

Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques

PATRICIA K. KUHL and DENISE M. PADDEN
University of Washington, Seattle, Washington

(Manuscript received 10/15/81; revised manuscript received 1/15/82; accepted for publication 2/15/82)

Two groups of macaques (*Macaca mulatta*) were trained to discriminate between two stimuli that differed in the voicing of the final consonant of a word. The stimuli were /b/ and /p/ in the context of the word 'bat'. The macaques were trained to discriminate between the two stimuli by responding to the stimuli with a specific behavior. The results showed that the macaques were able to discriminate between the two stimuli at the phonetic boundaries, indicating that they have a phonetic representation of the voicing feature.

The present study was designed to test the hypothesis that macaques have a phonetic representation of the voicing feature. The stimuli were /b/ and /p/ in the context of the word 'bat'. The macaques were trained to discriminate between the two stimuli by responding to the stimuli with a specific behavior. The results showed that the macaques were able to discriminate between the two stimuli at the phonetic boundaries, indicating that they have a phonetic representation of the voicing feature.

The results of the present study indicate that macaques have a phonetic representation of the voicing feature. The macaques were able to discriminate between the two stimuli at the phonetic boundaries, indicating that they have a phonetic representation of the voicing feature.

Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques

PATRICIA K. KUHL and DENISE M. PADDEN
University of Washington, Seattle, Washington

Discrimination of speech sounds from three computer-generated continua that ranged from voiced to voiceless syllables (/ba-pa/, /da-ta/, and /ga-ka/) was tested with three macaques. The stimuli on each continuum varied in voice-onset time (VOT). Pairs of stimuli that were equally different in VOT were chosen such that they were either within-category pairs (syllables given the same phonetic label by human listeners) or between-category pairs (syllables given different phonetic labels by human listeners). Results demonstrated that discrimination performance was always best for between-category pairs of stimuli, thus replicating the "phoneme boundary effect" seen in adult listeners and in human infants as young as 1 month of age. The findings are discussed in terms of their specific impact on accounts of voicing perception in human listeners and in terms of their impact on discussions of the evolution of language.

Studies of human infants as young as 1 month of age have demonstrated that their discrimination of sounds from a continuum ranging from one phonetic unit to another (e.g., from /ba/ to /pa/) is typically enhanced the region of the adult-defined "boundary" between categories and is relatively poor within categories (Eimas, 1974a; Eimas, 1975; Eimas, Siqueland, Jusczyk, & Vigorito, 1971). This differential discriminability is conducive to the phonetic categorization of speech sounds and has been interpreted as supporting the idea that innate factors play an important role in the development of speech perception (Eimas, 1974b; Eimas & Tartter, 1979; Jusczyk, 1981; Kuhl, 1979a; Morse, 1974).

Although a number of authors agree that the infant's abilities are probably innately determined, the precise nature and origins of the infant's predispositions are not known. It has been suggested that these effects reflect the operation of a mechanism specifically designed to detect the acoustic properties of speech sounds (Eimas, 1974b). The possibility has also been raised, however, that these specific "phoneme-boundary effects" might

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, Department 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 authors thank A. N. Meltzoff for critical comments on an earlier version of this manuscript. The authors are affiliated with the Department of Speech and Hearing Sciences and the Regional Primate Research Center, University of Washington, Seattle, Washington 98195.

be attributable to the infant's more general auditory perceptual abilities (Kuhl, 1978, 1979b).

Experiments examining the nature and origins of these effects have centered on tests of their specificity—to speech as opposed to nonspeech (Miller, Wier, Pastore, Kelly, & Dooling, 1976; Pisoni, 1977), and to humans as opposed to animals (Kuhl, 1981; Kuhl & Miller, 1975; Morse & Snowdon, 1975; Sinnott, Beecher, Moody, & Stebbins, 1976; Waters & Wilson, 1976). A particular advantage of comparative experiments is that if animals demonstrate the phoneme-boundary effect, one can argue convincingly that the effect does not necessitate speech-specific mechanisms. This in turn reduces our need, in the absence of other data, to argue that such mechanisms exist. Moreover, animal data contribute to discussions examining the potential role of auditory constraints in the evolution of language. Demonstrations of the phoneme-boundary effect in non-human species raise the possibility that the phonetic inventory reflects certain psychoacoustic constraints (Kuhl, 1979b; Kuhl & Miller, 1975; Stevens, 1972).

There are five published experiments that compare the perception of speech sounds by human and animal listeners. Three of the five studies (Kuhl & Miller, 1975, 1978; Sinnott et al., 1976; Waters & Wilson, 1976) were aimed at determining whether the location of a perceptual boundary for animal listeners coincided with the location of the phonetic boundary in human listeners. These studies did not directly assess discriminability along the continuum, but provided support for the notion that certain nonhuman species perceptually partition speech continua in the phonetic boundary region as defined by human listeners (see Kuhl, 1979b, for a review).

For example, Kuhl and Miller (1975) obtained identification functions for humans and chinchillas using three stimulus sets, a bilabial (/ba/ to /pa/), an alveolar

(/da/ to /ta/), and a velar (/ga/ to /ka/) continuum. The stimuli on each continuum varied in voice-onset time (VOT), the time in milliseconds between the release of the articulatory constriction and the onset of laryngeal voicing. The animal subjects were trained in a conditioned-avoidance paradigm. They learned to respond differentially to good synthetic exemplars of /da/ (0 msec VOT) and /ta/ (+80 msec VOT), taken from the alveolar continuum. When performance on these endpoint stimuli was consistently above 95% correct, the stimuli between 0 VOT and +80 VOT (in 10-msec steps) were presented as generalization stimuli. The results demonstrated that the perceptual boundary for animal and human subjects did not differ significantly. In addition, studies using the other two stimulus sets showed that the exact location of the perceptual boundary depended upon the place of articulation (bilabial, alveolar, or velar) of the voiced-voiceless pair, just as it does for human listeners (Kuhl & Miller, 1978).

Two studies directly addressed discriminability of stimulus pairs from speech-sound continua with animals. Morse and Snowdon (1975) used a heart rate procedure to examine rhesus monkeys' discrimination of stimuli from a synthetic /bae-dae-gae/ continuum. The procedure involved 20 presentations of one stimulus followed by 20 presentations of a second stimulus. Heart rate typically habituates to the first stimulus and dishabituates when the stimulus is changed. The results demonstrated that the rhesus monkeys discriminated both the within-category and between-category pairs. Both groups demonstrated significantly more dishabituation than did the control group, which was presented with a single stimulus repeated 40 times. However, the degree of dishabituation was significantly greater for those animals presented with between-category comparisons—that is, with stimuli falling on opposite sides of the human boundary—than for those subjects presented with within-category pairs, thus suggesting that the phoneme boundary effect may exist in animals.

The second experiment (Kuhl, 1981) assessed differential discriminability along a speech continuum ranging from /da/ to /ta/ in the chinchilla using procedures typically associated with acuity studies in psychophysics. This study provided a direct estimate of the just noticeable difference in VOT (Δ VOT) that could be detected by the animal at various VOTs along the continuum. The rationale for such an experiment derived from the argument that sensitivity to a stimulus change should be maximum at the location of the phonetic boundary and minimum near the center of the stimulus category. The results of the study showed that the smallest Δ VOT values, indicating greatest sensitivity, occurred at the VOT values nearest the phonetic boundary, whereas Δ VOT was maximum at points most distant from the boundary region.

The present experiment was designed to extend these discrimination data to different voiced-voiceless

continua (bilabial, alveolar, and velar series). The testing technique was one that has been used in experiments on adult listeners (e.g., Wood, 1976), a same-different discrimination technique.

METHOD

Subjects

Three (two male, one female) juvenile Japanese monkeys (*Macaca fuscata*) served as subjects. They were between 1 and 3 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.

Stimuli

The voiced and voiceless stop consonants were computer synthesized at the Haskin's Laboratories (New Haven, Connecticut) on the parallel resonance synthesizer according to the parameter specifications described by Abramson and Lisker (1970) for bilabial (/ba-pa/), alveolar (/da-ta/), and velar (/ga-ka/) contrasts.

Naturally produced voiced and voiceless stop consonants are distinguished by the timing of the onset of laryngeal vibration (voicing) relative to the release of the constriction in the supralaryngeal musculature. In voiced stops (/ba/, /da/, /ga/), the onset of voicing precedes the release of the articulatory constriction by some to 5 to 40 msec, whereas in voiceless stops, the release of the articulatory constriction precedes the onset of voicing by more than 25-40 msec. The precise difference in the timing of these two articulatory events depends upon the place of articulation of the stop consonant (Lisker & Abramson, 1964).

There are a number of acoustic features that differentiate synthetically generated voiced and voiceless syllables; the timing difference described above is evidenced in the relative onsets of periodicity (voicing) and the burst of energy that occurs when the articulatory constriction is released. Other acoustic factors, such as the presence or absence of low-frequency energy in the first formant at the onset of voicing (Liberman, Delattre, & Cooper, 1958; Lisker, 1975; Stevens & Klatt, 1974; Summerfield & Haggard, 1977), the presence or absence of aspiration noise, and the aperiodic signal that results when air rushes through the open glottis (Fisher-Jorgenson, 1954), as well as its intensity (Repp, 1979), are also known to be important acoustic cues for the perception of the voicing feature.

In the synthetically generated stimuli created by Abramson and Lisker (1970) and used in this as well as in many other speech experiments reported in the literature, these cues covary. While the stimuli are said to vary in VOT, a name that emphasizes the timing cue, the additional acoustic factors just mentioned also vary in the stimuli.

Abramson and Lisker's (1970) data showed that when adult listeners are asked to label synthetic stimuli that vary in VOT along a continuum ranging from voiced to voiceless stimuli, the locations of the phonetic boundaries (the 50% point on the identification function) are at approximately +22 msec VOT for the bilabial stimuli, +35 msec VOT for the alveolar stimuli, and +42 msec VOT for the velar stimuli. Using standard identification procedures, we verified that adults tested in our situation produced similar results.

In this study, three pairs of stimuli were tested from each continuum, two within-category pairs (one voiced and one voiceless) and one between-category pair. Each stimulus pair differed by 20 msec VOT. The between-category pair was chosen such that it straddled the phonetic boundary on each continuum, and the two within-category pairs were chosen to be immediately adjacent on both sides. Table 1 lists the nine pairs tested. These values were chosen to coincide with those used by investigators in infant studies (Eilers, Gavin, & Wilson, 1979;

Table 1
The Nine Stimulus Pairs Tested in the Experiment

	Bilabial	Alveolar	Velar
	VOT	VOT	VOT
Within Voiced	0 vs. +20	+5 vs. +25	+10 vs. +30
Between Voiced-Voiceless	+20 vs. +40	+25 vs. +45	+30 vs. +50
Within Voiceless	+40 vs. +60	+45 vs. +65	+50 vs. +70

Note—All VOT in milliseconds.

Eimas et al., 1971; Lasky, Syrdal-Lasky, & Klein, 1975; Streeter, 1976). Spectrograms of stimuli from these three series have been published (Kuhl & Miller, 1978) along with more complete acoustic analyses of the signals. Each syllable was 434 msec in duration and had a fundamental frequency that was constant at 114 Hz until the last 100 msec, during which the fundamental fell to 70 Hz.

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 by a two-channel tape deck (TEAC, Model A-2300S) 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 1 ft in front of the animal. A red light was adjacent to the green light. An automatic feeder under computer control 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 (Grason-Stadler, Model 1284) was used to adjust the intensity levels of the signals during the training phase of the experiment. After training, all stimuli were presented at a constant level (65 dB SPL). Information concerning each trial was printed on an electronic data terminal (Texas Instruments, Model 700).

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 foreperiod (VFP), which ranged from .01 to 1.2 sec, 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 sec and keypressing 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-sec 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).¹ 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 sec timed from the onset of the third stimulus), producing a "correct rejection," and to release the key during the 1.7-sec 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-sec trial interval on an S trial and thus produced a "false-positive" response or failed to release the key during the 1.7-sec trial interval on a D trial and thus produced a "miss" response, no food reinforcement was delivered and a 7-sec 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 kept off until the animal released the key for .5 sec; after this time interval, the light again began to blink, indicating to the animal that a trial could be initiated.

Trial Structure

The nine stimulus pairs were presented in a randomized-block design using repeated measures. The animals were tested on each stimulus pair in random order for a 1.5-min period (approximately 20 trials when the animal was working steadily); during that time, S trials (AA pairs) and D trials (AB pairs) for that stimulus pair occurred with equal probability. Each 1.5-min trial block was separated by a 5-sec pause. In a typical 50-min session, each stimulus pair was tested three times to provide approximately 60 trials per day per stimulus pair.

Preliminary Training

The basic 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 gradually 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 sec in duration and the animal consistently held the key down until the stimulus had been presented and released the key as soon as the sound had been 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 had succeeded at this stage in training, S trials (AAAA) and D trials (AABB) were run with equal probability and with all the contingencies previously described in effect.

The last step remaining 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/), pairs of identical vowels differing in pitch contour (rise vs. fall), and syllable pairs differing in the initial consonant (/sa/ vs. /fa/; /va/ vs. /sa/). When performance on these training stimuli was consistently above 80% correct, discrimination testing began. The training period ranged from 3 to 9 months for individual animals.

RESULTS

The discrimination data for each animal were organized in separate 2 X 2 stimulus-response matrices like

Table 2
Stimulus-Response Matrix Computed for Each Stimulus Pair

Response	Stimulus	
	Different	Same
Different	Hit P(D/D)	False-Positive P(D/S)
Same	Miss P(S/D)	Correct-Rejection P(S/S)

those shown in Table 2. As indicated, the conditional probability of a "hit" is the probability of responding "different," that is, of releasing the response key, when the members of the stimulus pair were different [P(D/D)]. Similarly, the conditional probability of a "false-positive" is the probability of responding "different" when the members of the stimulus pair were 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 animal were based on approximately 120 trials (six blocks) for each stimulus pair. Only data from the first two sessions for each animal were used so we could assess performance in the absence of training; studies on adults have shown that protracted training with feedback can produce ceiling effects that obscure potential peaks in discriminability (e.g., Carney, Widen, & Viemeister, 1977).

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. Thus, a score of 50% correct represents chance. This measure takes into account the animals' responses on both S and D trials. The mean percent-correct scores are plotted in Figure 1.

As Figure 1 shows, the animals performed better, for all three test series, on the between-category pair than on either of the two within-category pairs. This group trend was shown for each of the three individual animals. In no instance did an animal perform better on the within-category contrasts than on the between-category contrast for a given test series. A three-way ANOVA

examining the main effects of stimulus contrast (within vs. between), place of articulation (bilabial, alveolar, or velar), and trial block (1-6) revealed significant effects for both stimulus contrast [$F(1,2) = 35.5$, $p < .03$] and place of articulation [$F(2,4) = 10.8$, $p < .02$] but not for trial block [$F(5,10) = 1.5$, $p < .25$]. Neither the two-way interactions nor the three-way interaction approached significance ($p < .20$ in all instances).

Examination of the percent-correct scores also revealed that performance on the within-category pairs tended to differ, with performance on the within-voiceless pair exceeding performance on the within-voiced pair for the bilabial and alveolar contrasts. However, this effect was shown to be due to a change in response bias rather than to a true difference in discriminability for within-category pairs. Recall that the animals were trained to discriminate S (AAAA) trials from D (AABB) trials and that the higher VOT value in a given pair served as the B stimulus.¹ Because of this, all animals tended to release the key more frequently in response to pairs containing a stimulus with a higher VOT value, even on an S trial. This could be seen in a variety of results: first, the animals tended to produce a greater number of "hit" responses to pairs of stimuli with high VOT values, but also greater numbers of "false-positive" responses. Second, animals tended to produce greater numbers of "early-release" responses (releasing before the end of the first two stimulus presentations and therefore before the actual start of the trial) when the A stimulus had a higher VOT value.

To separate potential effects of response bias from those associated with true changes in discriminability, two sets of discriminability/response-bias measures were calculated using the data from the 2×2 stimulus-response matrices. The two measures of discriminability were the d' parameter of signal-detection theory (Green & 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 the β of signal-detection theory (Green & Swets, 1966) and $\ln b$, a distribution-free index of response bias.

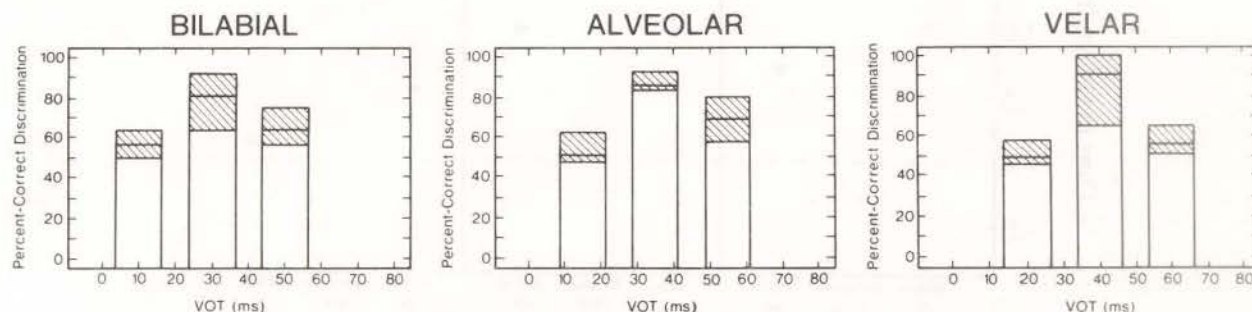


Figure 1. Average percent-correct scores—[(probability of hit + correct-rejection responses)/2 × 100]—for the nine stimulus pairs tested in the experiment. The hatched area shows the range of performance obtained for each pair.

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

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

Its value is zero at chance and increases with the accuracy of performance. Figure 2 plots $-\ln \eta$ as a function of the stimulus pair. For each place of articulation, the $-\ln \eta$ index is greater for the between-category pair than it is for either of the within-category pairs, indicating greater sensitivity. The d' analysis revealed an identical pattern of results. This increase in discriminability for between-category pairs is similar to that shown by Wood (1976), who tested human listeners on pairs of stimuli differing by 20 msec on a bilabial VOT continuum and reported his data in terms of the $-\ln \eta$ index of discriminability. Wood's data showed comparable within-category discriminability, but slightly greater between-category discriminability, when compared with the data obtained here.

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

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

When there is no response bias, $\ln b$ is equal to zero; it becomes increasingly positive with increasing bias toward S responses (holding the key) and increasingly negative with increasing bias toward D responses (releasing the key). The $\ln b$ index for each stimulus pair is provided in Figure 3. The measure of response bias (β) of signal-detection theory produced a pattern of similar results. The data indicate that the animals dem-

onstrated 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 three of the nine pairs were easily discriminable, plus the fact that half of all trials presented were S trials, which required a holding response. The density of reinforcement, therefore, was actually greater for holding responses than for lifting responses. This would tend to cause animals to refrain from lifting the response key unless they were quite sure that the members of the stimulus pair were different.

In addition to the overall tendency toward "same" responses, the animals demonstrated a systematic change in response bias with increasing VOT. Since lifting the key was associated with reinforcement more frequently in the presence of signals with higher VOT values, animals tended to release the key more frequently when the stimulus pair contained a stimulus with a higher VOT value. Recall, however, that only for the between-category pairs was this tendency associated with greater discriminability. The improved discriminability for within-voiceless pairs seen in the percent-correct measure can therefore be attributed to response bias rather than to a true increase in discriminability. In contrast, the peak in discriminability for pairs straddling the boundary represents a true increase in discriminability.

These response-bias data can be compared with those obtained on human listeners by Wood (1976), who also used the $\ln b$ index. He found a significant shift toward "same" responses for within-category contrasts and a significant shift toward "different" responses for

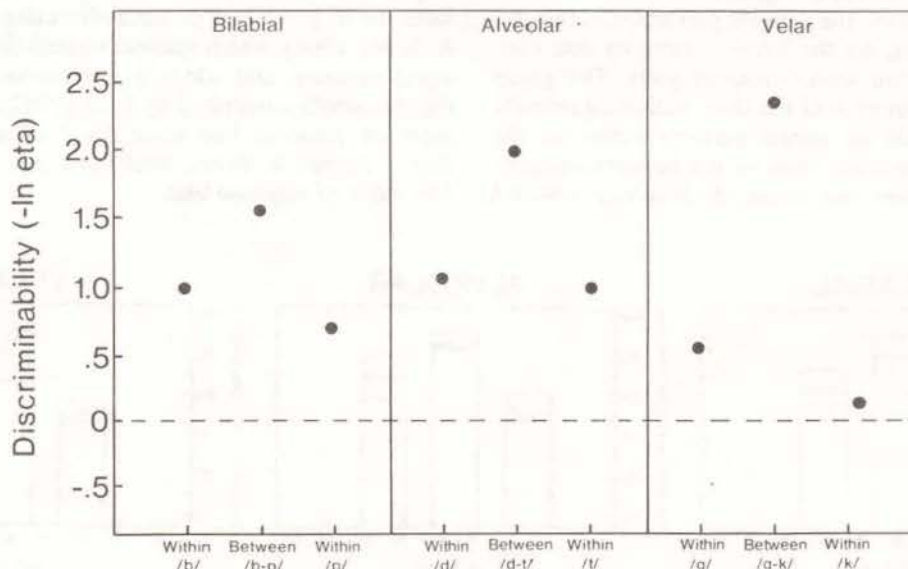


Figure 2. Average data for a distribution-free index of discriminability ($-\ln \eta$) for the nine stimulus pairs. Higher numbers indicate greater sensitivity (see text for additional details).

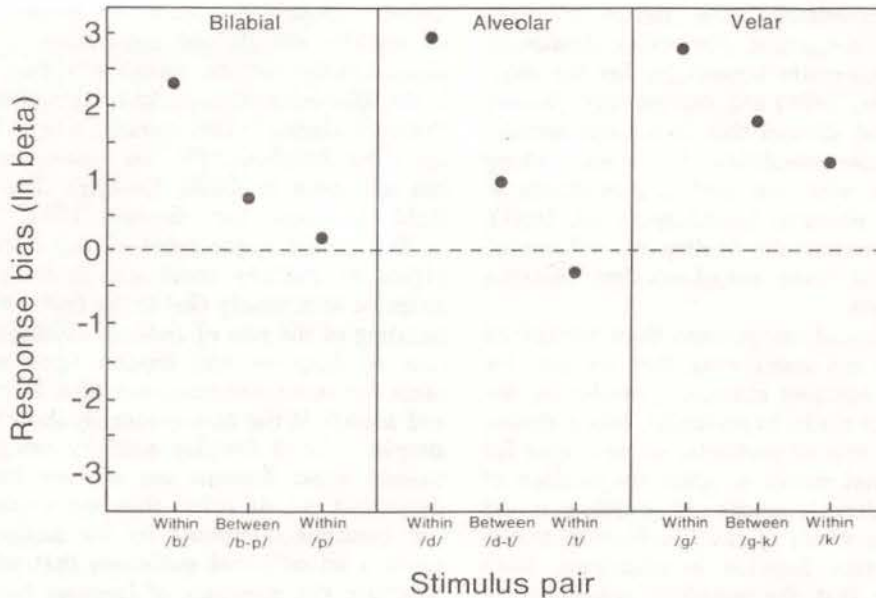


Figure 3. Average data for a distribution-free index of response bias ($\ln b$) for the nine stimulus pairs. Positive numbers indicate a bias toward "same" responses (holding the key), and negative numbers indicate a bias toward "different" responses (releasing the key).

between-category contrasts. We obtained a similar shift in response bias from "same" responses to "different" responses when comparing the voiced within-category pairs with the between-category pairs (Figure 3), but not for the voiceless within-category pairs. We attribute this difference to our specific trial structure.¹

DISCUSSION

In the present study, we trained monkeys to respond on a same-different task and then tested them with pairs of stimuli from a physical continuum that ranged perceptually from voiced to voiceless sounds. The pairs of stimuli were chosen such that they were separated by an equal physical difference in VOT on each of the three continua tested and such that some were perceived to be phonetically identical by adult human listeners while others were perceived to be phonetically different. Measures of discriminability demonstrated that monkeys discriminated sounds that were phonetically different (i.e., straddled the phonetic boundary) significantly better than they discriminated sounds that were phonetically identical (i.e., fell on one side of the boundary). This was true for all three speech continua studied.

The fact that animal listeners demonstrate relatively good discriminability at the boundaries between phonetic categories and relatively poor discriminability within categories, just as human adults and infants do, demonstrates that the phoneme-boundary effect is not exclusive to human listeners. The data raise two important theoretical issues: (1) the relevance of animal data to the interpretation of human data, both adult and

infant, and (2) the role played by auditory constraints in the evolution of language.

Regarding the first issue, the relevance of animal data to interpretations of human adult and infant data, we argue that systematic comparisons among adult, infant, and animal studies will aid in developing strong theories concerning the nature and origins of the mechanisms underlying phonetic perception. Comparisons between human adults and infants demonstrate the degree to which the infant demonstrates an initial capacity to partition an acoustic continuum in a phonetically appropriate way. Comparisons between humans and animals suggest the degree to which effects should be attributed to general auditory perceptual mechanisms rather than to mechanisms evolved specifically for processing speech information. The issue of whether phonetic perception involves mechanisms that are speech-specific will not be resolved with a single comparison. And, given that the initial comparisons between humans and animals have revealed many striking similarities (Kuhl, 1981; Kuhl & Miller, 1975, 1978; Morse & Snowdon, 1975; Waters & Wilson, 1976), as well as some differences (Sinnott et al., 1976), the answer to the speech-specificity issue will not be a simple yes or no. Rather, it will be a determination of the level at which special mechanisms must be invoked to account for the data.

The comparisons of interest form a hierarchy. To date we have examined whether animals tested in labeling tasks perceptually partition speech continua at the phonetic boundaries and whether any peaks in discriminability are consistent with the loca-

tions of phonetic boundaries. The results of these studies on the voiced-voiceless distinction confirmed the existence of appropriate boundaries for the chinchilla (Kuhl & Miller, 1978) and the monkey (Waters & Wilson, 1976), and showed that chinchillas demonstrate differential discriminability for stimuli along a /da-ta/ continuum, with best performance shown at the location of the phonetic boundary (Kuhl, 1981). The present data extend the finding of differential discriminability to all three voiced-voiceless continua in a nonhuman primate.

Given that these initial comparisons show similarities between the human and animal data, then comparative tests involving more complex examples provided by the adult literature can be made. In particular, future studies should examine the role of particular acoustic cues for voicing that have been shown to affect the location of the phonetic boundary in adults. An example is the determination of the effect of the first-formant transition on the boundary location in macaques, since studies have shown that the boundary systematically shifts as the frequency of the first formant is lowered in adult listeners (Lisker, 1975; Summerfield & Haggard, 1977). Recently, experiments have been undertaken with infants that address the potential interaction between these two kinds of acoustic cues (first-formant duration and VOT) for voicing perception (Miller & Eimas, Note 1). The results suggested that the infant's discrimination of speech sounds was influenced by both acoustic cues, as it is for adults. It will now be important to determine if animals show similar effects.

The continued comparison of adult, infant, and animal data using the same stimuli and comparable methods should eventually identify the precise examples for which mechanisms specific to speech must be invoked to account for the data, and the extent to which those mechanisms are functional at birth. Adult experiments that isolate the role of individual acoustic cues and specify the extent to which they govern the boundary locations, alone or in combination, will be helpful. Until the set of rules for combining the cues for voicing perception are determined and the experiments that are definitive tests for the use of those rules are identified, the most powerful comparative and developmental experiments cannot be run.

As definitive examples are tested, we will know exactly how far we can push the argument that the adult and animal data are comparable. It is possible that a full account will suggest that animals use simpler rules for perceptually grouping stimuli, separating them on the basis of an acoustic principle such as, for the voicing contrast, the relative timing of two acoustic events, whereas human listeners employ a more complex set of rules. More complex rules might involve taking into account the values of other acoustic cues. It is also possible, however, that effects as complex as the recently observed "trading relations" (Best et al., 1981)

derive from general rules about the perceptual grouping of auditory stimuli, and are inherent in the functional characteristics of the auditory system. Pushed to its limits, this latter account holds that speech sounds form "natural classes." This notion, which has been developed by Rosch (1973) for certain visual categories, has also been modified for application to speech (see Kuhl, *in press*, and Stevens, 1981, for discussion).

The second major point of this discussion, the role played by auditory constraints in the evolution of language, is intrinsically tied to the first. That is, our understanding of the role of auditory constraints in the evolution of language will depend upon what eventually turns out to be common, and what divergent, in human and animal. If the data eventually show that animals use simpler rules in forming auditory categories for speech sounds while humans use a more intricate context-dependent set of rules, then we would conclude that the constraints imposed by the auditory system provided a set of broad guidelines that served to initially structure the acoustics of language but did not solely determine them. These constraints could have taken the form of a set of natural psychophysical boundaries (Kuhl & Miller, 1975) whose inherent characteristics included poor discriminability among stimuli falling on one side of such boundaries but good discriminability for stimuli straddling them. Given that these natural psychophysical boundaries were determined by the mammalian auditory system, it would have been natural for the acoustics of language to reflect these constraints (Kuhl, 1979b; Stevens, 1981). But even if one admits to the existence of natural psychophysical boundaries and their role in the evolution of speech-sound categories, the question of how complete an explanation this provides for the perception of speech-sound categories in humans still remains. Since speech categories are represented by diverse acoustic events, a complete account based solely on auditory constraints would require one to argue that not only boundaries, but also category centers, are determined by the functional characteristics of the auditory system (see Kuhl, *in press*, Kuhl & Padden, 1983, and Stevens, 1981, for discussion).

In summary, we have shown that animals display the tendency to partition continua in ways that are conducive to the phonetic discrimination of voiced and voiceless stimuli. This was shown in an identification task using stimuli from a voiced-voiceless continuum; the animals behaved as though they perceived an abrupt change in the quality of the stimulus at precisely the point at which many languages separate the categories (Kuhl & Miller, 1975, 1978; Waters & Wilson, 1976). Also, animals demonstrate poor discriminability for within-category acoustic variants and good discriminability for between-category acoustic variants. This was seen in previous studies (e.g., Kuhl, 1981) and in this experiment. Further studies will be required to determine

exactly how far the analogy extends. Their outcomes have important implications for models of speech processing and for understanding the evolution of language.

REFERENCE NOTE

1. Miller, J. L., & Eimas, P. D. *Contextual perception of voicing by infants*. Paper presented at the meeting of the Society for Research in Child Development, Boston, April 1981.

REFERENCES

- ABRAMSON, A., & LISKER, L. Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the Sixth International Congress of Phonetic Science, Prague, 1967*. Prague: Academia, 1970.
- CARNEY, A. E., WIDIN, G. P., & VIEMEISTER, N. F. Non-categorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 1977, **62**, 961-970.
- EILERS, R. E., GAVIN, W., & WILSON, W. R. Linguistic experience and phonemic perception in infancy: A cross-linguistic study. *Child Development*, 1979, **50**, 14-18.
- EIMAS, P. D. Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, 1974, **16**, 513-521. (a)
- EIMAS, P. D. Linguistic processing of speech by young infants. In R. Schiefelbusch & L. Lloyd (Eds.), *Language and perspectives—Acquisition, retardation, and intervention*. Baltimore: University Park Press, 1974. (b)
- EIMAS, P. D. Auditory and phonetic coding of the cues for speech: Discrimination of the /r-/l/ distinction by young infants. *Perception & Psychophysics*, 1975, **18**, 341-347.
- EIMAS, P. D., SIQUELAND, E. R., JUSCEYK, P., & VIGORITO, J. Speech perception in infants. *Science*, 1971, **171**, 303-306.
- EIMAS, P. D., & TARTTER, V. C. On the development of speech perception: Mechanisms and analogies. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 13). New York: Academic Press, 1979.
- FISHER-JORGENSEN, E. Acoustic analysis of stop consonants. *Miscellanea Phonetica*, 1954, **2**, 42-59.
- GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.
- JUSCEYK, P. W. Infant speech perception: A critical appraisal. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech*. Hillsdale, N.J.: Erlbaum, 1981.
- KUHL, P. K. Predispositions for the perception of speech-sound categories: A species-specific phenomenon? In F. D. Minifie & L. L. Lloyd (Eds.), *Communicative and cognitive abilities—Early behavioral assessment*. Baltimore: University Park Press, 1978.
- KUHL, P. K. Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, 1979, **66**, 1668-1679. (a)
- KUHL, P. K. Models and mechanisms in speech perception: Species comparisons provide further contributions. *Brain, Behavior and Evolution*, 1979, **16**, 374-408. (b)
- KUHL, P. K. Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America*, 1981, **70**, 340-349.
- KUHL, P. K. Categorization of speech by infants. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming buzzing confusion*. Hillsdale, N.J.: Erlbaum, in press.
- KUHL, P. K., & MILLER, J. D. Speech perception by the chin-chilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 1975, **190**, 69-72.
- KUHL, P. K., & MILLER, J. D. Speech perception by the chin-chilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, 1978, **63**, 905-917.
- KUHL, P. K., & PADDEN, D. M. Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America*, 1983, in press.
- 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 Psychology*, 1975, **20**, 215-225.
- LIBERMAN, A. M., DELLATRE, P. C., & COOPER, F. S. Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, 1958, **1**, 153-167.
- LISKER, L. Is it VOT or a first-formant transition detector? *Journal of the Acoustical Society of America*, 1975, **57**, 1547-1551.
- LISKER, L., & ABRAMSON, A. S. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 1964, **20**, 384-422.
- LUCE, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology*. New York: Wiley, 1963.
- MILLER, J. D., WIER, C. C., PASTORE, R. E., KELLY, W. J., & DOOLING, R. J. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, 1976, **60**, 410-417.
- MORSE, P. A. Infant speech perception: A preliminary model and review of the literature. In R. Schiefelbusch & L. Lloyd (Eds.), *Language perspectives—Acquisition, retardation, and intervention*. Baltimore: University Park Press, 1974.
- MORSE, P. A., & SNOWDON, C. T. An investigation of categorical speech discrimination by rhesus monkeys. *Perception & Psychophysics*, 1975, **17**, 9-16.
- PISONI, D. B. Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, 1977, **61**, 1352-1361.
- REPP, B. H. Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, 1979, **22**, 173-189.
- ROSCH, E. Natural categories. *Cognitive psychology*, 1973, **4**, 328-350.
- SINNOT, J. M., BEECHER, M. D., MOODY, D. B., & STEBBINS, W. C. Speech sound discrimination by monkeys and humans. *Journal of the Acoustical Society of America*, 1976, **60**, 687-695.
- STEVENS, K. N. The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David, Jr., & P. B. Denes (Eds.), *Human communication: A unified view*. New York: McGraw-Hill, 1972.
- STEVENS, K. N. Constraints imposed by the auditory system on the properties used to classify speech sounds: Evidence from phonology, acoustics, and psychoacoustics. In T. Myers, J. Laver, & J. Anderson (Eds.), *Advances in psychology: The cognitive representation of speech* (Vol. 7). Amsterdam: North-Holland, 1981.
- STEVENS, K. N., & KLATT, D. H. Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*, 1974, **55**, 653-659.
- STREETER, L. Language perception of two-month-old infants shows effects of both innate mechanisms and experience. *Nature*, 1976, **259**, 39-41.
- SUMMERFIELD, Q., & HAGGARD, M. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, 1977, **62**, 435-448.

WATERS, R. S., & WILSON, W. A., JR. Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. *Perception & Psychophysics*, 1976, 19, 285-289.

WOOD, C. C. Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America*, 1976, 60, 1381-1389.

NOTE

1. In typical same-different formats with human listeners, S trials consist of both 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 nine stimulus pairs (i.e., repeated measures), we chose the restricted format described above, in which S trials consist of AA pairs and D trials consist of AB pairs. The B stimulus in any given pair was the stimulus with the higher VOT value. This format tended to increase the animal's response bias, but the analysis procedure allowed the separation of response bias and discriminability.

(Manuscript received March 25, 1982; revision accepted for publication September 23, 1982.)