

# Enhanced discriminability at the phonetic boundaries for the place feature in macaques

Patricia K. Kuhl and Denise M. Padden

*Department of Speech and Hearing Sciences and Regional Primate Research Center, University of Washington, Seattle, Washington 98195*

(Received 23 April 1982; accepted for publication 25 October 1982)

Discrimination of speech-sound pairs drawn from a computer-generated continuum in which syllables varied along the place of articulation phonetic feature (/b, d, g/) was tested with macaques. The acoustic feature that was varied along the two-formant 15-step continuum was the starting frequency of the second-formant transition. Discrimination of stimulus pairs separated by two steps was tested along the entire continuum in a same-different task. Results demonstrated that peaks in the discrimination functions occur for macaques at the "phonetic boundaries" which separate the /b-d/ and /d-g/ categories for human listeners. The data support two conclusions. First, although current theoretical accounts of place perception by human adults suggest that isolated second-formant transitions are "secondary" cues, learned by association with primary cues, the animal data are more compatible with the notion that second-formant transitions are sufficient to allow the appropriate partitioning of a place continuum in the absence of associative pairing with other more complex cues. Second, we discuss two potential roles played by audition in the evolution of the acoustics of language. One is that audition provided a set of "natural psychophysical boundaries," based on rather simple acoustic properties, which guided the selection of the phonetic repertoire but did not solely determine it; the other is that audition provided a set of rules for the formation of "natural classes" of sound and that phonetic units met those criteria. The data provided in this experiment provide support for the former. Experiments that could more clearly differentiate the two hypotheses are described.

PACS numbers: 43.70.Dn, 43.70.Ve, 43.66.Gf, 43.63.Th, 43.80.Lb

## INTRODUCTION

Studies of human adults have demonstrated that discrimination of pairs of stimuli from a /bæ-dæ-gæ/ continuum is discontinuous, with best discrimination occurring in the region of the /bæ-dæ/ and /dæ-gæ/ boundaries (Mattingly *et al.*, 1971; Pisoni, 1973). This phenomenon of enhanced sensitivity at the locations of phonetic boundaries, termed the "phoneme-boundary effect" by Wood (1976), has been replicated for the place feature in two- to three-month-old infants (Eimas, 1974). This suggests that infants are innately predisposed to partition speech-sound continua in ways that are conducive to the phonetic classification of the sounds.

The existence of these phoneme-boundary effects in human infants has not been questioned (Eimas and Tartter, 1979; Kuhl, 1979a; Jusczyk, 1981); however, the extent to which they can be unequivocally attributed to mechanisms that are speech-specific is still unclear (Kuhl, 1978, 1979b). The reason for the uncertainty is the fact that these specific effects are not exclusive to human listeners—comparative studies on human and animal listeners have shown that animals display similar tendencies, at least for the voicing feature (Kuhl, 1981; Kuhl and Padden, 1983).

The purpose of the present experiment was to provide additional comparative data on a second phonetic feature—place of articulation. While discrimination of synthetic stimuli varying in an acoustic cue that is sufficient to indicate place has been tested in comparative experiments (Morse

and Snowden, 1975; Sinnott *et al.*, 1976), the data do not provide sufficient evidence to establish the presence or absence of the phoneme-boundary effect. Morse and Snowden used a heartrate habituation-dishabituation technique to test three pairs of stimuli on a three-formant /bæ-dæ-gæ/ continuum. They reported that both between-category and within-category pairs were discriminated by macaques but that dishabituation was significantly greater for between-category pairs. This is compatible with the idea that the effect exists in animals, but Morse and Snowden rightly argued that the data were equivocal because of the difficulty in equating degree of heartrate dishabituation with degree of discriminability.

Sinnott *et al.* (1976) showed that macaques could learn to discriminate naturally produced /ba/ and /da/ syllables. They also tested discrimination using a synthetic /ba-da/ continuum, but tested pairs in an AX format in which an endpoint stimulus (A) was combined with all other stimuli on the continuum (X). This allowed a comparison of the psychometric functions for humans and animals but did not address the issue of differential discriminability along the continuum.

In the present experiment, macaques were trained on a same-different task and then tested on a variety of pairs drawn from a two-formant /bæ-dæ-gæ/ continuum. These stimuli do not contain the full set of acoustic cues that occur in natural stimuli (Fant, 1973; Stevens and Blumstein, 1981; Kewley-Port, 1982). Namely, they do not contain formants above the third, nor "bursts." It has been suggested that such

cues are essential in providing an acoustic complex sufficient to differentiate stop consonants (Stevens and Blumstein, 1981; Kewley-Port, 1982). For example, Stevens and Blumstein (1978) developed a "template" model which described unique spectral shapes for each place of articulation. The templates were derived from static spectral sections (25.6 ms in duration) taken at the onset of naturally produced syllables. Tests of the model demonstrated that the templates correctly accepted or rejected new syllables that were naturally produced about 85% of the time. However, two-formant stimuli are not correctly classified by these templates. Stevens and Blumstein argue that the isolated second-formant transitions might be "secondary" cues, learned by association with the primary template-specified cues. If so, then macaques who have not been trained to differentiate full-cue synthetic or naturally produced syllables should fail to produce the phoneme-boundary effect for a place continuum that is cued only by second-formant transitions.

This experiment was designed to determine (1) whether animal listeners demonstrated enhanced discriminability at any location on a two-formant place continuum; and (2) if enhanced discriminability did occur, whether it coincided with the locations of phoneme boundaries defined by human adults.

## I. METHOD

### A. Subjects

Three (two male, one female) juvenile monkeys (two *Macaca fuscata*, one *Macaca nemistrina*) were tested. They were between one and three years of age at the onset of training. Each of the animals was housed in an individual cage at the University of Washington's Regional Primate Research Center. They had access to water in their home cages at all times and were fed once daily at the completion of the experimental session. The human listeners were three adults (aged 25–30) who had not had extensive experience listening to synthetically generated speech.

### B. Stimuli

The two-formant /bæ-dæ-gæ/ stimuli were computer synthesized at the Haskins Laboratories in New Haven on the parallel resonance synthesizer according to the parameter specifications described by Mattingly *et al.* (1971). The

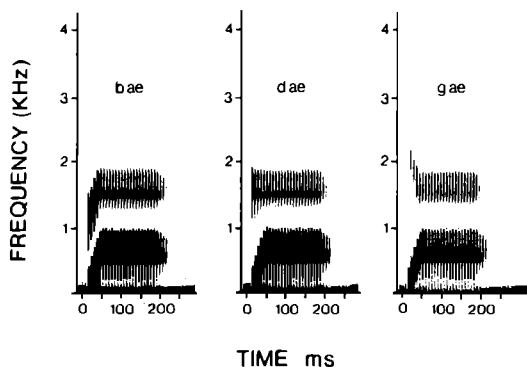


FIG. 1. The first, seventh, and fifteenth members of the two-formant /bæ-dæ-gæ/ continuum used in the experiment. The acoustic cue being manipulated in the series is the starting frequency of the second formant transition.

continuum contained 15 stimuli. Figure 1 displays spectrograms of the first, seventh, and 15th members of the series; they are perceived as /bæ/, /dæ/, and /gæ/, respectively. As shown, the second formant of number 7 is nearly constant; in numbers 1–6 it is rising and in numbers 8–15 it is falling. The stimuli were synthesized with a 15-ms period of closure voicing represented by a low-amplitude first formant centered at 150 Hz, followed by a 40-ms transitional period during which the two formants moved toward the steady-state frequencies appropriate for the vowel with the first two formants located at 740 and 1620 Hz, respectively. The acoustic feature that was varied to generate the continuum was the starting frequency of the second-formant transition. It was varied in approximately equal steps from 1150 to 2230 Hz. The stimuli were 245 ms in duration with a fundamental frequency that was constant at 90 Hz.

Discrimination was tested for seven pairs of stimuli, each separated by two steps on the continuum (pairs 1–3, 3–5, 5–7, 7–9, 9–11, 11–13, 13–15). These pairs coincide with those tested by Mattingly *et al.* (1971) for human adult listeners. Eimas (1974) tested infants on pairs 1–4, 2–5, 3–6, 7–9, and 9–11, and showed significant effects for 2–5, 3–6 (in the /b-d/ boundary region), and 9–11 (in the /d-g/ boundary region).

Neither our human or animal listeners had prior experience in an identification task using these stimuli. Identification functions were obtained on our human listeners after discrimination testing to confirm the boundary locations specified in previous studies. During identification testing, the 15 stimuli were presented singly and the listeners wrote their responses. Ten responses to each stimulus were obtained.

### C. Apparatus

The experiment was conducted in a double-walled, sound-proof IAC booth. During testing, the animals were restrained in primate chairs. Audio signals were delivered through a single earphone (TDH-49 with MX-41/AR cushion) to the animal's right ear. A response key was located directly in front of the chair and a green light was mounted at eye level one foot in front of the animal. A second red light was adjacent to the green light. An automatic feeder delivered 2 cc of applesauce through a rubber tube located near the animal's mouth. A small laboratory computer (Raytheon, 706) controlled the delivery of sound and all of the appropriate contingencies during the experiment. A programmable attenuator (Grasson-Stadler, model 1284) was used to adjust the intensity levels of the signals during the training phase of the experiment. Information concerning each trial was printed on an electronic data terminal (Texas Instruments, model 700) following each trial.

### D. Procedure

A positive-reinforcement procedure was employed. The animal initiated trials by depressing the response key when the green light was blinking. As soon as the animal depressed the key, the light stopped blinking and was on steadily. If the animal held the key for the duration of a

variable-fore-period (VFP), which ranged from 0.01 to 1.2 s, a trial was presented. If the animal released the key before the end of the VFP, a time-out period (TO) occurred, during which the green light was turned off and the red light was turned on for 7 s and key-pressing responses failed to initiate trials. Animals were tested for 1 h each day.

Two kinds of trials, same (S) and different (D), were run with equal probability. During S trials, four identical stimuli were presented at 1-s intervals measured onset to onset (e.g., AAAA). During D trials, the first two stimuli were identical to the stimuli presented during S trials, but the last two stimuli were different (e.g., AABB). In typical same-different formats with human listeners, S trials consist of AA and BB trials, and D trials consist of both AB and BA trials. We have not been able to train our animals to do the latter kind of task with more than a single stimulus pair, and since the design involved the collection of data from each animal on all seven stimulus pairs (i.e., repeated-measures), we chose the format described above. In order to be reinforced, the animal was required to continue to depress the key for the full duration of the S trials (1.7 s timed from the onset of the third stimulus), producing a "correction rejection," and to release the key during the 1.7-s trial interval (also timed from the third stimulus) on D trials, producing a "hit" response. If the animal incorrectly released the key during the 1.7-s trial interval on an S trial producing a "false-positive" response or failed to release the key during the 1.7-s trial interval on a D trial producing a "miss" response, no food reinforcement was delivered and a 7-s TO period occurred. A TO period also occurred if the animal released the response key during the presentation of the first two stimuli on either S or D trials (an "early-release" response). At the completion of each trial the green light was turned off and remained off until the animal released the key for 0.5 s; after this time interval, the light again began to blink indicating to the animal that a trial could be initiated.

Humans were tested in the same sound-proof booth with the same earphone. Applesauce was not delivered for correct responses but the feeder produced an audible noise to provide them with identical feedback concerning the accuracy of their responses.

### E. Trial structure

The seven stimulus pairs were presented in a randomized-block design using repeated measures. Listeners were tested on each stimulus pair for a 1.5-m period (approximately 20 trials when they are initiated steadily); during that time S and D trials for that stimulus pair occurred with equal probability. Each 1.5-m trial block was separated by a 5-s pause. The order of stimulus pairs was randomized within a session. In a typical 40-m session, each stimulus pair was tested three times to provide approximately 60 trials per day per stimulus pair. The data were collected in four sessions.

### F. Preliminary training

The procedures used to train the animals were similar to those described by Sinnott *et al.* (1976). Briefly, the animal was placed in a primate chair each day and trained using

standard shaping procedures to press and release the response key for food reinforcement. The animal was trained to depress the key until a sound (the eventual B stimulus) was presented, and then to release the key for reinforcement. The interval prior to the presentation of the B stimulus (VFP) was slowly lengthened, but continued to be varied from trial to trial to prevent the animal from "timing" his release response rather than listening for the stimulus. When the VFP was approximately 3 s in duration, and the animal consistently held the key down until the stimulus was presented and released the key as soon as the sound was presented, a second stimulus (the A stimulus), attenuated by 50 dB, was introduced prior to the B stimulus. The animal continued to be reinforced for releasing the bar when B was presented, and was given a TO period for releasing to A, as the intensity of A was systematically increased until it equaled the intensity of B. After the animal succeeded at this stage in training, S trials (AAAA), and D trials (AABB), were run with equal probability and all the contingencies previously described were in effect.

The last step in the pretraining period was the block-to-block variation in the stimulus pair being tested. The stimulus pair used during training consisted of a vowel contrast (/a/ vs /i/), and the stimulus pairs used to adapt the animal to the randomized-block design consisted of additional vowel contrasts (/a/ vs /ɔ/) as well as pairs of identical vowels differing in pitch contour (rise versus fall) and syllable pairs differing by the initial consonant (/sa/ vs /ʃ a/; /va/ vs /ʃa/). When performance on these training stimuli was consistently above 80% correct, discrimination testing began. The training period ranged from three to nine months for individual animals.

## II. RESULTS

### A. Human identification and discrimination data

Figure 2 shows the pooled identification data for the human listeners. Percent labeled /bæ/, /dæ/, or /gæ/ is referenced to the right-hand ordinate. The data show that the human listeners perceived three distinct phonetic categories on the continuum. They identified stimuli 1-4 as /bæ/, stimuli 5-11 as /dæ/, and 12-15 as /gæ/.

The discrimination data for each human and animal listener was organized in separate 2×2 stimulus-response matrices like those shown in Table I. As indicated, the condi-

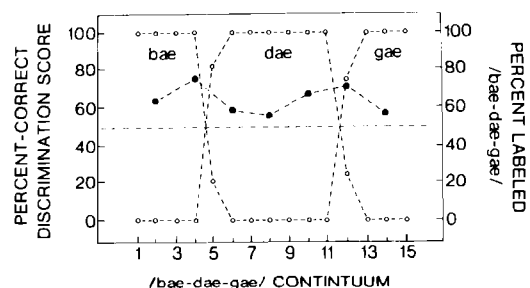


FIG. 2. Pooled identification data (referenced to the right ordinate) and discrimination data (referenced to the left ordinate) for adult human listeners. The discrimination data are plotted at the midpoint for the pair tested (e.g., performance on the 1-3 pair is plotted at stimulus #2).

TABLE I. Stimulus-response matrix computed for each stimulus pair.

		Stimulus	
		Different	Same
Response	Different	Hit P(D/D)	False-positive P(D/S)
	Same	Miss P(S/D)	Correct-rejection P(S/S)

tional probability of a hit is the probability of responding different (releasing the response key) when the stimulus pair was different [P(D/D)]. Similarly, the conditional probability of a false-positive is the probability of responding "different" when the stimulus pair was actually the same [P(D/S)]. Conditional probabilities for "miss" responses [P(S/D)] and "correct rejection" responses [P(S/S)] are simply 1-P(D/D) and 1-P(D/S), respectively. The matrices for each listener were based on an average of 120 trials (six blocks) for each stimulus pair. The absolute number varied from 113-131. This is the case because listeners were allowed to initiate as many trials as possible in each 1-min block and that number varied slightly from block to block.

A number of analyses were conducted using these stimulus-response matrices. The simplest was a percent-correct measure, calculated by adding the probabilities of hits and correct rejections, dividing by two, and multiplying by 100. This measure takes into account the listeners' responses on both S and D trials.

The human discrimination data shown as the percent-correct score for each pair tested is plotted in Fig. 2, referenced to the left-hand ordinate. The data show that discrimination of these two-formant stimuli by untrained listeners is difficult; performance ranged from 57.0% correct to 74.7% correct. Best performance was produced on stimulus pairs 3-5, 9-11, and 11-13. The similarity on the latter two pairs was attributable to individual differences, with some listeners performing well on pair 9-11 and not on pair 11-13, while other listeners demonstrated the reverse. These differences in the locations of peak discriminability were correlated with the locations of the phonetic boundaries for individual listeners.

**B. Animal discrimination data**

Figure 3 shows the pooled percent-correct discrimination data for the three animal listeners. The two dashed verti-

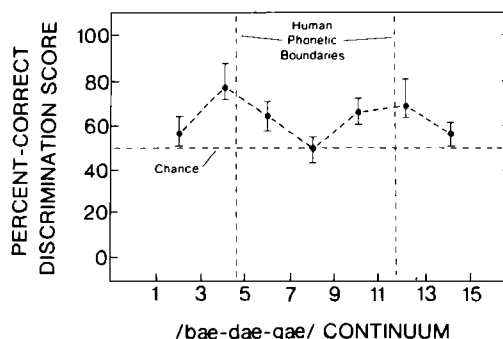


FIG. 3. Average discrimination data for the three animals. The phonetic boundaries for human listeners are marked by dashed vertical lines. The ranges of performance for the three animals are shown by the brackets.

cal lines show the locations of the adult defined boundaries between the /b-d/ and /d-g/ categories. The brackets represent the range of performance for the three animals. Mean percent-correct discrimination scores for the seven stimulus pairs ranged from 55.9% correct to 77.4% correct. Best performance occurred on pairs 3-5, 9-11, and 11-13. As with humans, the location of a peak in performance in the region of the /d-g/ boundary varied with the individuals. For two animals, best performance was produced on stimulus pair 11-13; for the third, best performance was produced on stimulus pair 9-11.

A two-way analysis of variance examined the main effects of groups (human versus animal) and stimulus pairs. The analysis demonstrated a nonsignificant effect for groups ( $F = 1.52; df = 1,4; p < 0.25$ ) and a significant effect for stimulus pair ( $F = 3.04; df = 6,24; p < 0.05$ ). No significant interaction occurred. The results of Newman-Keuls comparisons for differences among stimulus pairs are shown in Table II. The results show that the 3-5 pair differed significantly from all other pairs, and that the 11-13 pair differed significantly from all but 5-7 and 9-11.

To separate potential effects of response bias from those associated with true changes in discriminability in the animal data, two sets of sensitivity/response-bias measures were performed using the data from the 2x2 stimulus-response matrices. The two measures of discriminability were the  $d'$  parameter of signal detection theory (Green and Swets, 1966) which assumes normal distributions and equal variance, and  $-\ln \eta$ , a distribution-free index of discriminability described by Luce (1963). The two measures of response bias were beta ( $\beta$ ) of signal-detection theory (Green

TABLE II. Newman-Keuls comparisons on the differences between stimulus pairs.

Pairs	7-9	1-3	13-15	5-7	9-11	11-13	3-5
7-9				b	b	a	a
1-3						b	a
13-15						b	a
5-7							b
9-11							b
11-13							b
3-5							

<sup>a</sup> $p < 0.01$ .  
<sup>b</sup> $p < 0.05$ .

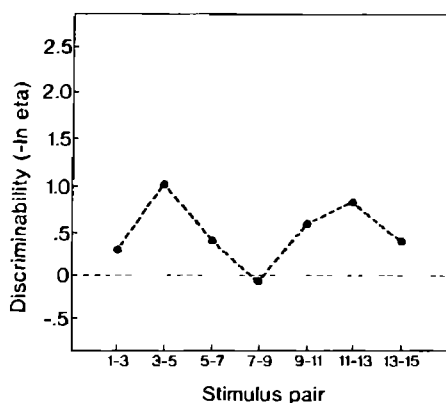


FIG. 4. Group data for animals showing a distribution-free index of discriminability. Higher numbers represent increased discriminability (see text for discussion).

and Swets, 1966), and  $\ln \beta$ , a distribution-free index of response bias. (See MacMillan *et al.*, 1977 for an argument that response bias cannot be completely separated from discriminability, even with these measures.)

The discriminability index  $-\ln \eta$  is described by the formula

$$-\ln \eta = \frac{1}{2} \ln \left[ \frac{P(D/D)P(S/S)}{P(S/D)P(D/S)} \right].$$

Its value is zero at chance and increases with the accuracy of performance. Figure 4 plots  $-\ln \eta$  as a function of the stimulus pair. The  $-\ln \eta$  index shows greater discriminability for the 3-5 and 11-13 pairs, indicating greater sensitivity than for other pairs. The  $d'$  analysis revealed an identical pattern of results.

The response-bias parameter  $\ln \beta$  is described by the formula

$$\ln \beta = \frac{1}{2} \ln \left[ \frac{P(S/S)P(S/D)}{P(D/S)P(D/D)} \right].$$

When there is no response bias,  $\ln \beta$  is equal to zero; it becomes increasingly positive with increasing bias towards S responses (holding the key), and increasingly negative with increasing bias towards D responses (releasing the key). The  $\ln \beta$  index for each stimulus pair is provided in Fig. 5. The measure of response bias of signal-detection theory ( $\beta$ ) produced a pattern of similar results. The data indicate that the animals demonstrated a general tendency toward S responses, regardless of the pair being tested. While hit and correct rejection responses were equally reinforced, this tendency toward holding the key was probably due to the fact that only two of the seven pairs were easily discriminable, plus the fact that half of all trials presented were S trials which require a holding response. The density of reinforcement, therefore, was actually greater for holding responses than for lifting responses. This would tend to cause the animal to refrain from lifting the response key unless the animal was quite sure that the stimulus pair was different.

In addition to the overall tendency toward same responses, the animals demonstrated a slight trend towards lifting the key for /dæ/ stimuli. This trend can be explained by the direction of stimulus change adopted for D-trials in the experiment. To make the task easier, stimuli were arranged such that the B stimulus in an AAB trial was closest

to stimulus #7 (/dæ/). For example, the B stimulus for the 1-3 pair was 3, and for the 11-13 pair it was 11. This stimulus arrangement should result in a slight response bias towards lift responses for B stimuli in the middle of the continuum, but if it were the sole determinant of performance in these animals, one would expect to see a correlated peak in the discriminability index at stimulus #7. Instead, we see two peaks in the discriminability index, neither of which is centered at stimulus #7.

### III. DISCUSSION

This study sought to answer two questions: (1) whether animals demonstrated enhanced discriminability at any locations on a two-formant /b-d-g/ continuum; and (2) if enhanced discriminability did occur, whether it coincided with the locations of phoneme boundaries defined by human adults.

We trained monkeys on a same-different task and then tested them with pairs of stimuli from a physical continuum varying in an acoustic cue that distinguishes the place feature for human adults. While the stimulus pairs were always separated by an equal physical distance on the continuum, the data produced here show that their auditory perceptual differences are not equivalent. Rather, discriminability is relatively poor for within-category pairs of stimuli and relatively good for between-category pairs. This results in two regions of enhanced discriminability on the place continuum, one at each of the two boundary regions separating the three phonetic categories. Thus, animals demonstrate the "phoneme-boundary effect" for a continuum varying only in the starting frequency of the second formant, an acoustic cue sufficient to identify the place feature in human adults.

The data are relevant to three related issues: (1) the acoustic cues for the place feature; (2) arguments concerning the degree to which animal data contribute to our understanding of human data; and (3) arguments concerning the role played by hearing in the evolution of language.

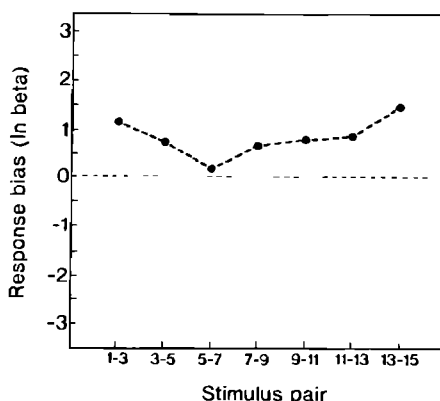


FIG. 5. Group data for animals showing a distribution-free index of response bias. Positive numbers indicate a tendency towards "same" (hold the key) responses, while negative numbers indicate a tendency towards "different" (release the key) responses (see text for discussion).

## A. Acoustic cues for place

The data on human listeners provided here confirm what has been shown in previous studies (Cooper *et al.*, 1952; Delattre *et al.*, 1955; Mattingly *et al.*, 1971). That is, human listeners can identify and discriminate two-formant synthetic stimuli. In order to account for the adult's ability to do so, Stevens and Blumstein (1981) hypothesized that human listeners use secondary cues acquired by an "incidental learning" process whereby a set of partial cues are associated with full cues. However, the animal data pose a problem for this account. Animals have not had experience with naturally produced exemplars of the categories, but still partition the continuum appropriately. The data thus suggest that the "secondary" cues are sufficient in and of themselves, without associative pairing with the "primary" cues, to produce the appropriate partitioning.

Similarly, the account offered by Stevens and Blumstein (1981) predicts that human infants will fail to differentiate two-formant stimuli appropriately until they have had sufficient exposure to stimuli that contain full cues and have learned to depend on a set of partial cues. The data provided by Eimas (1974) and more recently by Walley (1979) show that infants are at least capable of partitioning two-formant stimuli appropriately.

The next question, then, is whether or not there is a difference in performance by adults, infants, or animals when two-formant are used as opposed to more complex stimuli. The fact that the naive human listeners in this study found the two-formant stimuli difficult to discriminate supports Stevens and Blumstein's idea, that these stimuli do not contain all of the cues that specify good examples of the categories. The data on infants are equivocal. Williams and Bush (1978) showed that infants between 6 and 12 weeks, tested using the high-amplitude sucking technique, discriminated a /d-g/ contrast using either the partial-cue or the full-cue stimuli. They argued that the full-cue stimuli were better exemplars for infants because the degree of sucking recovery was greater for the infants presented with the full-cue stimuli. However, the magnitude of sucking recovery has not been shown in previous studies to correlate with the degree of perceptual difference between the two stimuli (see Kuhl, 1979a for review), and the differences in recovery between the two groups in the Williams and Bush study narrowly failed to reach significance, so no firm conclusions could be drawn. Walley (1979) provided a stronger test using partial-cue (two-formant without a burst) versus more complex stimuli (five-formant with a burst). She tested six-month-olds using the operant head-turn technique and found no differences in the degree of discriminability for /b-d-g/ contrasts cued partially or more completely. Perhaps an experimental design such as that employed by Kuhl (1979a), which tests the degree to which infants generalize a head-turning response to novel instances from a category, would reveal differences. It could be the case, for example, that the degree of generalization to novel stimuli would differ depending upon whether or not the infant was initially trained to differentiate "good" as opposed to "poor" exemplars from place categories. Comparative experiments could

be similarly designed to test whether animals treat two-formant and more complex stimuli differently.

## B. Contribution of animal data to human data

These data demonstrate that animal listeners produce the "phoneme-boundary effect." Such effects in human adults and infants have been interpreted as support for the existence of speech-specific mechanisms. Does the fact that the effect can be reproduced in animals alter the interpretation of the human data?

Perhaps it is best to state the most obvious limitation first. It is not the case that animal data rule out explanations of human behavior. Animals may use a simpler set of acoustic cues to guide discrimination and perceptual grouping of stimuli while humans may use a more complex set of acoustic cues, and rules for their combination, to determine discrimination and categorization. The eventual answer to questions related to speech-specific mechanisms for phonetic categorization will undoubtedly not be a simple yes or no. Rather, it will be a determination of the "level" at which speech-specific mechanisms operate. The goal is to push the human-animal analogy to its limits, eventually demonstrating empirically the examples in which the human and animal data diverge. In comparative tests to date the focus has been on the animal's tendency to demonstrate the phoneme-boundary effect using stimuli from a continuum. Given that the initial comparisons have confirmed these specific effects for voicing (Kuhl, 1981; Kuhl and Padden, 1983) and is here demonstrated for place, then comparative tests involving more complex examples can be undertaken.

This approach is best illustrated for the voicing feature. The phoneme-boundary effect for voicing in animals (Kuhl, 1981; Kuhl and Padden, 1983) could be related to the detection of single acoustic parameters. Differential discriminability could be based on either (1) a simultaneity versus nonsimultaneity threshold (Miller *et al.*, 1976; Pisoni, 1977); (2) the presence or absence of either a first-formant transition at voicing onset (Stevens and Klatt, 1974) or the presence or absence of low-frequency energy at voicing onset (Lisker, 1975; Summerfield and Haggard, 1977); or (3) the degree of aspiration (Repp, 1979). The stimuli used to date (described by Abramson and Lisker, 1970) do not allow a differentiation of the use of these cues in isolation from the use of a more complex set of rules for the combination of cues, but experiments on adult listeners have provided some good examples for further tests. For instance, manipulation of the presence or absence of low-frequency energy at voicing onset systematically alters the location of the boundary on a VOT continuum in adults (Summerfield and Haggard, 1977). This effect has recently been demonstrated in human infants (Miller and Eimas, 1981), and will therefore make a good test case for comparative experiments.

This kind of systematic approach to adult, infant, and animal data is made more difficult in the case of the place feature by our lack of experimentation on the acoustic features that alter the locations of phonetic boundaries. Until experiments on adults demonstrate the human listener's rules for the categorization of stimuli varying in the place

feature, the critical comparative experiments cannot be undertaken.

### C. Auditory perceptual constraints and the evolution of language

It has been suggested (Kuhl and Miller, 1975) that man's auditory perceptual system provided a set of "natural psychophysical boundaries" which influenced the selection of candidates for a phonetic repertoire. While this discussion focuses on the potential role of auditory constraints, independent of those provided by the articulatory mechanism, this by no means excludes the potential contribution of articulatory constraints. We emphasize those related to audition simply to invite debate on three theoretical perspectives: (1) audition did not provide a strong selective pressure, independent of articulation, on the choice of a phonetic inventory; (2) audition provided an independent pressure, but one that served to *initially structure* rather than *solely determine* the selection of the inventory, or (3) audition *per se* directed the selection of the inventory by providing a set of "natural classes" for auditory stimuli that the articulatory mechanism evolved to achieve. We take the posture that the first is least plausible given the available data, that the second is consistent with the data now in hand, but that the third should not be ruled out.

The first argues that audition did not play an independent role in the evolution of language. The most powerful data suggesting that this is not the case are those presented here and in other studies in which animals have been shown to appropriately categorize and discriminate speech sounds (Burdick and Miller, 1975; Baru, 1975; Kuhl and Miller, 1975, 1978; Morse and Snowdon, 1975; Waters and Wilson, 1976; Kuhl, 1981; Kuhl and Padden, 1983). Data suggesting that human listeners demonstrate perceptual discontinuities when listening to nonspeech sounds that are similar to those seen for speech (Miller *et al.*, 1976; Pisoni, 1977) have also been interpreted in support of the notion that the phonetic inventory was influenced by auditory constraints. Taken together these data support the claim that audition may have played an independent role in shaping the acoustics of language.

The second posture argues that the auditory system provided a set of broad guidelines which initially structured the selection of phonetic candidates but did not determine them precisely. These guidelines might have taken the form of "natural psychophysical boundaries" (Kuhl and Miller, 1975). These boundaries would have served to separate sounds along a number of auditory dimensions. Examples of such dimensions might include (1) the relative timing of two events with simultaneous and nonsimultaneous events being maximally distinct; (2) rise-time, with rapid as opposed to slow being maximally distinct, and (3) spectral shape, with parameters like diffuse/compact and spectral gravity interacting to produce maximally distinct classes. These properties of sounds could have provided a set of perceptual boundaries whose functional characteristics produced poor discriminability for stimuli falling on either side of a boundary and good discriminability for stimuli straddling a boundary.

This account would not go so far as to suggest that man's auditory system dictated a detailed set of rules for the interactions of acoustic cues in the perception of speech. It would simply stipulate that man capitalized on a set of acoustic properties, but then elaborated on them, perhaps to take into account constraints inherently imposed by the joint actions of the articulators. These articulatory constraints might have resulted in a set of nonequivalent acoustic events that would mandate a mechanism that perceptually equated them—perhaps a mechanism such as that proposed by the original motor theory (Lieberman *et al.*, 1967). The data to date on the perception of speech by animals do not contradict this general explanation.

The third account argues for a deterministic role for audition in the evolution of language. It predicts an even closer correspondence between the human and animal data, a correspondence not as yet warranted by the data because the relevant experiments have not been undertaken. In its broadest form, this posture advocates the view that speech sounds form "natural classes." That is to say, they represent an optimum combination of acoustic cues. The natural class theory argues that the perceptual equivalence of different acoustic cues in speech are not the result of a special mechanism that recognizes their association in articulation. Rather, it argues that the "trading relations" seen in studies of adult speech perception (e.g., Best *et al.*, 1981) occur because the cues produce equivalent auditory effects when measured in terms of the degree to which the entire signal approaches the specification of a particular prototype. The strong version of this account predicts, therefore, that perceptual effects as complex to explain as "trading relations" (Best *et al.*, 1981) would obtain in animals.

### IV. SUMMARY

This study demonstrated differential discriminability of sounds along a two-formant place of articulation continuum in a nonhuman primate. The data demonstrated that animals partition the continuum into categories consistent with the phonetic identity of the sounds. This suggests that a place continuum cued only by the second-formant transition can be appropriately partitioned in the absence of exposure to a more complex set of cues. The data also support the notion that auditory constraints could have provided a selective pressure in the absence of articulatory ones to provide an initial structuring of the acoustics of language. It remains to be determined whether audition exerted an even stronger influence by providing the rules for the formation of "natural classes" of sounds to which speech conformed. Experiments demonstrating the existence of "trading relations" for speech cues in animals would provide support for this latter interpretation.

### ACKNOWLEDGMENTS

This research was supported by a grant to P. K. Kuhl from the National Science Foundation (BNS 8022492). We wish to thank K. L. Morgan-Landan and M. Kulick for assistance in running animals during the training phase of the experiment. We gratefully acknowledge J. M. Miller, De-

partment of Otolaryngology, for donating space for the conduct of these experiments, the cooperation of the Haskins Laboratories (NIH contract NIH-71-2420) in providing the synthetic stimuli and the Regional Primate Research Center (NIH Grant RR 00166) for the use of core facilities and the help of core staff throughout the conduct of this research. The author thanks A. N. Meltzoff for critical comments on an earlier version of this manuscript.

- Abramson, A., and Lisker, L. (1970). "Discriminability along the voicing continuum: cross-language tests," in *Proceedings of the 6th International Congress of Phonetic Science, Prague 1967* (Academia, Prague), pp. 569-573.
- Baru, A. V. (1975). "Discrimination of synthesized vowels /aa/ and /i/ with varying parameters in dog," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, London).
- Best, C. T., Morrongoello, B., and Robson, R. (1981). "Perceptual equivalence of acoustic cues in speech and nonspeech perception," *Percept. Psychophys.* **29**, 191-211.
- Burdick, C. K., and Miller, J. D. (1975). "Speech perception by the chinchilla: Discrimination of sustained /a/ and /i/," *J. Acoust. Soc. Am.* **58**, 415-427.
- Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J. (1952). "Some experiments on the perception of synthetic speech sounds," *J. Acoust. Soc. Am.* **24**, 597-606.
- Delattre, P. C., Liberman, A. M., and Cooper, F. S. (1955). "Acoustic loci and transitional cues for consonants," *J. Acoust. Soc. Am.* **27**, 769-773.
- Eimas, P. D. (1974). "Auditory and linguistic processing of cues for place of articulation by infants," *Percept. Psychophys.* **16**, 513-521.
- Eimas, P. D., and Tartter, V. C. (1979). "On the development of speech perception: Mechanisms and analogies," in *Advances in Child Development and Behavior. Vol. 13*, edited by H. W. Reese and L. P. Lipsitt (Academic, New York).
- Fant, G. (1973). *Speech Sounds and Features* (MIT, Cambridge).
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Wiley, New York).
- Jusczyk, P. W. (1981). "Infant speech perception: A critical appraisal," in *Perspectives on the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Erlbaum, Hillsdale, NJ).
- Kewley-Port, D. (1982). "Measurement of formant transitions in naturally produced stop consonant-vowel syllables," *J. Acoust. Soc. Am.* **72**, 379-389.
- Kuhl, P. K. (1978). "Predispositions for the perception of speech-sound categories: A species-specific phenomenon?" in *Communicative and Cognitive Abilities—Early Behavioral Assessment*, edited by F. D. Minifie and L. L. Lloyd (University Park, Baltimore).
- Kuhl, P. K. (1979a). "Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories," *J. Acoust. Soc. Am.* **66**, 1668-1679.
- Kuhl, P. K. (1979b). "Models and mechanisms in speech perception: Species comparisons provide further contributions," *Brain Behav. Evol.* **16**, 374-408.
- Kuhl, P. K. (1981). "Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories," *J. Acoust. Soc. Am.* **70**, 340-349.
- Kuhl, P. K., and Miller, J. D. (1975). "Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants," *Science* **190**, 69-72.
- Kuhl, P. K., and Miller, J. D. (1978). "Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli," *J. Acoust. Soc. Am.* **63**, 905-917.
- Kuhl, P. K., and Padden, D. M. (1983). "Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques," *Percept. Psychophys.*
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). "Perception of the Speech Code," *Psychol. Rev.* **74**, 431-461.
- Lisker, L. (1975). "Is it VOT or a first-formant transition detector?" *J. Acoust. Soc. Am.* **57**, 1547-1551.
- Luce, R. D. (1963). "Detection and recognition," in *Handbook of Mathematical Psychology*, edited by R. D. Luce, R. R. Bush, and E. Galanter (Wiley, New York).
- Macmillan, N. A., Kaplan, H. L., and Creelman, C. D. (1977). "The psychophysics of categorical perception," *Psychol. Rev.* **84**, 452-471.
- Mattngly, I. G., Liberman, A. M., Syrdal, A. K., and Halwes, T. (1971). "Discrimination in speech and nonspeech modes," *Cognitive Psychol.* **2**, 131-157.
- Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., and Dooling, R. J. (1976). "Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception," *J. Acoust. Soc. Am.* **60**, 410-417.
- Miller, J. L., and Eimas, P. D. (1981). "Contextual perception of voicing by infants," Paper presented at the Biennial Meeting of the Society for Research in Child Development, Boston, 1981.
- Morse, P. A., and Snowdon, C. T. (1975). "An investigation of categorical speech discrimination by rhesus monkeys," *Percept. Psychophys.* **17**, 9-16.
- Pisoni, D. B. (1973). "Auditory and phonetic memory codes in the discrimination of consonants and vowels," *Percept. Psychophys.* **13**, 253-260.
- Pisoni, D. B. (1977). "Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops," *J. Acoust. Soc. Am.* **61**, 1352-1361.
- Repp, B. H. (1979). "Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants," *Lang. Speech* **22**, 173-189.
- Sinnott, J. M., Beecher, M. D., Moody, D. B., and Stebbins, W. C. (1976). "Speech sound discrimination by monkeys and humans," *J. Acoust. Soc. Am.* **60**, 687-695.
- Stevens, K. N., and Blumstein, S. E. (1978). "Invariant cues for place of articulation in stop consonants," *J. Acoust. Soc. Am.* **64**, 1358-1368.
- Stevens, K. N., and Blumstein, S. E. (1981). "The search for invariant acoustic correlates of phonetic features," in *Perspectives On the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Erlbaum, Hillsdale, NJ).
- Stevens, K. N., and Klatt, D. H. (1974). "Role of formant transitions in the voiced-voiceless distinction of stops," *J. Acoust. Soc. Am.* **55**, 653-659.
- Summerfield, Q., and Haggard, M. (1977). "On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants," *J. Acoust. Soc. Am.* **62**, 435-448.
- Walley, A. (1979). "Infants' discrimination of full and partial cues to place of articulation in stop consonants," in *Research On Speech Perception*, Prog. Rep. **5**, Indiana University, 85-145.
- Waters, R. S., and Wilson, W. A., Jr. (1976). "Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants," *Percept. Psychophys.* **19**, 285-289.
- Williams, L., and Bush, M. (1978). "Discrimination by young infants of voiced stop consonants with and without release bursts," *J. Acoust. Soc. Am.* **63**, 1223-1226.
- Wood, C. C. (1976). "Discriminability, response bias, and phoneme categories in discrimination of voice onset time," *J. Acoust. Soc. Am.* **60**, 1381-1389.