Perception & Psychophysics
1991, 50 (2), 93-107

Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not

PATRICIA K. KUHL
University of Washington, Seattle, Washington

Many perceptual categories exhibit internal structure in which category prototypes play an important role. In the four experiments reported here, the internal structure of phonetic categories was explored in studies involving adults, infants, and monkeys. In Experiment 1, adults rated the category goodness of 64 variants of the vowel /i/ on a scale from 1 to 7. The results showed that there was a certain location in vowel space where listeners rated the /i/ vowels as best instances, or *prototypes*. The perceived goodness of /i/ vowels declined systematically as stimuli were further removed from the prototypic /i/ vowel. Experiment 2 went beyond this initial demonstration and examined the effect of speech prototypes on perception. Either the prototypic or a nonprototypic /i/ vowel was used as the referent stimulus and adults' generalization to other members of the category was examined. Results showed that the typicality of the speech stimulus strongly affected perception. When the prototype of the category served as the referent vowel, there was significantly greater generalization to other /i/ vowels, relative to the situation in which the nonprototype served as the referent. The notion of a *perceptual magnet* was introduced. The prototype of the category functioned like a perceptual magnet for other category members; it assimilated neighboring stimuli, effectively pulling them toward the prototype. In Experiment 3, the ontogenetic origins of the perceptual magnet effect were explored by testing 6-month-old infants. The results showed that infants' perception of vowels was also strongly affected by speech prototypes. Infants showed significantly greater generalization when the prototype of the vowel category served as the referent; moreover, their responses were highly correlated with those of adults. In Experiment 4, Rhesus monkeys were tested to examine whether or not the prototype's magnet effect was unique to humans. The animals did not provide any evidence of speech prototypes; they did not exhibit the magnet effect. It is suggested that the internal organization of phonetic categories around prototypic members is an ontogenetically early, species-specific, aspect of the speech code.

Human perceptual systems group stimuli into categories, many of which exhibit internal structure. Evidence of internal structure and organization derives from data indicating that all members of a category are not perceived as equal. Category goodness is a matter of degree, where some members are perceived as better exemplars, more representative or prototypic, than others (Rosch, 1975).

Work on the internal structure of categories and prototypes of categories has typically been done using stimuli in the visual domain (e.g., color or physical objects). Studies of visual categories show that good exemplars of a category have privileged status; they are more quickly encoded, they are more durably remembered, and they are often preferred over other members of the category (Garner, 1974; Goldman & Homa, 1977; Mervis & Rosch, 1981; Rosch, 1975, 1977).

The focus of the experiments reported here is the underlying psychological structure of speech categories. The questions are: Do speech categories exhibit internal structure, and if so, do prototypes[1] play a role in structuring speech categories? There is some evidence relevant to these issues. For example, it has been demonstrated that members of a phonetic category differ in perceptual potency; certain stimuli are more effective adaptors in selective adaptation experiments (Miller, Connine, Schermer, & Kluender, 1983; Samuel, 1982) and are more effective competitors in dichotic competition experiments (Miller, 1977; Repp, 1977). Moreover, when listeners are asked to rate the category goodness of individual members, it has been shown that individual exemplars vary in the degree to which they are perceived as good exemplars of the category (Grieser & Kuhl, 1989; Kuhl, 1986; Miller & Volaitis, 1989). Thus, the work suggests that speech stimuli, like complex stimuli from other domains, are graded, both quantitatively (certain members are more effective than others) and qualitatively (certain

---

This research was supported by a grant to the author from the National Institutes of Health (DC 00520). The author thanks D. Padden and E. Stevens for animal testing and statistical analyses, and Andrew Meltzoff for insightful comments on an earlier draft of the paper. Correspondence should be sent to Patricia K. Kuhl, WJ-10, University of Washington, Seattle, WA 98195.

stimuli are perceived as better exemplars than others). These data provide evidence that speech categories may be organized with reference to a good exemplar, a prototype of the category.

The claim has been made that speech prototypes are represented in long-term memory (Grieser & Kuhl, 1989; Kuhl, 1990; Oden & Massaro, 1978; Samuel, 1982). In this laboratory, work has focused on how phonetic prototypes—representations of phonetic events stored in long-term memory—function in perception. An equally important goal has been to study the origins of prototypes, from both an ontogenetic and a phylogenetic perspective.

Regarding ontogeny, our initial studies showed that *infants' perception* of vowel categories is affected by typicality, as defined by *adult* speakers of the language (Grieser & Kuhl, 1989). In the Grieser and Kuhl experiment, infants were tested in a task in which either a "good" vowel (a vowel judged to be a prototypic exemplar by adults) or a "poor" vowel (nonprototypic) was used as a referent to which novel vowels from the category were compared. The results showed that infants' generalization to novel vowels was affected by the goodness of the referent stimulus. In other words, when the prototype (as opposed to the nonprototype) served as the referent vowel, infants generalized to a significantly larger number of the novel vowels, effectively perceiving a larger speech category. Moreover, the finding that the prototype resulted in greater generalization suggested that phoneme categories may be represented by prototypes in human infants just 6 months of age.

The purpose of the present experiments was to extend these results to further examine the nature, function, development, and species specificity of speech prototypes. Speech prototypes were investigated in three populations—human adults, human infants at 6 months of age, and Rhesus monkeys—using identical stimuli and only minor variations in the techniques and procedures used to test the three populations. Previous experiments have shown that nonhuman animals exhibit some of the speech effects that have been demonstrated in young human infants, such as categorical perception (see Kuhl, 1987, 1988, for review). Although categorical perception effects may not be species-specific, the processing of speech signals that depend on a phonetic level representation should not be demonstrable in nonhuman animals. One goal, therefore, was to determine whether the effect of stimulus goodness seen in human infants was also found in an animal, or whether human and animals' perception of speech diverged at this level of analysis.

The present experiments examined four questions: (1) Do exemplars of a vowel category judged as belonging to the same phonetic category nonetheless vary in perceived category goodness (typicality) to adult speakers of the language? (2) Does perceived typicality affect adults' perceptual organization of the speech category? (3) Does typicality as established by adults differentially affect *infants'* perceptual organization of vowel categories? (4) Is the perceptual effect of typicality attributable to basic auditory processes common to monkey and man, or is the effect of typicality unique to human adults and infants?

EXPERIMENT 1

Conducting prototype experiments for speech categories required (1) a set of stimuli from a single speech category that varied acoustically so that we could examine how acoustic variation affected perceived typicality, and (2) a method that could be used to test whether the category goodness of a vowel stimulus differentially affected perception of other stimuli in the category.

A new set of stimuli was generated for use in these experiments. The stimuli provided twice the degree of acoustic variation that was present in the set of stimuli used by Grieser and Kuhl (1989). The stimuli were generated and goodness ratings were obtained from adult listeners in Experiment 1. These stimuli were then used to test how speech prototypes function in perception for adults (Experiment 2), infants (Experiment 3), and monkeys (Experiment 4).

Method

Subjects

Sixteen adults, with normal hearing, participated in Experiment 1. There were 8 subjects in each of two conditions. All had some training in phonetics and were students at the University of Washington. They ranged from 20 to 41 years of age, with a mean of 27.2 years. Each subject was paid $5 for participating in the experiment.

Rationale for the Design of the Experiments

The first hypothesis being tested was whether, for adults, vowel stimuli varied systematically in perceived goodness; however, the purpose of Experiment 1 went beyond this. The goal was to create a set of stimuli that revealed how quality judgments were affected by distance from the category prototype. We therefore wanted stimuli that varied in quantifiable steps.

Moreover, stimuli that varied in quantifiable steps were necessary to achieve the second goal in this set of experiments, which was to examine the role played by the category prototype in structuring the category. The plan was to select a prototype and a nonprototype stimulus from a vowel category and measure stimulus generalization around each of them using stimuli that varied in quantifiable steps. The specific question was whether listeners (adults, infants, and animals) tested on the prototype would show a greater degree of generalization to other stimuli in the category than they would when tested on a nonprototype of the category. If speech categories have *no* internal structure, then the degree of generalization to variants surrounding the prototype and the nonprototype should not differ; generalization should simply depend on psychophysical distance from each of the two vowels. However, if the category *is* internally structured, generalization should differ significantly; specifically, the prototype should show broader generalization to other members of the category than should the nonprototype.

Stimuli

The /i/ vowel, as in *peep*, was chosen for use in these studies. There were two reasons for this. First, the vowel /i/ is used universally in the world's languages; it is one of the three "point" vowels (/i/, /a/, and /u/), the vowels that are at the articulatory and acoustic extremes of the vowel space (Jakobson, Fant, & Halle, 1969). If prototypes for vowels exist, these three vowels would seem to be ideal candidates. K. N. Stevens's (1972, 1989) "quantal theory" asserts that these three vowels are acoustically more stabile than other vowels. Second, previous data from our lab had shown that infants correctly categorize perceptually diverse instances of /i/ vowels, ones spoken by men, women, and children (Kuhl, 1979; Kuhl, 1985a, for review). Infants' ability to categorize vowels according to phonetic category makes it possible to go further and explore whether these categories show internal structure and organization.

The prototype and nonprototype vowels. Grieser and Kuhl (1989) synthesized a large set of /i/ vowels for the purpose of selecting a prototype and a nonprototype from that category. The /i/ vowel stimuli covered the entire range of formant frequencies produced by typical male speakers (Peterson & Barney, 1952). We selected an /i/ vowel that was consistently judged by adult speakers of the language as the best /i/ (see rating procedure below) and designated it as the *prototype* (P) /i/. A second vowel was chosen from among the set of vowels that had been synthesized. This /i/ vowel was consistently judged by adults as a relatively poor exemplar of an /i/ vowel and was designated the *nonprototype* (NP) /i/. It is of fundamental importance to the logic of the study that this relatively poor exemplar was always judged as /i/ rather than as some other vowel by adults; both the P and the NP were easily identified as exemplars of the /i/ category.

Variants around P and NP /i/ vowels. We created a set of category variants that surrounded the P and the NP, forming four orbits around each of the stimuli (Figure 1). The distance between the four orbits and the P (or the NP) vowel was equated in psychophysical terms. The metric used to equate the psychophysical distance of the orbits from P and NP was the mel scale (S. S. Stevens, Volkmann, & Newman, 1937).

The formants of vowels have often been plotted on a linear frequency scale (e.g., Peterson & Barney, 1952). There are two reasons for converting formant frequencies to their corresponding pitch values on a mel scale (Fant, 1973). First, the mel scale is essentially linear at low frequencies and logarithmic at high frequencies; this corresponds well to spatial location on the ear's basilar membrane. Second, the minimum perceptible shift in formant pitch is the same order of magnitude for all three formants when formants are converted to mels. The mel scale was originally invented to equate the magnitude of a perceived change in pitch at different frequencies. Using the mel scale in the present case allowed the creation of stimuli that differed from the P and the NP in uniform steps. Thus, the variants around the P and the NP were scaled in a common metric, and this equated psychophysical distance between the four orbits and their respective vowel targets.

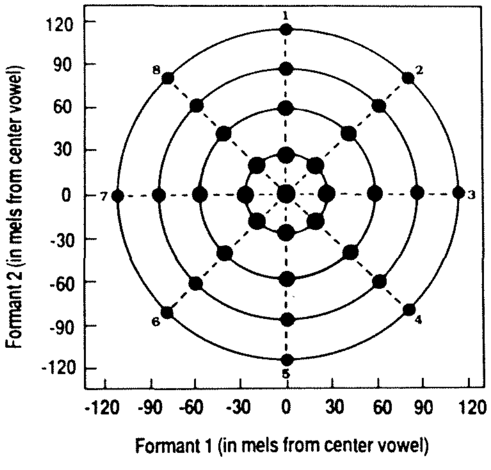

Figure 1. Formant frequency values in mels for stimuli surrounding a center vowel stimulus. The stimuli form four orbits and eight vectors around the center stimulus. The stimuli on each orbit are a specified distance in mels from the center vowel (30, 60, 90, or 120 mels, starting from the first orbit); the eight stimuli on each orbit differ in the direction and amount of formant frequency change.

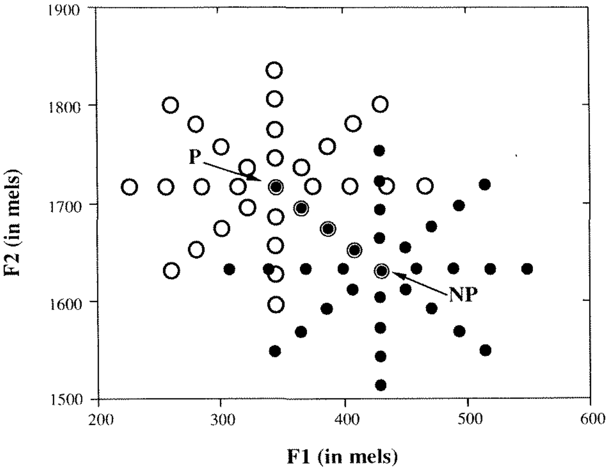

Figure 2. The prototype /i/ vowel (P) and variants on four orbits surrounding it (open circles) and the nonprototype /i/ vowel (NP) and variants on four orbits surrounding it (closed circles). The stimuli on one vector were common to both sets.

Formant frequency values were chosen from a mel-scaled vowel space such that there were 32 variants, eight located on each of four orbits (O_1–O_4) around the two vowel stimuli. O_1 was located 30 mels from the target vowel; O_2 through O_4 were located at 60, 90, and 120 mels, respectively (Figure 1). To illustrate how the stimuli were made, consider stimuli on Vector 1 around P. Here the P's second formant was increased to create new variants. The amount of change was dictated by the orbit the new variant was on: to create the variant on O_1, the P's second formant was increased by 30 mels; on O_2 the P's second formant was increased by 60 mels, and so on. The formula used to relate absolute frequencies to their mel-scale equivalents was the one recommended by Fant (1973): $y = k \log (1 + f/1000)$, where y is the mel-scale value, k is a constant, and f is the formant frequency in hertz.

The entire set of variants orbiting P and NP is shown in Figure 2. Note that the stimuli along one vector were common to both the P and the NP sets.

Synthesis. The stimuli were created using Klatt's (1980) cascade-parallel speech synthesizer, which was simulated on a DEC PDP 11/34 computer. Amplitude contours, fundamental frequency contours, formant frequency values, and formant bandwidth values were entered to produce vowel stimuli with five formants. The variants were created by manipulating the values of the first and second formants; the values of the third, fourth, and fifth formants remained constant for all vowels at 3010, 3300, and 3850 Hz, respectively. The bandwidths of the first three formants were set at the values recommended by Klatt (1980). The stimuli were 500 msec in duration. The fundamental frequency for all stimuli began at 112 Hz, rose to 132 Hz over the first 100 msec, and dropped to 92 Hz over the next 400 msec to produce a rise-fall contour. The stimuli were presented at 68 dB SPL, measured on the A scale of a sound-level meter (Bruel and Kjaer, 2106) placed in the approximate position of the subject's head.

Adult goodness ratings. Quantitative ratings of category goodness (typicality) of each variant were obtained using a 7-point rating scale (7 = *a good exemplar*, one representative of the /i/ vowel category as a whole; 1 = *a poor exemplar*, one not representative

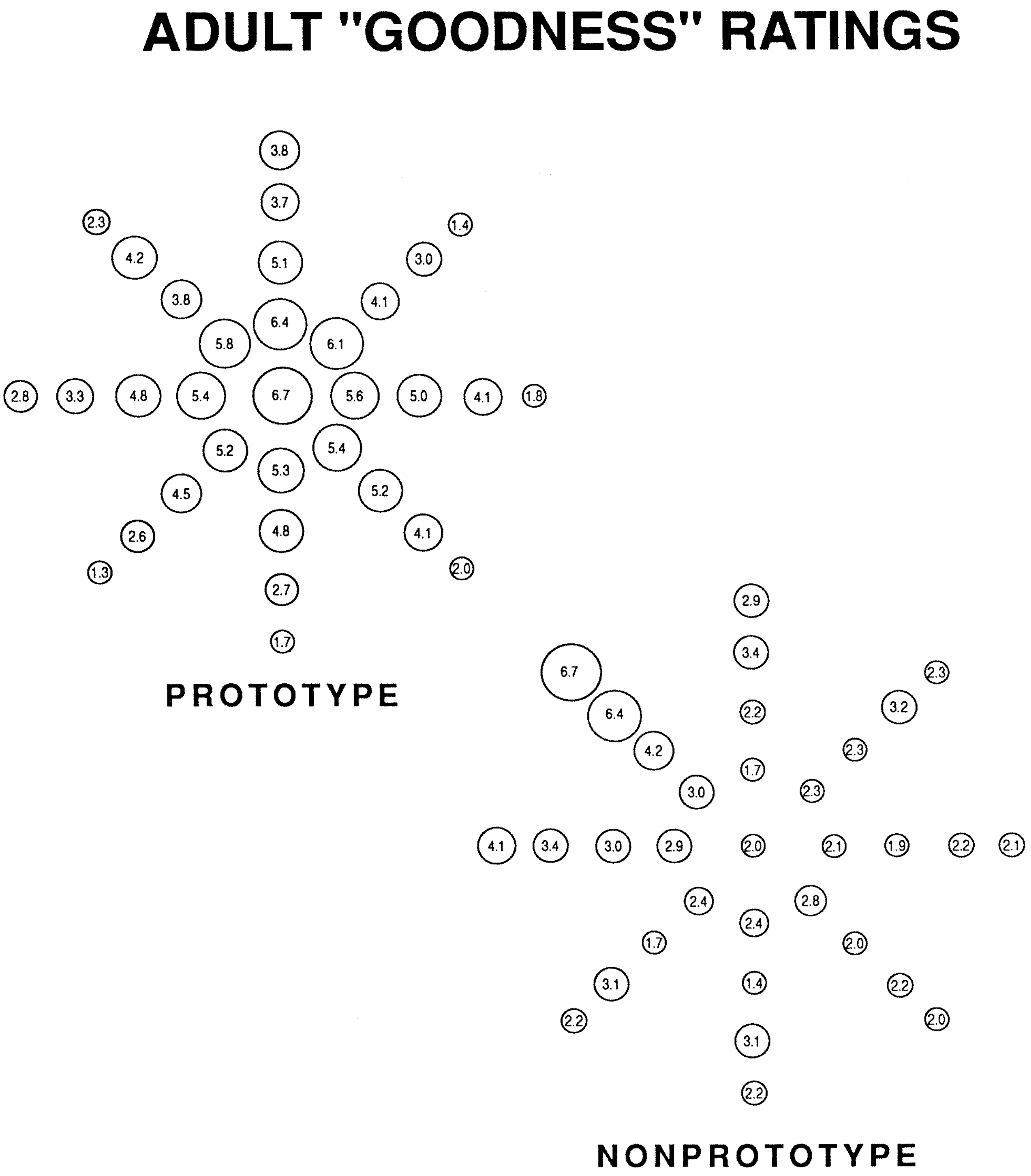

Figure 3. Category goodness (typicality) ratings for the prototype /i/ vowel, the nonprototype /i/ vowel, and the variants surrounding each of the two vowels. Typicality was judged by adults using a scale from 1 (a poor exemplar) to 7 (a good exemplar). The size of the circles correlates with the degree of goodness, with larger circles indicating better exemplars.

of the category as a whole). Sixteen adults, 8 in each condition (P or NP), listened to the variants surrounding one of the two vowels, presented in random order. No reference for a good exemplar of the /i/ vowel was provided to them. The subjects were given a card on which the word *peep* was written; they were told to rate the perceived goodness of each stimulus as an exemplar of the vowel contained in the word on the card.

The adults were tested, one at a time, in a sound-treated testing suite. They listened to the stimuli over the same loudspeaker (Electrovoice SP-12) that was used in the discrimination tests on adults (Experiment 2) and on infants (Experiment 3). The task was computer-controlled and self-paced; the adults pressed a button to present a stimulus and then circled a number from 1 to 7 on their answer sheet following each stimulus. Each adult rated each stimulus five different times.

Results

The average ratings for the stimuli are plotted in Figure 3. The size of the circles represents the relative goodness, or typicality, of each stimulus. As shown, the P was given an average rating of 6.7 and the NP was given an average rating of 2.0. The ratings for the stimuli nearest the P tended to be highest; the ratings consistently decreased with increases in distance from P. Conversely, stimuli in the orbits around NP received relatively low ratings, with an increase in ratings as they neared the region of the vowel space occupied by P.

Category goodness ratings were subjected to a two-way analysis of variance (ANOVA), with repeated measures on the second factor, which examined the effects of condition (P vs. NP) and distance (O_1–O_4). The results showed that the stimuli surrounding P received significantly higher ratings than did the stimuli surrounding NP, as reflected in the highly significant main effect for condition [$F(1,14) = 432.7, p < .001$]. There was also a main effect of distance [$F(3,42) = 168.7, p < .001$], and because the effect of distance was much more symmetrical for stimuli orbiting P when compared with NP, a significant condition × distance interaction effect was also obtained [$F(3,42) = 391.3, p < .001$]. Follow-up tests for simple effects revealed that the effect of distance was highly significant for both groups considered individually (both $ps < .001$).

Discussion

In Experiment 1, adult speakers of English provided category goodness ratings for 64 /i/ vowel stimuli. The results showed that the set of /i/ vowels varied in perceived category goodness. The ratings were highly consistent across listeners. These ratings revealed that all /i/ vowels are not perceived to be equal by adults; some are perceived as better exemplars of the category than others. Stimuli in a particular region of the vowel space stood out as the best exemplars of the category; moreover, the ratings declined evenly and symmetrically around the best instances.

EXPERIMENT 2

The hypothesis was that if vowel categories were internally structured, the typicality of the exemplar used as a referent for the vowel category should affect perception of other members of the category. Specifically, if the prototype was more representative of the category as a whole, then it should be perceived as more similar to other members of the category than would a nonprototype from the category. The prototype should thus produce a broader generalization gradient when it serves as the referent of the category to which other members are compared.

Method

Subjects

Sixteen adults, with normal hearing, participated in Experiment 2. There were 8 subjects in each of the two conditions (P and NP). None of these adults had been tested in Experiment 1. The subjects tested in Experiment 2 had some training in phonetics and were students at the University of Washington. They ranged from 23 to 38 years of age, with a mean of 29.8 years. Each subject was paid $5 for participating in the experiment.

Stimuli

The same stimuli used in Experiment 1 were used in Experiment 2. In the P condition, the prototype /i/ vowel served as the referent stimulus and its 32 surrounding variants served as the comparison stimuli in a discrimination task. In the NP condition, the nonprototype /i/ vowel served as the referent stimulus and its 32 surrounding variants served as the comparison stimuli in the discrimination task.

Equipment and Test Apparatus

The test suite consisted of a sound-treated booth and an adjoining control room. The booth contained a table and chair, loudspeaker, and a visual reinforcer. A button box was affixed to the table. A loudspeaker (Electrovoice SP-12) was located at a 45° angle to the left of the adult. The visual reinforcer (see Experiment 3) sat on top of the loudspeaker. In the control room, the experimenter operated a PDP-11/34 computer with terminal and printer. The program was self-paced and presented trials to the subject on a variable-interval schedule that mimicked a typical session with an infant. The subject's button responses were monitored, and all contingencies were delivered automatically.

Procedure

Each adult sat at a table and listened to stimuli presented over a loudspeaker. The subject was instructed to press a button when the *referent* speech sound, which was repeated continuously once per second, was changed to a *comparison* speech sound for 4.5 sec. Two kinds of trials were run, each with a probability of .5. During *test* trials, the referent vowel was changed to a comparison stimulus and button presses that occurred during the 4.5-sec observation interval were reinforced with a visual signal (the same used for infants; see Experiment 3). During *control* trials, the referent vowel was not changed, but the adults' responses were monitored during the 4.5-sec observation interval; these control trials were run to assess the probability of false-positive responses. The order of trials was quasirandom, with no more than three test or three control trials occurring in a row. Using the response data, four possible outcomes were scored. On test trials, a *hit* was scored if a response occurred and a *miss* was scored if no response occurred. On control trials, a *correct rejection* was scored if no response occurred and a *false positive* was scored if a response occurred. Only hit responses were reinforced with the visual stimulus.

The subjects were informed that all of the stimuli belonged to the same phonetic category and that they were to respond to any change that they heard in the stimulus. The discrimination test consisted of 128 trials (32 test stimuli × 2 trials each = 64 test trials, plus 64 control trials).

The hypothesis was that, in the P condition, when the adults listened to the prototype /i/ as the referent vowel, fewer differences would be detected between the novel variants and the P because the subjects would perceive the novel variants as more similar to the prototype /i/. In other words, discrimination would be more difficult in the P condition than in the NP condition, even though psychophysical distance was controlled across the two conditions. To operationalize this, two measures were of interest. First, the *overall percent-correct scores* for the two conditions provided an indicator of overall difficulty of discriminating within-category variants from the two referent vowels. Second, the number of miss responses at each distance provided a direct measure of subjects' ability to detect a change in the referent vowel. Miss responses indicated that adults perceived the referent vowel and the comparison vowel as similar; in other words, they were *generalization* responses. The *generalization score* was the percentage of all test trials in which subjects produced a generalization response.

Results

The results strongly support the hypothesis that adults' perception of within-category vowel differences is affected by typicality. Overall, adults were highly accurate at detecting within-category vowel differences. Across both conditions, overall percent-correct scores were above 75%. However, discrimination performance varied significantly, depending on the typicality of the referent exemplar. When a stimulus perceived as having high category goodness was used as the referent vowel in the discrimination task, overall percent-correct scores were significantly lower, indicating difficulty in perceiving differences between the prototype and other members of the category. Generalization scores were significantly higher in this condition, indicating that adults perceived the prototype as more similar to its surrounding variants when compared with the nonprototype in relation to its surrounding variants.

Examining first the overall discrimination performance, the percent-correct scores [(percent hits + percent correct rejections)/2] achieved by adult subjects in the two conditions are displayed in Figure 4. As shown, in the P condition, adults achieved an overall score of 78.6% correct, whereas, in the NP condition, adults achieved a score of 90.5% correct. A *t* test for independent groups was used to compare the overall percent-correct scores for the two conditions. The analysis revealed that this difference was highly significant [$t(14) = 6.89$, $p < .001$].

A more detailed look at the results can be seen in Figure 5, which presents the generalization scores. Generalization scores were derived from test trials; specifically, the average percentage of test trials in which subjects indicated that they perceived the two stimuli as the same was calculated. The results provide support for the hypothesis of internal structure, with the prototype stimulus being perceived as more similar to other variants than was the nonprototype stimulus. Adults in the P condition produced higher generalization scores at each distance (O_1, 28.3%; O_2, 5.1%; O_3, 2.9%; O_4, 0.7%) when compared with the adults in the NP condition (O_1, 5.8%; O_2, 2.6%; O_3, 0.9%; O_4, 0.4%). A two-way ANOVA with repeated measures on the last factor was conducted, examining the effects of condition (P vs. NP) and distance (O_1–O_4). The results revealed a highly significant effect of condition [$F(1,14) = 45.7$, $p < .001$], reflecting the fact that at each distance generalization was higher for the prototype condition. As expected, the effect of distance was also highly significant [$F(3,42) = 12.9$, $p < .001$], indicating that, for both groups, generalization decreased as the comparison stimulus moved further away from the referent vowel (from O_1 to O_4). Stimuli further away from the referent vowel are physically less similar to the referent vowel and are more discriminable. The condition × distance interaction was also significant [$F(3,42) = 6.8$, $p < .001$]. Follow-up tests for simple effects showed that the effect of distance was highly significant for each group considered individually (both *p*s $< .001$).

Discussion

Experiment 2 was a test of the hypothesis that phonetic categories exhibit internal structure. Specifically, the rationale was that if speech categories exhibit internal structure, then tests of discrimination between members of the same phonetic category would reflect that structure: When

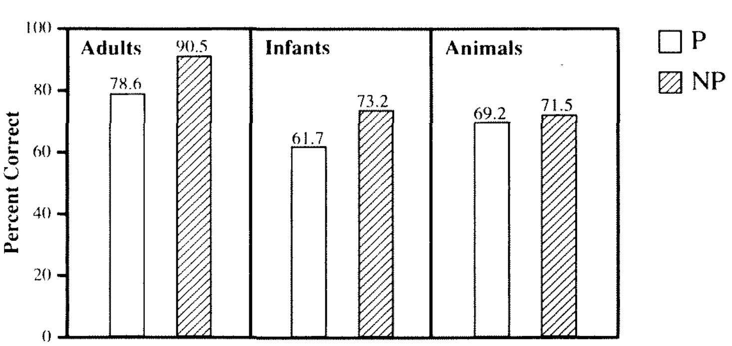

Figure 4. Average overall percent-correct scores achieved by adults (Experiment 2), infants (Experiment 3), and monkeys (Experiment 4) in the prototype (P) and the nonprototype (NP) conditions. For adults and infants (but not monkeys), there is a statistically significant difference between scores in the two conditions, with overall percent-correct scores being higher in the nonprototype condition.

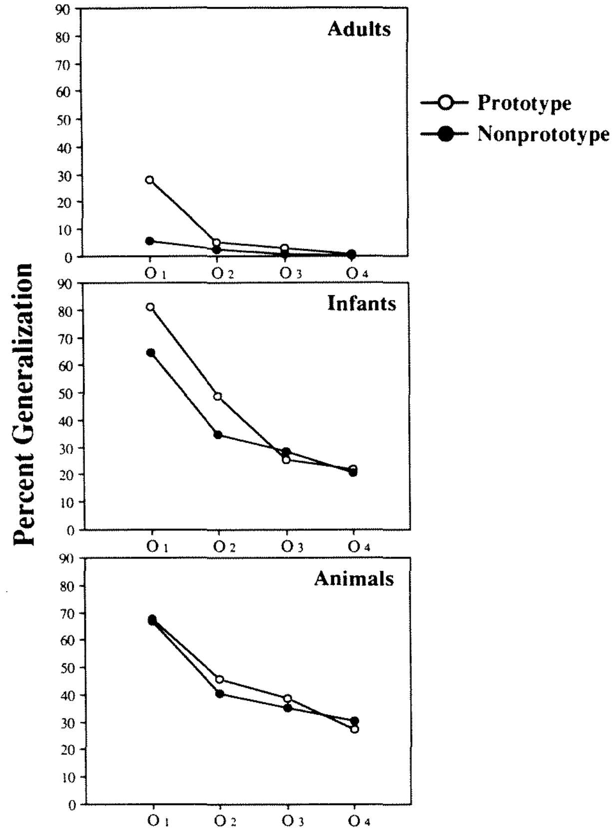

Figure 5. Average generalization scores shown for stimuli surrounding the prototype and the nonprototype by adults (Experiment 2), infants (Experiment 3), and monkeys (Experiment 4). For adults and infants (but not monkeys), there is a statistically significant difference between generalization scores in the two conditions, with scores being higher in the prototype condition.

a "good" (prototypic), as opposed to a "poor" (nonprototypic), exemplar of the phonetic category served as the referent stimulus to which other stimuli from the category were compared, the more central stimulus would be perceived as more similar to other members of the category. In the absence of internal structure, the two referents of the category would be expected to behave similarly in tests of stimulus generalization because psychophysical distance was equated. Thus, the hypothesis of Experiment 2 was that the vowel designated as the prototype /i/ vowel (on the basis of the goodness ratings of Experiment 1) was the more central of the two and would thus be perceived as more similar to other members of the category.

The data demonstrated that adults' perception was strongly affected by typicality. The measure of stimulus generalization showed that, at each distance from the referent, listeners produced greater generalization to the stimuli surrounding the prototype stimulus. These results confirm the hypothesis that listeners perceive the prototype stimulus as more similar to other members of the category than is the nonprototype of the category.

The findings suggest two things: (1) the members of phonetic categories are graded with regard to their representativeness or typicality, and (2) the best instances of a phonetic category play a special role in perception. Both findings support the contention that phonetic categories are internally structured.

It is worth examining the idea that phonetic categories are *perceptually* graded. The results of Experiment 1 showed quite convincingly that the members of a speech category are not perceptually equivalent. Specific members were perceived as better members than others. Representativeness was thus graded. Moreover, goodness ratings declined in a consistent fashion, with a particular location in vowel space resulting in better ratings. A systematic and symmetrical decline in goodness ratings was seen as stimuli moved from a specific location, suggesting that the category is organized with reference to category goodness.

It is also worth examining the notion that phonetic categories are *functionally* graded. The results of Experiment 2 demonstrated that the members of speech categories are not functionally equivalent. The prototype stimulus was perceived as more similar to other members of the category than was the nonprototype of the category. This finding suggests that prototypes play a special role in the perception and organization of speech categories. Our working hypothesis is that a prototype acts like a perceptual "magnet": Surrounding members of the category are perceptually assimilated to it to a greater degree than would be expected on the basis of real psychophysical distance. Relative to a nonprototype of the category, the distance between the prototype and surrounding members is effectively decreased; in other words, the perceptual space appears to be "warped," effectively shrunk around the prototype. The prototype of the category thus serves as a powerful anchor for the category, and the prototype's functional role as a perceptual magnet for the category serves to strengthen category cohesiveness.

Important questions arise from this finding: How does the perceptual magnet effect come about? Do infants exhibit the effect? And if so, is it because speech prototypes are specified by special innately determined speech mechanisms, or might the effect be inherent in the basic auditory processing of these stimuli? These questions were addressed in two further experiments, one exploring the ontogeny of the effect and the other examining its phylogeny. In Experiment 3, infants were tested to see whether they demonstrated the perceptual magnet effect; in Experiment 4, nonhuman animals were tested to determine whether they demonstrated the effect.

EXPERIMENT 3

Experiment 3 was designed to replicate the discrimination tests using the prototype and nonprototype stimuli with infant subjects. The methods and procedures used with the infants duplicated those used with adults with only slight changes, such as in the nature of the response (adults pressed a button, infants produced a head-turn response), that were necessitated by the age of the subjects.

Method

Subjects

Thirty-two normal full-term infants were tested (16 in each of the two conditions, P and NP). The infants ranged in age from 6.0 to 7.0 months old ($M = 6.53$ months). Thirteen infants (6 from the P group and 7 from the NP group) were eliminated from the experiment: 6 infants could not be conditioned, and 7 did not return for all the test sessions and therefore did not complete the experiment.

Stimuli

The same stimuli used in Experiment 2 were used in Experiment 3. Infants in the P condition were tested with the prototype /i/ vowel and its 32 surrounding variants; infants in the NP condition were tested with the nonprototype /i/ vowel and its 32 surrounding variants.

Procedure

A *head-turn conditioning procedure*, a technique commonly used in tests of infant speech perception (see Kuhl, 1985b, for review), was used to test the infants. The procedure involved conditioning an infant to produce a head-turn response for a visual reinforcer when one speech sound (the *referent*), which was repeated continuously once per second, changed to a different speech sound (the *comparison*) for 4.5 sec. The visual reinforcer consisted of a toy animal that moved when activated (a bear pounded a drum, a monkey clapped cymbals). The toy animal was housed in a box made of darkened Plexiglas that was normally opaque; when the reinforcer was activated, the lights in the box were lit and the animated toy became visible.

The two kinds of trials described earlier in tests on adults were run. During *test* trials, the referent vowel was changed to a comparison stimulus and head-turn responses that occurred during the 4.5-sec observation interval (as signaled by the experimenter's vote) were reinforced automatically by the computer. During *control* trials, the referent vowel was not changed, but infants' head-turn responses were monitored during the 4.5-sec observation interval to assess the probability of false-positive responses. The four possible outcomes of the two types of trials were scored exactly as they had been in the adult tests. On test trials, a *hit* was scored if a head-turn occurred; a *miss* was scored if infants failed to turn. On control trials, a *correct rejection* was scored if infants refrained from turning; a *false positive* was scored if a head-turn occurred. Only hit responses were reinforced with the visual stimulus. The infants' false-positive responses were never reinforced.

For the P group, the prototype /i/ served as the referent vowel and the 32 variants surrounding it served as the test stimuli; for the NP group, the nonprototype /i/ vowel served as the referent vowel and the 32 variants surrounding it served as the test stimuli (Figure 2). The question was whether the infants in the P group would show greater generalization to variants surrounding the P relative to the infants in the NP group, just as the adults did. The alternate hypothesis was that infants would be unaffected by the typicality of the referent stimulus; they would then be expected to show only a distance effect with equal generalization to variants surrounding the two referent vowels (P and NP).

Equipment and Test Apparatus

The test suite consisted of the same sound-treated experimental room and the adjoining control room used in the tests on adults. The room arrangement for infant tests has been described in full detail elsewhere (Kuhl, 1985b).

The infant, the caretaker, and an assistant sat in the experimental room. The caretaker, who held the infant, was seated at the table. The assistant sat at a 45° angle to the right of the caretaker and maintained the infant's attention by manipulating one of several silent toys. When the assistant determined that the infant was ready for a trial (that is, the infant was focused on the toy and not crying, fussing, or vocalizing), she pressed the button on the box, which activated the computer and initiated a trial. The assistant did not know what kind of trial would be initiated and therefore could not bias the experiment by her knowledge of the infant's state.

The loudspeaker was located at a 45° angle to the left of the infant. The visual reinforcer was above the loudspeaker at the infant's eye level. A video camera, located directly above the reinforcer, allowed closed circuit monitoring of the infant's head-turn responses. In the control room, the experimenter viewed a Sony television monitor and operated an audio cassette player and a PDP-11/34 computer with terminal and printer. The cassette player was used to provide classical music over headphones to the assistant and the caretaker so that they could not hear the stimuli being presented to the infant and therefore bias the infant's head-turn responses in any way. The experimenter, who could not hear the stimuli presented to the infant, recorded the infant's head-turn responses by pressing a key on the computer. The computer presented stimuli to the infant and controlled all of the contingencies, depending on the infant's behavior.

Test Phases

Training phase. The training phase involved two stages: conditioning and discrimination. During conditioning, only test trials were presented, wherein the referent vowel changed to a comparison vowel. During these initial test trials, a single comparison stimulus was used for each infant. The comparison stimulus was one of the eight stimuli located on O_4 surrounding the referent vowel, chosen independently for each infant to achieve a counterbalanced design.

Head-turn responses were initially shaped by turning on the visual reinforcer halfway through the 4.5-sec observation interval. The reinforcer was bright and made noise, causing the infant to turn away from the assistant and look at the reinforcer. Eventually, detection of the change in the vowel was sufficient by itself to cause the infant to produce a head turn; the reinforcer was not activated unless a full head-turn response was observed. Criterion performance during the conditioning phase was three test trials in a row in which the infant produced a head-turn response before the end of the 4.5-sec observation interval. This had to occur before the end of 35 trials in order to progress to the next stage of training.

When the three-in-a-row criterion was met, the infant progressed to the discrimination phase of training, in which two changes occurred. First, test and control trials were run with equal probability, with the stipulation that no more than three trials of either type could occur sequentially, as was stipulated in the adult tests. Second, the comparison stimuli presented on test trials now included all of the eight stimuli on O_4. During the discrimination phase, the infant was reinforced only if the head-turn response was correctly produced during the observation interval of a test trial; they were never reinforced during control trials. Criterion performance was seven out of eight consecutive correct responses within the first two sessions. The infants failing to meet this criterion were not tested further.

Generalization test. In the generalization phase, the test occurred. During generalization, each infant was tested with all of the 32 /i/ variants surrounding his or her background vowel (either the prototype or the nonprototype). The 32 variants served as the comparison stimuli presented on test trials; they were presented in random order. Each of the 32 variants was presented on two different trials (32 stimuli × 2 = 64 test trials). In addition, 64 control trials were randomly interspersed, so generalization consisted of a total of 128 trials per infant, as was the case for adults. Typically 30-40 trials were run each day in a session lasting about 30 min.

Results

The primary question of interest was whether infants' responses to the prototype and nonprototype stimuli mirrored those of adults. The same two measures used to assess adult performance were used to assess infant performance. First, the overall percent-correct measure was calculated. This measure assessed overall discrimination accuracy using performance on both test and control trials [(percent hits + percent correct rejections)/2]. Second, the direct measure of stimulus generalization used in the assessment of adult performance was also used here. It involved the calculation of infants' generalization scores —the average percentage of test trials in which the infants indicated that they perceived the referent vowel and the variant as the same.

Training Phase Data

The mean number of trials to criterion during conditioning was 13.6 for the P group and 16.7 for the NP group, a difference that was nonsignificant [$t(30) = 1.50$, $p > .15$]. In the discrimination phase, the mean number of trials to criterion was 22.3 for the P group and 26.3 for the NP group, a difference that also was not significant [$t(30) = 1.45$, $p > .15$].[2]

Generalization Test

During the generalization phase, both groups achieved overall percent-correct scores that were significantly above chance (50% correct) [both $ts(15) > 4.5$, $ps < .001$]. As was the case for the adult data, the overall percent-correct measure revealed a significant difference between performance in the P and the NP group, with the NP group discriminating surrounding variants more accurately (73.2% correct) than did the P group (61.7% correct) [$t(30) = 3.87$, $p < .001$]. In the measure of overall percent correct, therefore, infants replicated the result obtained with adults.

Infants' generalization scores are shown in Figure 5, which presents the scores for each condition at each distance. As shown, at each distance, the infants in the P group produced higher generalization scores (O_1, 81.4%; O_2, 48.8%; O_3, 25.3%; O_4, 22.0%) than did the infants in the NP group (O_1, 64.6%; O_2, 34.7%; O_3, 28.7%; O_4, 20.6%). These scores were submitted to a two-way ANOVA with repeated measures on the last factor, examining the main effects of condition (P and NP) and distance (O_1–O_4). As was found with the adults, higher generalization scores were obtained by the infants in the P group relative to the infants in the NP group, and this difference was highly significant [$F(1,30) = 46.3$, $p < .001$]. As expected by straightforward stimulus generalization, the effect of distance from the referent vowel was highly significant [$F(3,90) = 1{,}069.9$, $p < .001$]. The group × distance interaction was also significant [$F(3,90) = 37.3$, $p < .001$]. Follow-up tests for simple effects revealed that the effect of distance was significant for each group (both $ps < .001$).

Analysis of Stimuli on the Common Vector

Both groups of infants (P and NP) had been tested on a subset of identical stimuli, those located on the vector that was shared by the two groups (Figure 2). The only difference between the infants in the two groups was that the infants in the P group heard the prototype stimulus as the referent and compared it with stimuli in the direction of the nonprototype stimulus, whereas the infants in the NP group heard the nonprototype stimulus as the referent and compared it with stimuli in the direction of the prototype. In brief, the only difference was the direction of stimulus change—the stimuli were identical.

We were very interested in whether the two groups of infants displayed different response profiles to these identical stimuli. A polynomial trend analysis was conducted on the group × distance interaction for these stimuli (see Bock, 1975, for the computational procedures). The results revealed that the profiles of the two groups differed, and this produced a significant difference in the quadratic trend [$F(1,30) = 6.82$, $p < .05$]. To hear a difference, the infants in the P group had to go farther away from the prototype than did the infants in the NP group. This directional asymmetry in the perception of within-category differences underscores the prototype's effect on perception: the prototype appears to function like a perceptual magnet.

Comparison of Goodness Ratings and Generalization Data For Adults and Infants

We wanted to test whether adults' goodness ratings provided in Experiment 1 were correlated with adults' and infants' generalization responses. Adults had provided ratings of the degree to which the 32 vowels surrounding the prototype were similar to the category "ideal"; the adults and infants tested in the P condition were given the same /i/ category ideal as a referent and produced generalization responses to the same 32 stimuli. Spearman rank correlation coefficients were calculated between adults' quality ratings for each stimulus (from Experiment 1) and the generalization scores for both the adult subjects (from Experiment 2) and the infant subjects (from Experiment 3). In all cases, the measures demonstrated highly significant positive correlations ($r_s = 0.88$, for adults' goodness ratings and adults' generalization responses; $r_s = 0.86$, for adults' goodness ratings and infants' generalization responses). Moreover, there was a high correlation between adults' generalization responses and infants' generalization responses ($r_s = 0.82$).

Discussion

Experiment 3 provided strong support for the hypothesis that infants' speech categories demonstrate internal structure. Moreover, the direction of the effect mirrored that shown by adult speakers of the language: Stimuli defined by adult speakers of the language as better exemplars of the phonetic category, prototypes of the category, resulted in greater generalization to other members of the category in infants. The prototype appears to function like a perceptual magnet, even for infants only 6 months old. The pattern of results seen here replicate and extend that obtained by Grieser and Kuhl (1989) using twice the number of stimuli and slightly different training procedures, thus demonstrating the robustness of the effect.

The power of the effect was revealed in the analysis of infants' responses to stimuli on the common vector. These particular stimuli served as test stimuli for both groups, and they were located at equal distances in mels from the two stimuli (P and NP). The only difference between the P and NP groups was the stimulus that served as the referent for the perceptual comparisons. The prototype group heard the prototypic ''good'' /i/ as the background referent vowel, whereas the nonprototypic group heard the nonprototypic ''poor'' /i/ as the referent vowel. The results showed that even when the test involved the same four stimuli, generalization from the prototype toward the nonprototype occurred readily, and generalization from the nonprototype toward the prototype did not. This directional asymmetry suggests that the prototype is an especially powerful perceptual anchor for the category; it pulls other stimuli toward the center of the category, effectively shortening the perceptual distance between stimuli at the outskirts of the category and the prototype center. We have thus referred to it as a perceptual magnet.

What might be the origins of these effects? How do infants come to share adults' definition of a ''good'' stimulus? How do prototypes get into the mind of the baby?

One possibility is that the effects seen here are inherent to human auditory perceptual systems. It is conceivable that certain locations in vowel space are particularly appropriate for the locations of vowel-category centers, because basic auditory perception works in such a way as to minimize the perception of stimulus differences in those locations (K. N. Stevens, 1972, 1989). In other words, certain sound patterns could produce this pattern of results due solely to basic auditory considerations. If this were the case, it would in some respects mirror the situation in color vision, where it has been reported that regardless of culture, adults and infants respond to certain hues as ''focal'' wavelengths for the basic color categories (Bornstein, 1981).

One test of the hypothesis that the perceptual magnet effect is attributable to inherent auditory abilities is to examine the effect in a nonhuman primate, one whose auditory abilities are well matched to those of man, such as the monkey. Previous work has demonstrated that effects such as categorical perception can be replicated in monkeys (Kuhl, 1987, 1988). Such results illustrate that the perception of discontinuities (boundaries) on speech--sound continua are not unique to man. The results reported here do not concern category boundaries, but rather the internal organization of a category and the definition of its ''center.'' Do both humans and monkeys demonstrate the perceptual magnet effect, or do human infants and monkeys differ in this respect?

EXPERIMENT 4

The purpose of Experiment 4 was to replicate the speech prototype test with monkeys, using the same stimuli and similar methods and procedures. Only minor changes were necessitated in the response (a key lift instead of the buttonpress used for adults or the head-turn used for infants) and the reinforcer (food instead of the visual reinforcer used for adults and infants).

Method

Subjects

Six male juvenile monkeys (Rhesus macaques) served as subjects. They were between 1 and 3 years of age. The animals were housed in individual cages 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 the normal ration of food daily at the completion of the experimental session.

Stimuli

The stimuli were exactly the same ones used in Experiments 2 and 3.

Apparatus

The experiment was conducted in a double-walled soundproof booth. During testing, the animals sat in primate chairs. Audio signals were delivered by computer (DEC 11-23) to 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 at midline. An automatic feeder under computer control delivered 2 cc of applesauce through a rubber tube located near the animal's mouth. The delivery of sound and all of the appropriate contingencies during the experiment were under computer control.

Procedure

The design of the experiment with the monkeys mirrored that conducted with the human infants. The exceptions were that the monkeys responded to the detection of a sound change by releasing a response key rather than turning their heads and were reinforced with applesauce rather than with the visual stimulus of a dancing toy animal. Because monkeys were testable for a longer period of time, they were tested in both the P and NP conditions in counterbalanced order.

The procedure used to test the monkeys was the same as that used in previous experiments of animals' perception of speech stimul in this laboratory and have been described in detail (Kuhl & Padden, 1982, 1983). Briefly, the monkey initiated trials by depressing a telegraph key. As soon as the key was depressed, the presentation of stimuli began. The monkey indicated detection of a sound change by releasing the key. Each animal was tested for 1 h each day.

As in the previous tests on adults and infants, two kinds of trials were run with equal probability, *test* and *control*. On all trials, fou stimuli were presented. During *control* trials, the referent vowe (either the prototypic /i/ or the nonprototypic /i/) was presented four times. During *test* trials, the first two stimuli consisted of two repetitions of the referent vowel and the second two stimuli consisted of two repetitions of one of the 32 variants that surrounded the referent.

As was the case with the tests on human infants, these two kind of trials had four potential outcomes. On test trials, a *hit* was scored if the monkey released the telegraph key during the 1.9-sec trial interval (timed from the beginning of the third stimulus); a *miss* was scored if the monkey failed to release the response key during this same period. On control trials, a *correct rejection* was scored if the monkey refrained from releasing the key for the full duration of the 1.9-sec trial interval; a *false positive* was scored if the monkey incorrectly released the key during this same period. Both hits and correct rejections were reinforced with a squirt of applesauce. Miss and false-positive responses were never rewarded with food, and a 7-sec time-out period occurred. A time-out period also occurred if the monkey released the response key during the presentation of the first two stimuli on either a test trial or a control trial (an *early release* response).

Test Phases

As in the tests on human infants, the tests on monkeys consisted of two phases: the training phase (consisting of conditioning and discrimination) and the generalization phase. These phases were identical to those used in the tests on human infants. The conditioning phase continued until the monkey learned to lift the key when the sound change occurred; the discrimination phase continued until the monkey scored above 75% correct for three consecutive test sessions. No differences in the number of sessions required to meet the three-session criterion was seen for monkeys tested in the P condition first, relative to those tested in the NP condition first.

When the three-session criterion was met, the experiment progressed to the generalization phase. During generalization, each test trial consisted of the presentation of one of the 32 variants surrounding the referent vowel (P or NP). As in the previous tests on infants and adults, each of the variants was presented on two different test trials for a total of 64 test trials. An additional 64 control trials were also randomly interspersed, bringing the total number of trials in the generalization test to 128.

Results

In each of the two conditions monkeys achieved overall percent-correct scores that were significantly above chance (50% correct) [both *t*s(5) > 4.5, *p*s < .001]. However, unlike the case for either human adults or infants, the overall percent-correct measure for monkeys revealed no difference between performance in the P and NP conditions (Figure 4). When tested on the P stimuli, monkeys performed as accurately (69.2% correct) as they did when tested in the NP group (71.5% correct) [$t(10) = -.908$, $p > .40$]. In the measure of overall percent correct, therefore, monkeys did not replicate the result obtained with human adults and infants.

Initial inspection of monkeys' generalization scores also revealed a very different pattern of results from those seen with adults and infants (Figure 5). As expected, the effect of distance was highly significant for the monkeys; generalization decreased as the stimuli moved from O_1 to O_4. However, the data also suggested an effect of direction of formant frequency change. Recall that the stimuli formed eight vectors around the referent vowel (Figure 1); the stimuli on each vector varied in the direction of change in the first and second formant. Neither the adult nor the infant data suggest any effect of vector on performance. In fact, the adult and the infant data revealed a striking degree of symmetry across vector, paralleling the goodness ratings seen in Figure 3. For the monkeys, this was not found. Monkeys had more difficulty detecting stimulus changes that involved decreases in the first or second formant frequencies than they did in detecting stimulus changes in the positive direction; this produced differences in the degree of generalization across vector.

To examine these factors, monkeys' generalization scores were submitted to a four-way ANOVA comparing order (P first vs. P second), condition (P vs. NP), distance (O_1–O_4), and Vector (V_1–V_8). The effect of order was not significant [$F(1,4) = 3.2$, $p > .15$], and the interactions of other factors with order were not significant (*p*s > .20), indicating that performance was unaffected by order. Of main interest was the effect of condition (P vs. NP) and the condition × distance interaction. Both were significant in the adult and infant data, but neither of the two was significant for the monkeys (*p*s > .20).

The two factors that affected monkeys' performance were distance and vector. The monkeys showed a highly significant effect that was due to distance [$F(3,12) = 34.3$, $p < .001$] and a highly significant effect of vector [$F(7,28) = 3.5$, $p < .01$], indicating that monkeys' generalization functions decreased with distance, as expected, but were asymmetrical across vector. The vector × condition interaction was also significant due to the fact that, in the P condition, the monkeys generalized more toward stimuli on Vectors 5, 6, and 7, whereas in the NP condition, generalization was greatest for stimuli on Vectors 4, 5, and 6.

Discussion

The results of Experiment 4 demonstrated that monkeys' perception of speech sounds was unaffected by category goodness. There was no speech prototype effect for the monkeys; generalization around the prototype and the nonprototype was equal. For the monkeys, generalization to the variants surrounding the prototype and the nonprototype was dictated by: (1) the psychophysical distance between the referent vowel and the surrounding variants, and (2) a factor that had no influence on the performance of adults and infants, the direction of formant frequency change (vector). The effects of distance and vector were exhibited equivalently across the prototype and the nonprototype conditions. The positive distance and vector effects are important because they show that the monkeys were responsive in the task; however, the fact that their responses to distance and vector were wholly unaffected by whether the prototype or the nonprototype served as the referent provides evidence that typicality played no role in their perception of stimuli from the category.

Human listeners were tested in one of the two conditions (P or NP), whereas the monkeys were tested in both conditions with the order of conditions counterbalanced across subjects. The statistical tests indicate that monkeys' responses were not influenced by order effects; it is therefore appropriate to compare the data gathered in the two different experimental designs. Moreover, in a variety of other speech perception experiments conducted in this laboratory using the within-subjects design, animals have provided evidence of results matching those of humans (see Kuhl, 1987, for review). It would not appear to be the case, therefore, that the within-subjects aspects of the design determines whether or not animals replicate the effects shown by human listeners in tests of speech perception. The inference is that the differences between human and animal performance shown on this test of speech prototypes reflect differences in the mechanisms/processes tapped by the two speech tasks.

The lack of an effect of typicality in animals suggests that basic auditory processes—ones inherent to auditory perceptual processing and common to animals and humans—do not underlie the perceptual magnet effect. This difference is of interest in light of previous research demonstrating many commonalities between young infants and animals in tests of speech perception, particularly those involving categorical perception (Kuhl, 1981; Kuhl & Miller, 1975, 1978; Kuhl & Padden, 1982, 1983; Morse & Snowdon, 1975; Waters & Wilson, 1976). We have thus established a point of dissociation between infants and animals in tests of speech perception. This will be useful for theory construction.

GENERAL DISCUSSION

The results of these experiments strongly suggest that phonetic categories are internally structured for human adults and human infants, but that they are not similarly structured for nonhuman animals. The data address three issues on the perception and representation of speech: (1) the cognitive organization and representation of speech categories by adults—specifically, whether or not speech categories are internally structured and the role the category prototype plays in structuring the category, (2) the perception and representation of speech categories ontogenetically—the specific issue here being the origin of speech category prototypes in early infancy, and (3) speech perception from a phylogenetic perspective—specifically, whether or not speech prototypes are evidenced in nonhuman animals.

Consider first the implications of the findings on adults. In Experiment 1, adults showed that they were able to provide subjective ratings of the overall adequacy or quality of a speech stimulus. Their ratings were uniformly symmetrical around a particular location in /i/ vowel space—a kind of "hot spot" that received consistently high ratings across listeners. But this finding alone, while suggesting that not all /i/s are alike, did not reveal the effect that category goodness had perceptually. Experiment 1 simply revealed that stimulus quality was graded.

Moreover, the results of Experiment 1 suggested that category prototypes for speech are mentally represented in adults. The adults were not provided with a model on which to base their judgments of category goodness; they had to rely on an internal standard of the "ideal" /i/ vowel (see also Kuhl, Williams, & Meltzoff, in press, for a further examination of adults' mental representation of speech units.) The fact that adults' ratings were so consistent suggests that adult listeners, at least those who speak the same dialect of English, have an internal standard for the vowel /i/ that is quite similar.

Experiment 2 addressed the effects of stimulus goodness on perception. Was adults' perception of phonetic stimuli affected by typicality? In these tests, the listeners discriminated members of the same vowel category using two different referent vowels, the prototype and the nonprototype. The results showed that adults' perception of speech stimuli was strongly affected by typicality: When the prototype of the category served as the referent, the other members of the category were perceived as more similar to it. The prototype perceptually assimilated near neighbors in the category, effectively reducing the perceptual distance between it and the other members of the category. The nonprototype of the category did not function in this way. The inference is that adults' speech categories are internally structured. Moreover, the data suggest that, for adults, category organization depends on goodness. Thus, the prototype of a speech category plays a special role in perception—it functions like a perceptual magnet.

This perceptual magnet effect may account for other findings in adult and infant speech perception, such as those related to the perception of sounds from a foreign language. A number of studies in adults have focused on the perception of speech sounds from a foreign language and on adults' attempts to learn a second language. These studies suggest that a speech sound from a foreign language that is similar, but not identical, to one in the subject's native language is perceptually assimilated to the native-language sound (Best, McRoberts, & Sithole, 1988; Flege, 1987). Moreover, research shows that infants tested at 10 to 12 months of age fail to detect the difference between two foreign-language sounds that they could discriminate earlier in life (Werker, Gilbert, Humphrey, & Tees, 1981). The results of these cross-language experiments in adults and infants may be attributable to the perceptual magnet effect demonstrated here for speech prototypes. In other words, the native-language prototype may be acting as a magnet in these two cases to produce the effects reported.

It is also of interest that the results reported here on the functional properties of speech prototypes can be tied to a variety of other findings in the cognitive sciences. For example, in experiments examining the semantic structure of inductive judgments about category members (Rips, 1975), subjects show that the typicality of the member strongly affects subjects' judgments. In Rips's (1975) study, subjects were told that one member of a species had an unknown disease. They were asked to estimate the incidence of the disease in other members of the species. The question was whether or not the representativeness of the instance initially reported as having the disease had an effect on generalization to other members of the species assumed to have the disease. The results showed that representativeness had a potent effect on generalization. Much greater generalization occurred when the initial instance was highly representative of the category as a whole. Related results have also been obtained in the area of social cognition (Weber & Crocker, 1983). In experiments on adults and children using lexical categories, Mervis and Pani (1980) have shown that, when the initial exemplar of a category is a particularly good one, people are likely to generalize appropriately to other members of the category. However, when the initial exemplar of the category is a poor exemplar of the category, people are not likely to generalize appropriately. Thus, studies on topics as diverse as inductive reasoning, social cognition, and lexical categories indicate that the function of representativeness or typicality is increased generalization—prototypes show increased generalization to other members of the category, relative to nonprototypes. This mirrors what I have here called the *perceptual magnet effect* for speech categories. The similarity between the perceptual magnet effect for speech and these other findings suggests that speech prototypes and those in other domains may function similarly. Studying speech prototypes may thus provide a particularly rich and detailed set of examples with which to test more general theories about prototypes and their function in cognition (cf. Medin & Barsalou, 1987; Mervis & Rosch, 1981; Quinn & Eimas, 1986; Rosch, 1975, 1977).

Next, consider the results of Experiments 3 and 4 on infants and monkeys. These tests shed light on the origins and nature of the perceptual magnet effect. The results of Experiment 3 demonstrated that infants' perception of speech categories was also strongly affected by prototypicality; they, too, demonstrated the perceptual magnet effect. The infants whose referent for a speech category was the prototype showed significantly broader generalization to novel members of the category, relative to the infants whose referent was the nonprototype. Moreover, the infants whose referent was the prototype generalized to surrounding vowels in a way that was very highly correlated with adults' category goodness ratings for the same stimuli. Thus, when the adults and the infants used the same ideal /i/ as a referent, their perceptual responses were remarkably similar. In other words, by 6 months of age infant perception of simple vowel quality is very much in line with that of the adult speakers of the language.

The perceptual magnet effect was strongly illustrated in the data on perception of stimuli from the common vector. Recall that the stimuli on the common vector were judged by both groups of infants. This is important because here the only difference is that infants in the prototype group are comparing the stimuli on the vector to the prototype /i/, whereas infants in the nonprototype group are comparing the stimuli to the nonprototype /i/. The infants showed directional asymmetries in their abilities to generalize to the stimuli located between the prototype and the nonprototype vowel; the good vowel assimilated these intermediate stimuli to a significantly greater degree than did the poor vowel.

The issue raised by these results is a developmental one: Where do vowel prototypes in 6-month-old infants come from? The findings on monkeys tend to rule out a basic-auditory-process explanation for the results seen in human infants. Two alternatives remain. First, infants at birth could be biologically endowed with mechanisms that define vowel prototypes for certain vowels (e.g., the "quantal" vowels) or for all of the vowels in all languages of the world. An experimental outcome supporting this view would be one in which human infants demonstrated the prototype effect for vowels they had never heard, such as those used in a foreign language.

A second alternative is that the effects are due to experience in listening to a specific language. Six-month-old infants have had considerable experience in listening to their native language, and the speech directed toward infants, often called "motherese," is special in many respects (Fernald, 1985; Fernald & Kuhl, 1987; Grieser & Kuhl, 1988). Infants may already have begun to organize their speech categories as a function of that experience. Effects of linguistic experience have previously been reported in 12-month-old infants (Werker et al., 1981; Werker & Tees, 1984); however, if the perceptual magnet effect discussed here is based on linguistic input, it would provide evidence that linguistic exposure affects infants' organization of phonology at a much earlier age than heretofore thought. A prediction emerges from the language-experience hypothesis: Infants would be expected to demonstrate the magnet effect only for vowels in their native language. A test of this hypothesis is currently underway in a collaborative experiment being conducted between researchers in Seattle and Sweden, and our preliminary results suggest that the perceptual magnet effect is strongly affected by linguistic experience (Kuhl, 1990; Kuhl, Williams, Lacerda, K. N. Stevens, & Lindblom, 1991). These new cross-language results support the hypothesis that by 6 months of age infants have had sufficient experience with the ambient language to form representations of at least some of the vowels of their native language.

Finally, the data provided here show that the phonetic categories of young infants are structured in a way that diverges from that of monkeys. The present test was different from ones used previously in comparing infants' and animals' reactions to speech signals (Kuhl, 1988). The present experiments tapped the internal structure and the psychological organization of speech categories—the "centers" of speech categories—whereas previous tests focused on the "boundaries" between speech categories. The suggestion provided here is that certain boundaries of speech continua may be inherent to basic auditory processing, but that the centers providing internal organization to speech categories require more than this. Isolating the particular level at which differences exist between species helps clarify both the evolutionary foundation and the subsequent emergence of the special quality of the speech system. The study of speech prototypes may thus provide a particularly fruitful way to examine the "initial state" of the speech code and reveal how this, in conjunction with exposure to the ambient language, leads to the adult's mature species-specific representation of speech.

REFERENCES

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 345-360.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.

Bornstein, M. H. (1981). Two kinds of perceptual organization near the beginning of life. In W. A. Collins (Ed.), *Aspects of the development of competence: The Minnesota Symposia on Child Psychology* (Vol. 14, pp. 39-91). Hillsdale, NJ: Erlbaum.

Fant, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior & Development*, **8**, 181-195.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior & Development*, **10**, 279-293.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, **15**, 47-65.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

Goldman, D., & Homa, D. (1977). Integrative and metric properties of abstracted information as a function of category discriminability, instance variability, and experience. *Journal of Experimental Psychology: Human Learning & Memory*, **3**, 375-385.

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*, **24**, 14-20.

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, **25**, 577-588.

Jakobson, R., Fant, C. G. M., & Halle, M. (1969). *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, **67**, 971-995.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, **66**, 1668-1679.

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America*, **70**, 340-349.

Kuhl, P. K. (1985a). Categorization of speech by infants. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming, buzzing confusion* (pp. 231-262). Hillsdale, NJ: Erlbaum.

Kuhl, P. K. (1985b). Methods in the study of infant speech perception. In G. Gottlieb & N. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 223-251). Norwood, NJ: Ablex.

Kuhl, P. K. (1986). Reflections on infants' perception and representation of speech. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 19-30). Hillsdale, NJ: Erlbaum.

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In P. Salapatek & L. Cohen (Eds.), *Handbook of infant perception: From perception to cognition* (Vol. 2, pp. 275-382). New York: Academic Press.

Kuhl, P. K. (1988). Auditory perception and the evolution of speech. *Human Evolution*, **3**, 19-43.

Kuhl, P. K. (1990). Towards a new theory of the development of speech perception. In H. Fujisaki (Ed.), *Proceedings of the International Conference on Spoken Language Processing* (Vol. 2, pp. 745-748). Tokyo: The Acoustical Society of Japan.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, **190**, 69-72.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, **63**, 905-917.

Kuhl, P. K., & Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics*, **32**, 542-550.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America*, **73**, 1003-1010.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1991). *Linguistic experience affects phonetic perception by 6 months of age*. Manuscript in preparation.

Kuhl, P. K., Williams, K. A., & Meltzoff, A. N. (in press). Crossmodal speech perception in adults and infants using nonspeech auditory stimuli. *Journal of Experimental Psychology: Human Perception & Performance.*

Medin, D. L., & Barsalou, L. W. (1987). Categorization processes and categorical perception. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 455-490). New York: Cambridge University Press.

Mervis, C. B., & Pani, J. R. (1980). Acquisition of basic object categories. *Cognitive Psychology*, **12**, 496-522.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, **32**, 89-115.

Miller, J. L. (1977). Properties of feature detectors for VOT: The voiceless channel of analysis. *Journal of the Acoustical Society of America*, **62**, 641-648.

Miller, J. L., Connine, C. M., Schermer, T. M., & Kluender, K. R. (1983). A possible auditory basis for internal structure of phonetic categories. *Journal of the Acoustical Society of America*, **73**, 2124-2133.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, **46**, 505-512.

Morse, P. A., & Snowdon, C. T. (1975). An investigation of categorical speech discrimination by rhesus monkeys. *Perception & Psychophysics*, **17**, 9-16.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, **85**, 172-191.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, **24**, 175-184.

Quinn, P. C., & Eimas, P. D. (1986). On categorization in early infancy. *Merrill-Palmer Quarterly*, **32**, 331-363.

Repp, B. H. (1977). Dichotic competition of speech sounds: The role of acoustic stimulus structure. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 37-50.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning & Verbal Behavior*, **14**, 665-681.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, **7**, 532-547.

Rosch, E. (1977). Human categorization. In N. Warren (Ed.), *Studies in cross-cultural psychology* (pp. 1-49). London: Academic Press.

Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, **31**, 307-314.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 51-66). New York: McGraw-Hill.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, **17**, 3-45.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America*, **8**, 185-190.

Waters, R. S., & Wilson, W. A., Jr. (1976). Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. *Perception & Psychophysics*, **19**, 285-289.

Weber, R., & Crocker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. *Journal of Personality & Social Psychology*, **45**, 961-977.

WERKER, J. F., GILBERT, J. H. V., HUMPHREY, K., & TEES, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, **52**, 349-355.

WERKER, J. F., & TEES, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, **7**, 49-63.

NOTES

1. The internal representations of speech used by the listener are referred to as *prototypes* in this paper. In doing so, I do not mean to make a statement about the specific form that these speech representations take. Speech representations could consist of average or modal values abstracted from the array of stimuli heard by listeners, or category "ideals," or even specific individual instances that have been experienced by listeners (see Medin & Barsalou, 1987, for a discussion of alternative conceptualizations). The central goal of this work is not to decide these issues, but to examine how speech representations function in perception.

2. In Grieser and Kuhl (1989), the mean number of trials to criterion was significantly higher for the NP group. We hypothesized that it might be due to the difficulty in remembering a nonprototype stimulus. In that study, the initial comparison stimulus presented to the NP group was the prototype /i/, a potentially more memorable stimulus than the NP stimulus that served as the referent. In the present experiment, infants in the NP group heard one comparison stimulus from O_4 during the conditioning phase, and then all eight stimuli from O_4 during the discrimination phase, giving them a range of comparison stimuli to listen to. This change in the procedure was aimed at making the two groups more comparable at the start of the generalization test, and it appears to have been effective in that there were no significant differences between the two groups of subjects in the number of trials taken to reach criterion in either the conditioning or the discrimination phases of the experiment.

(Manuscript received January 4, 1989;
revision accepted for publication March 22, 1991.)