAS
Brown University, Providence, Rhode Island 02912

Infants, 2 and 3 months of age, were found to discriminte stimuli along the acoustic continuum underlying the phonetic contrast [r] vs. [l] in a nearly categorical manner. For an approximately equal acoustic difference, discrimination, as measured by recovery from satiation or familiarization, was reliably better when the two stimuli were exemplars of different phonetic categories than when they were acoustic variations of the same phonetic category. Discrimination of the same acoustic information presented in a nonspeech mode was found to be continuous, that is, determined by acoustic rather than phonetic characteristics of the stimuli. The findings were discussed with reference to the nature of the mechanisms that may determine the processing of complex acoustic signals in young infants and with reference to the role of linguistic experience on the development of speech perception at the phonetic level.

There is now considerable experimental evidence indicating that very young, prearticulate infants are able to use a phonetic code in perceiving speech. This conclusion has been drawn from studies, using synthetic speech stimuli, that have measured the infant's ability to discriminate segmental distinctions based on differences in the distinctive phonetic features of voicing (Eimas, in press, a; Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Lasky, Syrdal-Lasky, & Klein, in press; Miller, 1974; Streeter, 1974; Moffitt & Pankhurst, Note 1) and of place of articulation (Eimas, 1974; Morse, 1972).

With regard to syllable-initial distinctions based on voicing ([b] vs. [p], for example), Lisker and Abramson (1964) have identified three major, and quite possibly universal, modes of voicing in the stop consonants: (1) the prevoiced stop in which the onset of voicing precedes the release burst, (2) the voiced stop in which voicing onset lags just slightly behind the release burst, and (3) the voiceless stop in which voicing lags appreciably behind the release of acoustic energy. English uses the voiced and voiceless modes, but there are numerous languages that use the prevoiced and voiced modes or all three modes. Lisker and Abramson (1970) and Abramson and Lisker (1970, Note 2) have been able to identify and vary the acoustic dimension (voice onset time, or VOT) that is effective in signaling voicing distinctions. For English-speaking listeners, the phonetic quality of synthetic speech patterns changed abruptly from a voiced to a voiceless stop as VOT lagged by approximately 25 msec behind the initial release. Discrimination of differences in VOT was found to be very nearly categorical, in that a given difference in VOT was readily detected when the two stimuli differed phonetically as well as acoustically, but the same difference was detected only slightly better than chance when the two stimuli differed only acoustically. From these data (and the data from quite different experimental paradigms, e.g., Studdert-Kennedy & Shankweiler, 1970; Wood, Goff, & Day, 1971), the inference has been made that speech, even meaningless syllables, undergoes some additional speech-specific processing. One function of this special processing is the extraction of a phonetic feature message or code which is then used for decisions about speech events at the segmental level (cf. Eimas, in press, a; Eimas & Corbit, 1973; Liberman, 1970; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Pisoni, 1973).

Eimas et al. (1971), among others cited previously, found that infants as young as 1 month of age discriminated differences in VOT in a nearly categorical manner. Given the considerable overall similarity of the adult and infant discriminability data, infants, apparently, also have access to a phonetic feature code for purposes of deciding whether two speech events are the same or different. In addition, there is evidence that the ability to perceive voicing distinctions in accord with a phonetic feature code during early infancy is independent of the infant's linguistic environment. Infants were found to distinguish categorically or nearly so between

The research reported here was supported by Grant HD 05331 from the National Institute of Child Health and Human Development, USPHS. Preparation of the manuscript was completed while the author was a Guggenheim Fellow at the Haskins Laboratories. I wish to thank Ms. Jane Carlstrom and Ms. Sheila Polofsky for their capable assistance in the collection and analysis of the data. I also wish to express my gratitude to Dr. Franklin S. Cooper and Dr. Alvin M. Liberman for their generosity in making the facilities of the Haskins Laboratories available for the preparation of stimulus materials, and to Dr. Joanne L. Miller for her very helpful comments on an earlier version of this paper.

the prevoiced stop and the voiced stop, even though this distinction is not phonemic or allophonic (i.e., context conditioned) in the parental language, whether English (Eimas, in press, a; Moffitt & Pankhurst, Note 1) or Spanish (Lasky et al., in press). The Lasky et al. study was particularly noteworthy: not only did Guatamalan infants categorize the VOT continuum into three classes of voicing, whereas their parents were most likely able to differentiate only two categories (Abramson & Lisker, Note 2), but it was also the case that the boundary value for the two categories that the infants and adults have in common differed substantially. [The results of the infant research locate the two boundaries between -20 and -60 msec and between +20 and +60 msec, whereas Williams (1974) has found the boundary value for monolingual speakers of Spanish to be - 4 msec.] The data strongly suggest that infants are capable of categorizing the VOT dimension into the three major classes of voicing and that they need not have had specific receptive experience with these categories before this capability is present. Moreover, specific experience provided by the linguistic environment may not determine phonetic boundary values during early infancy.

Experiments with adult listeners on the perception of the acoustic correlates of place of articulation ([b] vs. [d] vs. [g], for example) revealed that the processing was again very nearly categorical, that is, linguistic in nature (Eimas, 1963; Liberman, Harris, Hoffman, & Griffith, 1957; Mattingly, Liberman, Syrdal, & Halwes, 1971). Of particular interest is the finding of Mattingly et al. that the nature of the processing of the cues for place distinctions varied with the context in which they were presented, being categorical in a speech context, as mentioned, and much more nearly continuous in a nonspeech context, that is, when the critical acoustic information is removed from the speech pattern and presented in isolation. Adult listeners typically perceived the sounds as nonspeech-like chirps or bleats. In a recent study, Eimas (1974) found that 2- and 3-month-old infants also perceived the acoustic correlates of place distinctions in a categorical manner when the cues were presented in a speech context, and moreover, the three categories used by the infants matched the three categories of place of articulation used by adult listeners. Finally, the infants' perception of the same acoustic information in a nonspeech context was essentially continuous: the discriminability function was determined by acoustic parameters and not by phonetic features.

The conclusions drawn from these studies were that infants, well before they can consistently articulate voicing or place distinctions, have access to mechanisms for the phonetic processing of speech. Moreover, given the extremely early age at which this system is operational, it is quite likely that it is part of the biological endowment of the human infant. Although this conclusion implies a strong genetic determination of phonetic categories and boundaries, it does not actually preclude the modification of the mechanisms underlying the categorization of speech. Indeed, the data of Lasky et al. (in press) demand that modifications in the loci of the phonetic boundary of infants from Spanish environments occur if there is to be effective communication between generations as there always is (at least on the phonetic level).

The present study investigated the infant's ability to process the distinction between syllable-initial [r] and [l], which is effectively signaled by the initial steady-state frequency of the third formant. These phonetic contrasts are members of the manner class called liquids and have properties of both consonants and vowels. Miyawaki, Strange, Verbrugge, Liberman, Jenkins, and Fujimura (1975) have shown that adult American speakers, for whom this contrast is phonemic /read/ vs. /lead/, for example), perceived the relevant acoustic dimension in an essentially categorical manner. However, listeners were able to use acoustic-auditory information more than is usually the case in making within-category difference judgments, especially for the acoustic variations of [l]. The authors also found that perception of the same acoustic information in a nonspeech setting was continuous, and hence not determined by phonetic features.

It is of interest that perception of these same speech sounds by Japanese adult listeners was quite different. The level of discriminability was depressed compared with that of the American listeners and there was an absence of a peak at the region of the phonetic boundary. Japanese, of course, does not distinguish between [r] and [l] either as phonemes or as context-conditional allophones. The discriminability function for the same acoustic information in a nonspeech setting was, on the other hand, virtually identical to that obtained with American listeners. If the nonspeech condition is an adequate procedure for presenting speech-relevant information in a context that is not speech-like, as Miyawaki et al. and others believe it is, then the differences shown by the Japanese listeners are clearly phonetic in nature.

An investigation of the perception of the [r-l] distinction in young infants serves a number of purposes: (1) It extends our knowledge of the speech processing capacities of the infant to a new class of sounds; (2) by presenting the relevant acoustic information in speech and nonspeech modes, it tests the generality of the hypothesis that there exist two systems for processing acoustic signals, an auditory system and a phonetic system; and (3) by comparing the infant data with the adult Japanese data, additional evidence on the role of linguistic experience on speech perception is obtainable. To accomplish these ends, 2- and 3-month-old infants were tested under two modes of stimulus presentation: a speech mode and a nonspeech mode. Within each presentation mode, there were found groups of infants: Group D, Group $S_{[r]}$, Group $S_{[l]}$, and Group C. In the speech condition, Group D received two to-be-discriminated synthetic speech stimuli that were exemplars of different phonetic categories, [r] and [l]. Groups $S_{[r]}$ and $S_{[l]}$ each received two stimuli from the same phonetic category, [r] and [l], respectively. Group C was administered a single stimulus and served as a control for nonstimulus-related changes in the dependent measure. In the nonspeech condition, the four groups received the same variations in the relevant acoustic dimension, but now this information was removed from the synthetic speech pattern and presented in isolation. Given that the acoustic differences are as nearly equal as possible between the stimuli of Groups D, $S_{[r]}$, and $S_{[l]}$ and between the stimuli of the two presentation modes, differences in the relative discriminability of stimulus pairs either across groups or across presentation modes cannot be attributed to stimulus factors per se, but rather must be attributed to differences in the manner in which the stimuli are processed.

METHOD

Procedure

Each infant was tested individually in a sound-shielded room, dimly illuminated by light from the rear-projected image of a colorful pattern. The image was located just below the speaker, which was approximately 45 cm from the infant's head and also served to maintain the infant's orientation to the speaker. The infant was placed in a reclining seat and given a blind nipple, which was held gently, but firmly, in place by one of the experimenters, who wore headphones and listened to recorded music during the course of the experiment. The second experimenter monitored the apparatus.

The actual experimental procedure was a modification of the technique developed by Siqueland and De Lucia (1969) and proceeded as follows. Prior to presentation of auditory stimulation, the high-amplitude sucking criterion and baseline rate of high-amplitude, nonnutritive sucking were established. The high-amplitude criterion was defined for each individual infant as the sucking amplitude that yielded a sucking rate of about 20 to 30 responses/min. As a consequence of this method of determining the amplitude criterion, relatively substantial changes in sucking rates in either direction were possible without serious contamination by floor or ceiling effects. Immediately after establishing the baseline rate, the auditory stimulus was made contingent on high-amplitude sucking. If the time between each criterion response was 1 sec or more, then each response produced one presentation of the stimulus, 500 msec in duration, and 500 msec of silence. If the infant produced a burst of high-amplitude responses with interresponse times less than 1 sec, as was usually the case, then each response recycled the timing apparatus and the 1-sec period began again. This limitation in feedback stimulation was imposed in order to prevent the occurrence of reinforcing events longer than 1 sec after the last criterion response of a sequence.

Presentation of feedback stimulation in this manner typically produces an increment in the infant's high-amplitude sucking rate, compared with the baseline rate, followed by a marked decrement in performance. Presumably the increment occurs as a result of the reinforcing power of novel stimuli, whereas the decrement occurs because of satiation or a loss in the reinforcing properties of the stimulus with increasing familiarization. When the sucking rate decreased by 20% or more for 2 consecutive minutes compared with the minute immediately preceding the first minute of satiation, the feedback stimulus was changed without interruption by switching the channel selector on the tape deck. In the case of control infants, at the moment in time when a change in stimulation should have occurred, the channel was switched and the experimental session continued but with the original stimulus, which had been recorded on both channels. The postshift period was 4 min in duration, and at its completion the experimental session was terminated.

Inasmuch as infants are highly responsive to novel stimuli, the presentation of a new, discriminably different stimulus would be expected to produce a different, that is, higher, rate of sucking compared with the performance of the control infants. Consequently, evidence that infants discriminated two stimuli is inferred when there is an increment in sucking associated with a change in stimulation greater than that shown by the control infants or a decrement less than that demonstrated by the control subjects. It should be noted, however, that the assumptions that novel stimuli are initially reinforcing and then less and less so with continued presentation need not necessarily be true in order to conclude from this procedure that infants are able to differentiate a pair of stimuli. All that is necessary for this inference is the observation that the second stimulus of at least one experimental treatment is responded to in a manner different from that exhibited by the control infants. It is of interest, however, that the significant changes have been in the form of increases in response rates and thus in accord with the presumed properties of novel stimuli.

Stimuli

The stimuli were 10 speech and 10 nonspeech patterns originally produced by means of the computer-controlled parallel resonance synthesizer at the Haskins Laboratories by Miyawaki et al. (1975). Each stimulus was 500 msec in duration with an initial steady-state portion of 50 msec, followed by a transition period of 75 msec, during which the formant frequencies were modulated from their initial to final steady-state values in a linear fashion. The final steady-state period was maintained for 375 msec.

For the speech stimuli, the frequency values for the first and second formants were the same for all 10 stimuli. The first formant had a center frequency of 311 Hz during the initial steady-state and a final steady-state frequency of 769 Hz. The center frequency for the second formant was 1,232 Hz for the entire syllable. The third formant varied in the center frequency of the initial steady state and hence in the starting frequency of the third-formant transition. The terminal frequency of the third formant was 2,525 Hz, which, in combination with the terminal requencies of the first two formants, resulted in the perceived vowel [a]. When the initial steady-state frequency of the third formant is low, the pattern is perceived as [r] plus the vowel [a], and when the initial frequency is high, the perception is [l] plus the vowel [a]. The initial third-formant frequency values of the 10 stimuli along with the differences in hertz between initial frequencies for each stimulus pair are given in Table 1.

In each of the three experimental groups, D, $S_{[r]}$, and $S_{[l]}$, two pairs of stimuli were used. In each pair of stimuli for Group D, Stimulus 1 was perceived by adult listeners as [r] and Stimulus 2 as [l]. For both pairs of stimuli for Groups $S_{[r]}$ and $S_{[l]}$, the two stimuli were perceived as [r] and as [l], respectively. The difference in starting frequency between Stimulus 1 and Stimulus 2 for the six pairs was equated as closely as the synthesizer would permit, and averaged 510 Hz for Group D and 496 Hz for Groups $S_{[r]}$ and $S_{[l]}$ together. The control infants each received one of the 10 stimuli, randomly selected.

The nonspeech stimuli were generated by setting the amplitudes for the first and second formants to zero throughout the syllable, thereby eliminating any energy in these frequency ranges. As a result, the isolated third-formant patterns were identical in every respect to the third formants in the speech stimuli. They were, however, perceived by adults not as speech but rather as

Table 1
Starting Frequencies (in Hz) of the Initial Steady-State Portion of the Third Formant

		Stimulus 1	Stimulus 2	Difference
Group D	Pair 1	2018*	2525	507
	Pair 2	2180	2694	514
Group $S_{[r]}$	Pair 1	1361	1849	488
	Pair 2	1524	2018	496
Group $S_{[l]}$	Pair 1	2694	3195	501
	Pair 2	2862	3363	501

**The first stimulus in each of the pairs for Group D was consistently identified as [r] by adult listeners, as were, of course, the stimuli for Group $S_{[r]}$. All of the remaining stimuli were perceived as [l] by adult listeners.*

high-pitched glissandos or bird-like chirps followed by a steady pitch. Nonspeech third-formant patterns were paired to correspond exactly with the stimulus pairs of the speech groups, with the consequence that the acoustic difference between pairs and across groups was the same for speech and nonspeech stimuli. Each nonspeech control infant likewise received one of the 10 nonspeech patterns, randomly selected.

All stimuli were excited by a periodic source for their entire duration. They also had a falling pitch contour and an overall amplitude contour that began 15 dB down and rose over the first 50 msec of the pattern.

The stimuli were recorded on high-quality magnetic tape, from which several continuous loops were made of each pair. Each 500-msec sound was separated from each other by 500 msec of silence. The experimental tape loops had Stimulus 1 on Channel 1 and Stimulus 2 on Channel 2. The control loops had the same stimulus on both channels.

Within each experimental group (speech and nonspeech Groups D, $S_{[r]}$, and $S_{[l]}$), half of the infants received the first pair and half received the second pair. In addition, half of the infants received Stimulus 1 during the familiarization stage and Stimulus 2 during the postshift period. The converse was true for the remaining infants.

Apparatus

The details of the apparatus have been described elsewhere (Eimas, 1974). In essence, each high-amplitude sucking response, which was recorded polygraphically and digitally, activated a power supply for 1 sec (or restarted the 1-sec cycle) that, in turn, produced a rapid increase in the intensity of the auditory feedback stimuli from an initially inaudible level to one approximately 15 dB above the background noise level of 63 dB SPL caused by the ventilation system and slide projector. All speech and nonspeech stimuli were presented at the same intensity level.

Subjects

The subjects were 64 2-month-old and 64 3-month-old infants from the greater Providence, Rhode Island, area. Half of the infants at each age level were males and half were females. It was necessary to test 294 infants in order to obtain complete data for 128 infants. The success rate of 44% is in line with the success rates (range 40%-50%) of the past 5 years of experimentaion with infants using the high-amplitude sucking procedure coupled with auditory feedback. There were no reliable differences in failures as a function of mode of presentation, treatment condition, age, or sex. The reasons for failure to complete testing were crying (32%), sleeping (32%), cessation of sucking or failure to suck initially (28%), failure to satiate, extremely erratic sucking patterns, and error (8%). Of the infants who failed to complete the experiment, 88% were eliminated before the change in stimulation. Of the 20 infants who were eliminated during the postshift phase, 75% were terminated for crying or falling asleep, usually during the first minute or two, and 25% were eliminated for failure to suck during the first minute of stimulus change. There were no reliable differences in the rate of failure as a function of experimental treatment during the postshift period.

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

RESULTS

The mean number of high-amplitude sucking responses is displayed in Figure 1 as a function of minutes, groups, and modes of stimulus presentation. An analysis of variance (Stimulus Mode by Groups by Minutes) of the sucking rates for the 5 min immediately preceding the shift in stimulation (or the point in time when a shift would have occurred in the case of the control infants) revealed statistically nonsignificant differences for groups and the interaction of Groups by Minutes. There was a reliable difference in the rate of sucking between infants who received the speech stimuli and those who received the nonspeech patterns [$F(1,120) = 5.7$, $p < .05$], with the speech stimuli producing the higher rate; 47.8 responses/min vs. 41.6 responses/min. Although an analysis of baseline scores revealed a nonsignificant difference between the two modes of presentation, the direction of the difference favored the speech mode: 27.5 responses/min vs. 24.3

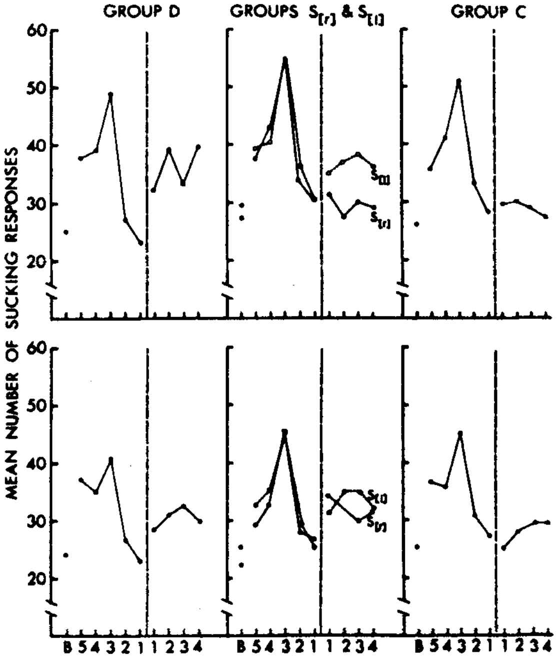

Figure 1. Mean number of sucking responses as a function of time, experimental conditions, and mode of stimulus presentation. The data from infants receiving speech patterns are shown in the upper functions. Time is measured with reference to the moment of the stimulus shift, which is marked by the vertical dashed line. The baseline rate of sucking is indicated by the letter "B."

responses/min. Consequently, conclusions to the effect that speech stimuli are the more preferred stimuli, evoke greater attention, or are better reinforcers must be made cautiously, especially since this effect was not found in an earlier study (Eimas, 1974). As has consistently occurred in this paradigm, there was a reliable increment in the response rate from baseline to the third minute before the stimulus change, and, of course, a reliable decrement in responding over the final 2 min of the familiarization phase ($p < .01$ for both measures for all eight groups). Neither of these measures differed reliably as a function of groups within each of the two stimulus conditions.

The analysis of postshift performance was based on difference scores (see Eimas, 1974, for the rationale behind this measure). For each infant, the mean response rate for the 2 min immediately prior to the stimulus shift (or comparable minutes in the case of the control infants) was subtracted from each of the four 1-min measures of postshift behavior. The mean change in response rate, averaged over minutes and infants in each of the eight groups, is shown in Table 2. Also displayed in Table 2 are the percentages of correct discriminative judgments for the same stimulus pairs for adult American listeners, which were obtained by Miyawaki et al. (in press). Consider the speech discrimination data first. In terms of the overall trend, the infant data corresponded remarkably well with the adult data: there was a virtually perfect correlation between the magnitude of recovery from satiation and the percentage of correct judgments. An analysis of variance of the infant speech data (Groups by Minutes) revealed a significant groups effect [$F(3,60) = 4.3$, $p < .01$] and nonsignificant differences for the effects of minutes and the interaction of minutes and groups. Individual comparisons showed that Group D differed from Group $S_{[r]}$ and Group C at the .02 level or better and that the latter two groups did not differ from each other. In addition, Group D differed from Group $S_{[l]}$, but only at the .06 level, and Group $S_{[l]}$ did not differ reliably from Group $S_{[r]}$ or Group C, although Group $S_{[l]}$ did have a higher level of postshift responding than either of the latter two groups, as might have been expected given the adult data. The data are consistent with the hypothesis that infants process speech largely in terms of the phonetic structure of the stimuli: there was a reliable recovery of the sucking response when the stimuli differed phonetically, but there was not a reliable recovery when the stimuli differed in acoustic information only.[1] However, there was some indication that the infants were able to use nonphonetic information in discriminating acoustic variations of the same phonetic category, although to a lesser extent than was the case for American adult listeners.[2]

Examination of the nonspeech recovery data reveals a very different pattern of behavior. As is apparent in

Table 2
Mean Recovery Data During the 4 Min of Postshift Stimulation for the [r-l] Distinction

	Experimental Conditions			
	Group D	Group $S_{[r]}$	Group $S_{[l]}$	Group C
Speech	11.3 (79%)*	−2.3 (44%)	3.4 (57%)	−1.6
Nonspeech	5.6 (85%)	4.2 (70%)	5.9 (75%)	−.9

**The numbers in parentheses represent the percentages of correct discriminations (chance equals 33%) for adult American English speakers (from Miyawaki et al., 1975). Within each experimental condition, the stimulus pairs were the same for the infant and adult subjects.*

Figure 1 and Table 2, there was relatively little difference with respect to the magnitude of recovery as a function of stimulus assignment. An analysis of the nonspeech data showed no reliable differences across the three experimental groups. Moreover, the general level of recovery of these groups was not very high; in fact, only when the data of the three experimental groups were combined did they differ from the results of the control groups [$t(62) = 2.0$, $p < .05$]. Thus, there was no evidence for the categorical perception of this continuum in a nonspeech context and there was only weak evidence that the nonspeech stimuli were discriminated by the infants. Exactly why the nonspeech patterns in the present study yielded less recovery than did the nonspeech patterns corresponding to variations in place of articulation (Eimas, 1974) is not known at the present time. Perhaps the [r-l] nonspeech stimuli were simply less discriminable or perhaps the second nonspeech pattern was a less potent reinforcer or focus of attention. Regardless of the reason or reasons for the lower level of performance by the nonspeech experimental groups, the major conclusion of the present experiment stands intact: perception of the speech patterns, representing the [r-l] distinction, was based to a large extent on a phonetic code, whereas perception of the same acoustic variations in a nonspeech context was perceived in accord with an acoustic-auditory code.[3]

DISCUSSION

The failure to find a peak in the discriminability function for the nonspeech stimuli in the region corresponding to the phonetic boundary is in opposition to the hypothesis that phonetic boundaries are situated at points of high discriminability along simple or complex acoustic continua and that permissible acoustic variations of phonetic categories are located at regions of low discriminability (cf. Stevens, Liberman, Studdert-Kennedy, & Ohman, 1969). Rather, the present data and the data from Eimas (1974) indicate that speech stimuli undergo some additional, speech-specific processing (cf.

Liberman, 1970; Liberman et al., 1967; Mattingly et al., 1971) that results in the categorization of speech signals into matrices of distinctive phonetic features. Furthermore, it is assumed that it is the output of this special process of categorization that, at least in part, makes some acoustic differences more discriminable than others.

As for the actual mechanisms that underlie the categorical processing of speech and that may be part of the biological structure of the human infant, there is experimental evidence from adult listeners that these mechanisms may take the form of feature detectors (Cooper, 1974; Eimas, Cooper, & Corbit, 1973; Eimas & Corbit, 1973; Tartter & Eimas, in press). These detectors have been assumed to have as their input information from lower level auditory detectors and to have as their output the equivalent of distinctive phonetic feature values. To accommodate the infant speech perception data, it is necessary to assume that phonetic detector systems capable of making most, if not all, of the phonetic distinctions found in human speech, developed over the course of human evolution; that they are operative very early in life; and that they require little, if any, exposure to speech to be set in operation. As mentioned previously, the data of Lasky et al. (in press), among others, strongly suggests that if receptive phonetic experience is necessary, it does not need to be specific. (For a more complete explication of how this model can accommodate the infant data, the reader is referred to Eimas, in press, a.)

A detector model of this nature can also be extended to account for the considerably greater categorical perception of the [r-l] contrast by infants than by adult Japanese speakers. The latter correctly discriminated the synthetic speech stimuli of Group D 49% of the time, of Group $S_{[r]}$, 44% of the time, and of Group $S_{[l]}$, 54% of the time (Miyawaki et al., 1975). It is only necessary to postulate that if specific phonetic experience is not forthcoming within some broadly defined critical period (cf. Lenneberg, 1967), the phonetic feature detectors subserving the contrast in question will undergo some loss of sensitivity or receptivity.[4] In line with this reasoning, it is noteworthy that of the 3 of the 24 Japanese listeners who were able to perceive the [r-l] contrast in essentially the same manner as did the American listeners, one attended an English-speaking school and another had lived in Germany between the ages of 12 and 16 years. The third subject had not had exposure to a language that uses the [r-l] distinction. However, since all of the Japanese listeners had studied English for 10 or more years, but in an educational system that emphasized reading and writing, it appears that specific phonetic competence in adulthood probably requires rather intensive and specific phonetic experience prior to maturity. In accord with this explanation of speech perception, there is the evidence from infant and adult listeners, particularly on the perception of voicing contrasts, indicating that the course of development of phonetic competence is one characterized by a loss of abilities over time, if specific experience is not forthcoming. This is in marked contrast to most theories of perceptual development (e.g., Gibson, 1969) that assume a gradually increasing ability to extract and utilize the relevant or critical information provided by the external world. It would be of considerable theoretical importance to determine whether this course of development is restricted to speech perception or whether there are analogous phenomena in the development of processing capacities for other forms of information.

REFERENCE NOTES

1. Moffitt, A. R., & Pankhurst, D. J. *Speech perception during infancy.* Unpublished manuscript, Carleton University, Ottawa, Canada, 1973.

2. Abramson, A. S., & Lisker, L. *Voice-timing perception in Spanish word-initial stops.* In Status Report on Speech Perception, January-June, 1972, Haskins Laboratories, New Haven, Connecticut, pp. 15-25.

3. 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.

REFERENCES

Abramson, A. S., & Lisker, L. Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the Sixth International Congress of Phonetic Sciences*, Prague, 1967. Prague: Academic, 1970. Pp. 569-573.

Cooper, W. E. Adaptation of phonetic feature analyzers for place of articulation. *Journal of the Acoustical Society of America*, 1974, 56, 617-627.

Eimas, P. D. The relation between identification and discrimination along speech and nonspeech continua. *Language and Speech*, 1963, 6, 206-217.

Eimas, P. D. Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, 1974, 16, 513-521.

Eimas, P. D. Speech perception in early infancy. In L. B. Cohen & P. Salapatek (Eds.), *Infant perception*. New York: Academic Press, in press. (a)

Eimas, P. D. Developmental aspects of speech perception. In R. Held, H. Leibowitz, & H. L. Teuber (Eds.), *Handbook of sensory physiology: Perception*. New York: Springer-Verlag, in press. (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.

Gibson, E. J. *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts, 1969.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, 1975, in press.

Lenneberg, E. H. *The biological foundations of language.* New York: Wiley, 1967.
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.
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.
Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. Discrimination in speech and nonspeech modes. *Cognitive Psychology*, 1971, **2**, 131-157.
Miller, J. L. Phonetic determinants of infant speech perception. Unpublished PhD dissertation, University of Minnesota, 1974.
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. P. 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 nonnutritive sucking in human infants. *Science*, 1969, **165**, 1144-1146.
Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., & Ohman, S. E. G. Cross-language study of vowel perception. *Language and Speech*, 1969, **12**, 1-23.
Streeter, L. A. *The effects of linguistic experience on phonetic Perception.* Unpublished PhD dissertation, Columbia University, 1974.
Studdert-Kennedy, M., & Shankweiler, D. P. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 1970, **48**, 579-594.
Targer, V. C., & Eimas, P. D. The role of auditory feature detectors in the perception of speech. *Perception & Psychophysics*, in press.
Williams, L. *Speech perception and production as a function of exposure to a second language.* Unpublished PhD dissertation, Harvard University, 1974.
Wood, C. C., Goff, W. P., & Day, R. S. Auditory evoked potentials during speech perception. *Science*, 1971, **173**, 1248-1251.

NOTES

1. As mentioned above, the acoustic differences between stimulus pairs were not equal: there was a greater difference in the pairs of Group D. Consequently, there exists the possibility that the greater recovery by Group D with speech stimuli was a function of the greater acoustic difference. A number of arguments can be made against this conclusion, however. First, the evidence on discrimination of simple frequency differences indicates that infants are relatively insensitive to small differences in frequency (Eimas, in press b). Thus, the 14-Hz difference of differences between the stimulus pairs of Group D and Groups $S_{[r]}$ and $S_{[l]}$ together most likely had little influence on the magnitude of recovery Second, if recovery of the speech groups was determined to a large extent by acoustic factors, then the rank ordering of the nonspeech groups should have at least matched that of the experimental speech groups and a greater spread of recovery scores should have occurred among the nonspeech groups. Third, there is some evidence that the infants did not use an even greater difference in acoustic information when it was available, but which was used by adult listeners (see the comparison of Groups $S_{[r]}$ and C and see Note 2 immediately below). While some of the recovery of the Group D speech subjects may have been determined by acoustic factors, it is extremely unlikely that the entire effect or even a major portion of the effect was determined in this manner.

2. The depressed use of acoustic information in making within-category discriminations by infants is best illustrated by the findings that adult listeners (Americans) showed an 11% difference from chance (which is probably statistically reliable) in discriminating the acoustic variations of [r], whereas the infants of Group $S_{[r]}$ and Group C produced virtually the same recovery score. A similar finding was obtained by Eimas (1974). It is possible that this finding is attributable to procedural differences. However, there is another and more interesting explanation. Fujisaki and Kawashima (Note 3) and Pisoni (1973) have noted that categorical perception may be at least partially a function of processes that affect the availability of the auditory representations of the acoustic inputs. If infants are less able to encode, store, or retrieve the auditory representations of speech stimuli than are adult listeners, then their within-category discrimination levels would be depressed relative to those of adult subjects. We are unable to state at this time exactly why auditory information may be less available than phonetic information and why this effect may be even more pronounced in infants than in adults (but see Eimas, 1974, for some speculations on the problem).

3. It may be that the difference between the speech and nonspeech discriminability functions in infants and adults does not reflect a difference in the way in which the two types of stimuli were processed, but rather reflects some complex form of masking or interference that occurred only with the speech patterns and then only for some patterns for American speakers but for all patterns with the Japanese speakers. However, until we are able to account for this complex masking in terms of what is known about auditory perception, this explanation remains a matter of speculation and certainly is not more acceptable than the hypothesis that there exist fundamental differences in the processing of speech and nonspeech signals and that the difference between Japanese and American speakers occurs at the level of phonetic processing (cf. Miyawaki et al., 1975).

4. As for the ability of Japanese adults to discriminate the relevant speech information in a nonspeech mode of presentation, there is no reason to believe that their lower-level auditory analyzers should have suffered any loss of sensitivity with development. It is undoubtedly the case that the natural environment provided many examples of sounds and transient sounds in the frequency range under question. However, that the Japanese listeners were unable to use the same auditory (even in a continuous manner) when the mode of presentation was speech presents an interesting and important problem.

(Received for publication April 14, 1975;
revision received August 5, 1975.)