Developmental Psychology
1989, Vol. 25, No. 4, 577–588

0012-1649/89/$00.75

Categorization of Speech by Infants: Support for Speech-Sound Prototypes

DiAnne Grieser and Patricia K. Kuhl
University of Washington

This study tested 6-month-old infants' categorization of speech stimuli to investigate whether infants organize speech categories around "prototypes." In Experiment 1, infants first discriminated single "good" exemplars from two different vowel categories. They were then tested with 64 novel exemplars, 32 from each vowel category. The test stimuli varied in the degree to which they conformed to adult-defined prototypes of the two categories. The results showed that infants correctly sorted the novel stimuli over 90% of the time. In Experiment 2, we trained two groups of infants, one with a good (prototypical) exemplar from a vowel category and the other with a poor (nonprototypical) exemplar. Then we tested both groups with 16 novel exemplars from that same vowel category. Generalization to novel members of the category was significantly greater following exposure to the prototypical exemplar. Results are consistent with a model of speech perception that holds that young human infants organize vowel categories around prototypes. This may contribute to their seemingly efficient processing of speech information, even in the first half year of life.

The ability of infants to group perceptually distinct speech sounds into categories has been demonstrated in a series of experiments (Kuhl, 1979, 1983, 1985a). These studies do not, however, elucidate the underlying structure and organization of speech categories. Might infants' phoneme categories, like infants' visual categories, be represented in terms of prototypes, "good stimuli" that best exemplify the category?

It is characteristic of human perceptual systems that they partition stimuli into cognitively efficient categories that facilitate the storage and retrieval of information. At least in adults, the formation of perceptual categories has been hypothesized to be based on the abstraction of category prototypes (Rosch, 1975). Prototypical stimuli are considered to be the best instances of a particular category; they embody more of the features that are critical to category membership. The prototype model holds that novel stimuli are assigned to perceptual categories on the basis of their degree of resemblance to such category prototypes.

Studies testing the prototype model as the underlying basis for categorization in children and adults have most often used visual stimuli. These studies have shown that subjects abstract category prototypes when presented with category exemplars that vary in stimulus "goodness," that is, stimuli that vary with regard to the degree to which they are prototypical of the category as a whole. They do so both for artificial categories, in which the category prototypes are designated by the experimenter (Garner, 1974; Goldman & Homa, 1977), and for natural categories, where the prototype appears to be inherent in the stimulus (Mervis & Rosch, 1981; Rosch, 1977). For example, the prototypes that represent color categories appear to be universal rather than culturally determined, in that there is cross-cultural agreement as to the "best exemplar" of each category.

A relatively new approach to the study of the perception and organization of categories is the use of infant subjects. Work with 16- to 21-month-old infants shows that they will readily sort physical objects into categories by putting all the objects of one kind into one pile and objects of another kind into a second pile (e.g., Gopnik & Meltzoff, 1987; Sugarman, 1983). Experiments using visual patterns demonstrate that even infants in the first year are capable of categorizing certain stimuli when measures of perceptual recognition are used. These studies show that when infants are visually familiarized with members of a specific category, they do not dishabituate to the presentation of a novel instance from the same category. For example, studies of facial recognition (Cohen & Strauss, 1979; Cornell, 1974; Fagan, 1976) have shown that 7-month-old infants are capable of abstracting general features of faces from a specific facial category (i.e., all male, all female, all same orientation, etc.) and are capable of using that information to generalize to novel photographs sharing those same features.

More recent work has begun to focus on the internal structure and representation of infants' visual categories (see Quinn & Eimas, 1986, for a review). Investigators have shown that young infants presented with exemplars of artificial categories abstract prototypical representations of the categories (Bomba & Siqueland, 1983; Sherman, 1985; Strauss, 1979). For example, Bomba and Siqueland showed that after experience with dot patterns arranged into simple forms, 3- to 4-month-old infants failed to dishabituate to a previously unseen prototype while dishabituating to a novel instance.

The research reported here was supported by grants to Patricia K. Kuhl from the National Science Foundation (BNS 81 03581) and the National Institutes of Health (HD18286). We thank D. M. Padden for valuable assistance in the final stages of data analysis and A. N. Meltzoff for helpful comments on earlier drafts of this article.

Correspondence concerning this article should be addressed to Patricia K. Kuhl, Department of Speech and Hearing Sciences, University of Washington (WJ-10), Seattle, Washington 98195.

Speech is a perceptual domain for which categorization is a critical issue. Understanding how adult listeners assign novel speech tokens to their appropriate phoneme categories, even very simple sounds such as vowels produced by different talkers, has proven very difficult. Computers, for example, have not been programmed to categorize speech sounds despite the fact that a great deal of time and money has been spent trying to achieve that goal (Klatt, 1986). The main problem for computers is that the acoustic features of speech sounds vary when factors such as the talker, the surrounding context, the sound's position in a syllable, or its rate are changed. Despite the acoustic variation introduced by these changes, human listeners identify the resulting sounds as members of the same phonetic category. Although human adults and even young infants can do this (Kuhl, 1979, 1983, 1985a), to date computers cannot.

Studies support the idea that speech-sound categories may be represented by prototypes in adults (Eimas & Miller, 1978; Miller, 1977; Oden & Massaro, 1978; Repp, 1976; Samuel, 1982). This research suggests that some sounds are better category exemplars than others. For example, Repp (1976) showed that certain members of a speech category dominate in dichotic competition experiments. Moreover, the results of selective adaptation experiments show that the magnitude of the adaptation effect depends on the relative goodness of the adapting stimulus (Miller, 1977; Miller, Connine, Schermer, & Kluender, 1983). Finally, experiments in which a number of different cues to speech sounds are covaried add general support to a model that is based on prototypes as opposed to one that is based on the presence or absence of specific features. Oden and Massaro's (1978) "fuzzy logical" model, which bears similarities to Rosch's (1977) "fuzzy" category prototype model and to Strauss's (1979) "averaging" prototype idea, has been supported by a number of experiments using speech (Massaro & Cohen, 1976; Sawusch & Pisoni, 1974).

The issue of speech-sound prototypes has not been addressed in infant studies so far. Previous work has established that, by 6 months of age, infants perceive discriminably different but phonetically identical sounds as members of the same category (Hillenbrand, 1983, 1984; Kuhl, 1979, 1980, 1983, 1985a). For example, Kuhl (1979, 1983) trained 6-month-old infants to respond differentially to single exemplars from two different vowel categories. The two vowels were spoken by the same male talker. Infants were then tested to see whether they would respond correctly to novel, discriminably different instances from the two vowel categories that were spoken by other talkers (females and children). In studies using vowel categories such as /a/, as in *pop,* compared with /i/, as in *peep* (Kuhl, 1979), and /a/, as in *cot,* compared with /ɔ/, as in *caught* (Kuhl, 1983), most infants correctly classified novel instances from the two vowel categories on the first trial. These studies show that infants are capable of correctly categorizing acoustically diverse exemplars of vowels. They treat discriminably different instances of a category as though they belong together. However, these studies do not suggest how infants do it; the studies were not designed to test theories of the underlying basis of infants' categorization abilities. The studies simply showed that speech categorization abilities exist in young infants.

In this report, two studies are described that test hypotheses concerning the internal structure and representation of infants' vowel categories. Specifically, the hypothesis of infant prototypes for speech is tested. Fundamental to that hypothesis is the notion that members of a category vary in goodness and that better members of a category are more representative of the category than others. To test this model, two experiments were carried out using stimuli that were computer-synthesized so that stimulus quality (*goodness,* as defined by adults) could be systematically manipulated. The previous studies on infants' categorization by Kuhl (1979, 1983) used variants of the categories that were all good exemplars. Varying stimulus goodness allowed us to ask two new questions in these experiments. The first (Experiment 1) focused on equivalence classification and categorization: Do infants, after learning to respond to single good exemplars from two different vowel categories, generalize to novel exemplars from the two categories that vary in stimulus goodness? The second (Experiment 2) focused on the prototype issue: Does the relative goodness of the stimulus that serves as the referent for the category differentially affect infants' formation of speech categories? More specifically, does exposure to a prototypical vowel produce broader generalization to other members of the category than does exposure to a nonprototypical vowel?

Experiment 1

Method

The technique used to test equivalence classification was a conditioned head-turn response. Infants were trained to produce a head-turn response when a single good exemplar from Vowel Category 1 was changed to a single good exemplar from Vowel Category 2. Then, infants were tested with novel stimuli from both vowel categories. The novel vowels differed from the training exemplars in controlled steps that resulted in variations in stimulus goodness, as judged by adult speakers of the language. Despite this variation in stimulus quality, adults perceive all variants of a given vowel category as members of that category. The question was whether infants would do the same.

Subjects

Sixteen 6-month-old infants (age range = 5.5–6.5 months; $M = 6.10$) served as subjects. The infants were normal and full-term, with no history of ear problems. An additional 7 infants were eliminated from the experiment for fussiness (2), failure to condition (3), and failure to complete 20 trials (2). Infants were chosen from a computerized subject pool. Their parents had previously returned cards indicating an interest in infant experiments and were subsequently contacted by telephone. Parents were paid $3 per session for up to four sessions of approximately one-half hour each.

Stimuli

Vowel centers and variants. Category centers for the vowel /i/ as in *peep* and /ɛ/ as in *pep* and 32 variants of each were computer-synthesized. The values of the first three formants for the /i/ and /ɛ/ vowel centers were taken from Peterson and Barney's (1952) average of 15 male speakers. The 32 variant stimuli, clustered in four rings around the vowel centers, were created by increasing or decreasing the values of the first two formant frequencies either separately or together. Values for these variants were selected by locating points at four different distances from the centers, along eight directional vectors (see Figure 1).

Use of the mel scale. Formant frequencies for the variants were derived using the mel scale (Stevens, Volkmann, & Newman, 1937). The mel scale is essentially linear at low frequencies and logarithmic at high frequencies and, at least in adults, it equates the magnitude of a perceived change in pitch across different frequencies. Use of the mel scale is important because it allowed us to create variants around two different vowel centers (/i/ and /ε/) that were equated in mels and, thus, equated in psychophysical distance from their respective vowel centers. Formant frequency values were generated by computer, using Fant's (1973) formula relating frequency values to their mel-scale equivalents: $y = k \log (1 + f/1000)$. The values chosen for the variants were created by locating points 30, 60, 90, and 120 mels along vertical, horizontal, and diagonal axes in both positive and negative directions. Figure 1 illustrates the four distances from the vowel centers as well as the eight vectors representing direction. Figure 2 plots the actual locations of the 64 stimuli used in the experiment, 32 variants around the /i/ vowel center and 32 variants around the /ε/ vowel center.

Synthesis. Klatt's (1980) cascade-parallel speech synthesizer, which models the human vocal tract, 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 (F). To create the variants around the two vowel centers, the values of F1 and F2 were manipulated; F3, F4, and F5 remained constant for each vowel category at 3010 Hz, 3300 Hz, and 3850 Hz, respectively. For each 500-ms stimulus, fundamental frequency (F0) began at 112 Hz, rose to 132 Hz over the first 100 ms, and dropped to 92 Hz over the next 400 ms to produce a rise–fall contour.

Loudness balancing. As expected, when synthesized to an equal amplitude value by the synthesizer, the /i/ variants sounded louder than the /ε/ variants (Lehiste & Peterson, 1959). To prepare the stimuli for

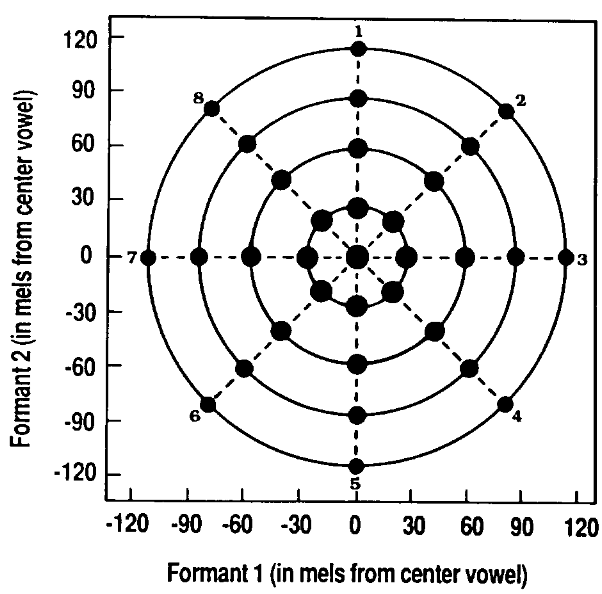

Figure 1. Formant frequency values of F_1 and F_2 for stimuli on four "rings" surrounding the center vowel stimulus. (Formant frequency values are converted to mel equivalents. The mel scale equates for differences in perceived changes in absolute pitch at different frequencies. The 32 variants are located at four distances from the center—30, 60, 90, and 120 mels—in eight different directions.)

presentation to the infants, each author balanced the loudness of the two vowels by listening to an alternating sequence of the good /i/ and the good /ε/, instructing the computer operator to lower the intensity of the /i/ stimulus in 1 dB steps until it was perceived to be equal in loudness to /ε/. Each author made these judgments while seated in the infant's test chair. Both authors' independent judgments showed that a 2 dB decrease in the good /i/ vowel stimulus was necessary to make it sound equivalent to the good /ε/ in loudness. The intensity of the /i/ variants was adjusted to match the intensity of the good /i/, and the intensity of the /ε/ variants was adjusted to match the intensity of the good /ε/. The good /ε/ was used for calibration purposes and was set for presentation at 60 dB SPL, as measured by a sound level meter (BRUEL & KGAER MODEL 2203) that was placed in the chair where the infant would be held by the caretaker.

Equipment and Test Apparatus

The test suite consisted of a sound-treated booth and an adjoining control room, which was arranged in a manner described in full detail elsewhere (Kuhl, 1985b). The booth contained a loudspeaker, a visual reinforcer, a table, 2 chairs, 2 sets of headphones, a button box, and a television camera. The infant, the caretaker, and the assistant sat in the booth. The caretaker, who held the infant, was seated directly across from the assistant. The assistant carefully maintained the infant's attention by manipulating one of several silent toys and held a button box that was used for starting trials. A loudspeaker (ELECTROVOICE SP 12) was located at a 90° angle to the left of the infant. The visual reinforcer, a mechanical animal housed in a dark Plexiglas box, sat on top of the loudspeaker at the infant's eye level. When activated, lights went on within the reinforcer box, and the animal (e.g., a bear pounding a drum, a monkey clapping cymbals) could be seen. A video camera, located directly above the reinforcer, allowed closed-circuit monitoring of the infant's head-turn response. In the control room, the experimenter viewed a Sony television monitor and operated an audiocassette player and a PDP-11/34 computer with terminal and printer. The cassette player was used to provide music over headphones to the assistant and the caretaker so that they could not hear the stimuli and, therefore, could not bias the infant's head-turn responses in any way. The computer presented stimuli to the infant and controlled all of the contingencies, depending on the infant's behavior.

Procedure

Head-turn conditioning technique. The head-turn procedure has been used widely in tests of infant speech perception and has been described previously in detail (Kuhl, 1985b). Briefly, the procedure involves conditioning an infant to produce a head-turn response for a visual reinforcer when a change from one speech sound to another speech sound occurs. Two kinds of trials are run. During test trials, the speech stimulus is changed and infants' head-turn responses are reinforced. During control trials, the speech stimulus is not changed but infants' head-turn responses are monitored. Head-turns that occur on control trials are considered "false-positive" responses and are used to assess the chance probability of head-turning.

This procedure is used to test infants' perception of speech categories by initially training infants on single sounds from two different speech categories and then testing them with novel, discriminably different variants from both of the categories. The hypothesis is that, if infants perceive the similarity among the members of Category 1 and among the members of Category 2, they will treat the novel stimuli from the two categories appropriately. That is, novel stimuli belonging to the category that was initially reinforced will result in head-turn responses, whereas equally novel stimuli belonging to the category that was not reinforced will not result in head-turn responses. The results of two previous experiments supported this hypothesis for vowel contrasts such as /a/ compared with /i/ and for /a/ compared with /ɔ/ (Kuhl, 1979, 1983).

Training phase: Conditioning. The training phase involved two stages: conditioning and discrimination. During conditioning, infants were trained to discriminate between good /i/ and good /ε/. One of the stimuli, either /i/ or /ε/ (counterbalanced across infants), was played from the loudspeaker once every second. This stimulus was the "background stimulus." When the infant was judged to be in a ready state (i.e., looking at the toys held by the assistant, not babbling, fussing, or turned toward the reinforcer), the assistant started a trial. The trial consisted of a 4.5-s observation interval during which the background stimulus changed to a new stimulus, which was repeated four times. This new stimulus was initially paired with the visual reinforcer by activating the reinforcer about halfway through the trial. Because the reinforcer (a mechanical animal) made noise, infants would turn toward it when it was activated. Eventually, infants learned to anticipate the reinforcer and turned as soon as the background sound was changed to the new sound. The criterion for this stage of testing was that the infant produce correct anticipatory head-turns on 3 consecutive trials. Infants who failed to meet the criterion of 3 in a row in 20 trials were not tested further.

Training phase: Discrimination. After meeting the conditioning criterion, infants began the discrimination stage of the training procedure. During discrimination, test trials occurred on half of the trials; the other half of the trials, selected randomly by the computer, were control trials. On test trials, the stimulus was changed, but, on control trials, it was not. During both types of trials, infants were monitored during the 4.5-s observation period to see whether head-turns occurred. Head-turn responses were judged by the experimenter, who watched the infant on the television monitor in the control room. A red light above the television screen was lit, signaling to the experimenter that an observation interval was occurring. During this period, the experimenter watched the infant and pressed a computer key if the infant produced a head-turn response. The experimenter could not hear the stimuli and, therefore, did not know whether the stimuli being presented to the infant came from Category 1 or Category 2, thus ruling out experimenter bias.

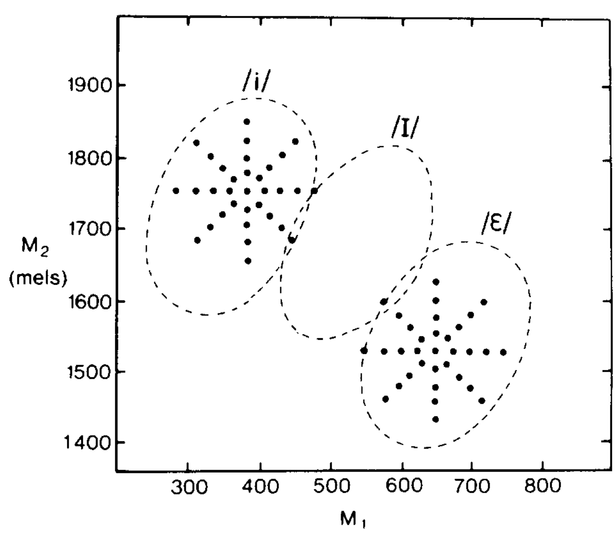

Figure 2. Stimuli for Experiment 1. (Thirty-two stimuli synthesized around each of the two vowels, /i/ and /ε/, converted to mels.)

Table 1
Stimulus Ensembles for Infants in Experiment 1

Subgroup/mel ring	Vector
1	
30	1
30	5
60	2
60	6
90	3
90	7
120	4
120	8
2	
30	2
30	6
60	3
60	7
90	4
90	8
120	5
120	1
3	
30	3
30	7
60	4
60	8
90	5
90	1
120	6
120	2
4	
30	4
30	8
60	5
60	1
90	6
90	2
120	7
120	3

There were four possible outcomes for the two types of trials. On test trials a "hit" was scored if a head-turn occurred, and a "miss" was scored if infants failed to turn. On control trials a "correct rejection" was scored if infants refrained from turning, and a "false positive" was scored if a head-turn occurred. Only correct responses on change trials (hits) were reinforced with the visual stimulus. Infants were tested in the discrimination phase until they met a performance criterion of 9 out of 10 consecutive correct trials.

Categorization test. At this point we considered the infant trained to produce a head-turn response to a single exemplar from Category 1 and to refrain from turning to a single exemplar from Category 2. What we now wanted to know was whether these two responses would generalize to *novel* members of the two vowel categories, particularly on the *first trial* during which the novel stimuli were presented. To test this, novel stimuli from the /i/ and /ε/ categories were presented to infants.

Infants were randomly assigned to one of four subgroups. Each subgroup was presented with 16 of the 64 novel stimuli, 8 from each of the two categories. For each infant to have been presented with all 64 stimuli would have required too many trials from each infant. The 16 stimuli were chosen so that each infant heard stimuli from each of the four mel rings (30, 60, 90, and 120 mels) and from each of the eight vectors. The stimulus set presented to each subgroup is presented in Table 1. A single novel stimulus, repeated four times, was presented during each 4.5-s trial so that infants' head-turn responses could be attributed to a single stimulus. Each of the 16 stimuli occurred on 4 different trials arranged in random order, with the stipulation that all 16 were tested once before any one stimulus could be tested again (a randomized-block design). No more than 3 trials from Category 1 or Category 2 could occur in a row. Each day, between 30 and 40 trials were given in a 30-min test session. At the beginning of each new session, 3 trials, using the original stimuli that were used during conditioning, were run. On these trials, the reinforcer was turned on at the end of the trial, even if the infant had not turned. These served as "refresher" trials and were not included in the data analysis.

Results

The strongest support for categorization comes from data obtained on the first trial, that is, data obtained the first time that each novel stimulus was presented. Previous work by Kuhl (1979, 1983) demonstrated that infants responded correctly on the first trial when the exemplars were all good. In the current experiment, the quality of the exemplars ("goodness") was varied, so we were particularly interested in the effect of goodness on infants' first responses. We also analyzed data beyond the first trial (total-trial data) to assess whether the total-trial results might be stronger than the first-trial data alone.

Accuracy of Responding

Discrimination training. Acquisition of the head-turn response occurred fairly quickly. The number of trials required by the 16 infants to meet the conditioning criterion of 3 in a row ranged between 6 and 16 trials, with an average of 8.5 trials. This rate of response indicates that the /i/–/ɛ/ distinction is easily discriminable by the infants when the single good category exemplars are contrasted. In the discrimination phase, 14 of the 16 infants met the criterion of 9 out of 10 correct on the first day of testing. Trials-to-criterion ranged between 9 (the minimum number required to meet the criterion of 9 out of 10) and 36, with a mean of 20.6 trials.

Categorization test: First trial data. First-trial data were obtained by analyzing the results for the first presentation of each novel stimulus from the two categories. These trials are important because they represent infants' spontaneous tendencies to categorize novel exemplars; they are not affected by reinforcement and, thus, are not the result of learning during the experiment.

The percentage of head-turn responses was subjected to a three-way (2 × 4 × 2) analysis of variance (ANOVA) for trial type (Category 1 vs. Category 2), subgroup (Groups 1–4), and reinforced vowel (/i/ vs. /ɛ/), with repeated measures on the trial-type factor.[1]

As predicted, only the trial-type factor, which distinguished stimuli from the two vowel categories, was significant. Infants produced a large number of head-turns to novel vowel stimuli from Category 1 (92.6%), but produced only a small number of head-turns to novel stimuli from Category 2 (10.2%). Thus, a highly significant main effect of trial type was obtained, $F(1, 8) = 266.6$, $p < 0.001$. Analysis of overall performance across the four subgroups, each of which received one-fourth of the total stimuli, showed that the subgroups performed equally well; thus, the main effect for subgroup was not significant, $F(3, 8) = 0.9$, $p > 0.20$. Infants performed equally well regardless of which stimulus served as the reinforced vowel, as shown by the nonsignificant vowel main effect, $F(1, 8) = 0.01$, $p > 0.20$. No significant interactions for Subgroup × Trial Type, Subgroup × Vowel, or Vowel × Trial Type were shown ($p > 0.20$, in all cases). The Subgroup × Vowel × Trial Type interaction was also nonsignificant ($p > 0.20$). Follow-up analyses were conducted that collapsed the data across subgroup and reinforced vowel to test the effect of trial type and distance. The results showed that the trial type factor was highly significant, $F(1, 15) = 358.5$, $p < 0.001$, and that the distance factor did not significantly affect overall performance, $F(3, 45) = 0.8$, $p > 0.20$. There were significantly more head-turns to Category 1 stimuli than to Category 2 stimuli at all four distances from the center, as shown by the lack of a significant interaction between distance and trial type, $F(3, 45) = 1.8$, $p > 0.10$.

Categorization test: Total-trial data. Highly similar effects were obtained for the total-trial data: 81.7% of the stimuli from Category 1 resulted in head-turn responses, but only 14.5% of the stimuli from Category 2 did. A three-way ANOVA for percentage of head-turns on the total-trial data by trial type, subgroup, and vowel, with repeated measures on the trial-type factor, showed that the main effect for trial type was highly significant, $F(1, 8) = 911.9$, $p < 0.001$. The subgroup main effect was again nonsignificant, $F(3, 8) = 1.6$, $p > 0.20$. Infants performed equally well regardless of which stimulus served as the reinforced vowel, as shown by the nonsignificant main effect of vowel, $F(1, 8) = 0.3$, $p > 0.20$. All two- and three-way interactions were nonsignificant ($p > 0.1$, in all cases). A follow-up two-way ANOVA that collapsed the data across subgroup and reinforced vowel was conducted to test the effect of trial type and distance in mels. The results showed that trial type was highly significant, $F(1, 15) = 498.4$, $p < 0.001$, and that the distance factor was not, $F(3, 45) = 1.4$, $p > 0.20$. The Trial Type × Distance interaction was also not significant, $F(3, 45) = 1.3$, $p > 0.20$.

[1] In a recent article, Hertzog and Rovine (1985) reviewed the statistical treatment of repeated measures data. They have suggested that before the mixed-model analysis of variance (ANOVA) F test is used to analyze repeated measures data, the assumption of "sphericity" (Huynh & Feldt, 1970) should be examined. Because formal tests on variances and sphericity are not especially powerful and are sensitive to nonnormality, Hertzog and Rovine recommended using the correction factor for violations of sphericity, ε, to make decisions about which tests are most appropriate. They suggested that when estimates of ε are greater than 0.9, univariate mixed-model assumptions are either correct or are only trivially violated and, therefore, that mixed-model F tests should be used. When estimates of ε fall below 0.75 they recommended the use of multivariate analysis of variance (MANOVA), which provides a more conservative test of the patterns of significance. For the repeated measures ANOVAs detailed in Experiments 1 and 2, the estimates of ε were all greater than 0.9, so the use of a mixed-model ANOVA is warranted in this case.

Latency of Responding

Mean latencies of hit responses for each condition were as follows: for conditioning, 2.3 s; for discrimination, 2.4 s; and for categorization, 2.2 s. A two-way ANOVA, with repeated measures on the last factor, compared latencies for reinforced vowel (/i/ and /ɛ/) and distance (30, 60, 90, and 120 mels), collapsed across the four subgroups of infants. No significant effects were observed. Mean latencies were similar when /i/ (2.24 s), as opposed to /ɛ/ (2.2 s), served as the background vowel, as shown by the nonsignificant main effect for vowel, $F(1, 15) = 0.3, p > 0.20$. Latencies also did not differ significantly by distance, $F(3, 45) = 1.5, p > 0.20$. The Vowel × Distance interaction was nonsignificant ($p > 0.20$).

Discussion

The present experiment advances our general knowledge of infants' speech categorization abilities by showing that young infants can sort speech stimuli into two categories when the stimuli vary in goodness. The data of Experiment 1 strongly support the conclusion that infants categorize novel /i/ and /ɛ/ stimuli after training on a single good exemplar of each category, even when the novel exemplars are no longer all "good" or prototypical exemplars of their respective categories. Infants appropriately assigned 32 /i/s and 32 /ɛ/s to their respective categories despite variations in whether each exemplar could be considered a prototype of the category by adults. In this sense, Experiment 1 is an extension of previous experiments (Kuhl, 1979, 1983) showing equivalence classification for speech sounds by young infants.

Analysis of the first-trial data was particularly noteworthy. Responses for the total set of trials presented during the experiment can be expected to show some effect of learning, but responses for the first presentation of each stimulus are not subject to the effects of learning. They reflect the infants' natural tendencies to classify the novel vowels correctly in the absence of reinforcement or learning. The data suggested that infants correctly sort the novel stimuli immediately into their appropriate categories.

The facility with which the infants performed this categorization task indicated that they readily perceived the similarity among /i/ variants and among /ɛ/ variants. The fact that neither the first-trial data nor the total-trial data was affected by vowel, subgroup, or distance further underscores the evidence that the infants were consistently picking out between-category differences. It is particularly interesting that infants performed equally well at each distance (in mels) from the background stimulus. Moreover, response latency for producing a head-turn was also similar regardless of distance from the center vowel. Rather than being more accurate or showing shorter latencies for previously experienced training stimuli and for those stimuli that are closer to them in vowel space, infants' classification responses were accurate and rapid no matter how prototypical the variant of the category happened to be.

These data raise an issue with regard to speech categorization by infants: Is each member of the category perceived to be equally good from the infants' perspective? The results of Experiment 1 might lead one to believe so. No differences in accuracy of responding nor in response latency were seen for stimuli varying in goodness. However, we do not know if infants' equivalent treatment of the members of the category is due to the process of speech categorization itself, that is, the existence of a prototype to which variants are assimilated, or whether it is due to infants' inability to hear the differences between the variants. Experiment 2 was designed to address both issues.

Experiment 2

The prototype hypothesis predicts that all members of a category are *not* perceived to be equivalent; some members are perceived to be better exemplars of the category. These prototypical exemplars resemble other members of the category to a greater degree, that is, they are more "redundant" (Garner, 1974). If the prototype model characterizes infants' perception of vowels, then the exemplar infants use as a referent stimulus for the category might make a difference in their formation of the category. Specifically, we hypothesized that if the referent stimulus were a good exemplar of the vowel, infants might be expected to show greater generalization to other members of the category than if a poor vowel exemplar had served as the referent stimulus. In other words, the prototype hypothesis was that a good exemplar would assimilate more novel variants of the vowel category than would a poor exemplar. Alternatively, if infants perceived all /i/ vowels as equivalent, then generalization to novel members of the category after exposure to a good as opposed to a poor exemplar of /i/ should be equal.

In Experiment 2, infants' generalization to other members within the same vowel category was tested under two conditions: (a) when infants were trained with a good /i/ as the background (referent) stimulus and were reinforced for discriminating it from other members of the category and (b) when they were trained with a poor /i/ as the background (referent) stimulus and were reinforced for discriminating it from other members of the category. This allowed two questions to be asked. First, are infants capable of detecting within-category differences? Second, are some exemplars better than others in that they represent a greater number of category members?

Method

Subjects

A new group of 32 normal full-term infants was tested, 16 in each of two conditions, the prototype and nonprototype conditions. The infants ranged in age from 5.5 to 6.5 months old ($M = 6.08$). An additional 15 infants (7 from the prototype group and 8 from the nonprototype group) were eliminated from the experiment: 9 infants could not be conditioned, and 6 infants did not complete the experiment. This attrition rate is approximately equal to that observed in Experiment 1, where 7 out of 23 infants were eliminated from the experiment. Each infant participated in from one to five sessions. Parents were paid $3 for each session.

Stimuli

Good and poor vowel stimuli. Two stimuli were chosen from the pool of 32 /i/ stimuli synthesized for Experiment 1. One of these stimuli, the good /i/, was the training stimulus used in Experiment 1. This stimulus was based on the average center formant frequencies for the vowel /i/ specified by Peterson and Barney's (1952) male speakers, which was perceived as a good representative of the /i/ vowel category. The poor /i/ stimulus was located nearer the /i/ boundary but still well within the category. That is, although this stimulus was a poor representative of the /i/ vowel category, it was definitely an /i/ rather than any other vowel.

Variants around good /i/ and poor /i/. A set of 32 variant stimuli was computer-synthesized, 16 around the good /i/ and 16 around the poor /i/. The 16 variants of the good /i/ were a subset of those created for Experiment 1. These 16 stimuli were located at four distances from the good /i/, at 30, 60, 90, and 120 mel intervals, along four of the eight vectors used in Experiment 1. The poor stimulus was one of these variants. Around this poor stimulus, 16 additional variants were synthesized at four distances along four vectors, so that the vowel space occupied by the good and poor sets of stimuli overlapped. The result was that both the good /i/ stimulus and the poor /i/ stimulus had a semicircle of 16 /i/ variants at equivalent mel distances from their respective centers (see Figure 3). Four stimuli along one vector were shared by the two sets of stimuli.

Synthesis. The stimulus values for the first and second formants were derived by locating the intersection of the first two formants on a mel-scaled vowel space and reconverting these values to hertz using a computer program. The stimuli were then synthesized using the Klatt (1980) synthesis program, in the same manner as described for Experiment 1. The values of the other stimulus parameters (amplitude contour, fundamental frequency contour, and formant bandwidths) were identical with those for Experiment 1. Each stimulus was 500 ms long.

Adult goodness ratings. Because we were interested in adults' responses to each of these stimuli, quantitative ratings were obtained from adults to each stimulus. Eight university students with educational backgrounds in phonetics listened to each of the stimuli in the prototype and the nonprototype sets, presented in random order. They provided numerical ratings of goodness on a scale from 1 to 7, with 1 indicating a poor exemplar and 7 indicating a good exemplar of the /i/ category. The adults were initially familiarized with the full set of stimuli, but no reference vowel was provided ahead of time to illustrate a good stimulus. *Good* was simply defined as sounding "natural and nondistorted."

Each adult sat in the laboratory and listened to the stimuli one at a time in a computer-controlled task. The adults listened over the same loudspeaker used in the infant tests. They circled a number from 1 to 7 after each stimulus was presented and pressed a button to present the next stimulus. Each adult judged each stimulus five times. The average goodness ratings for each stimulus are plotted in Figure 4. The size of the circles and squares indicates the relative goodness. As shown, the good /i/ (prototype) stimulus was given an average rating of 7 and the poor /i/ (nonprototype) stimulus was given an average rating of 1.3. As shown, the average ratings for stimuli near the good /i/ tended to receive the highest ratings, and ratings consistently decreased with an increase in distance from the good /i/. Conversely, stimuli surrounding the poor /i/ received relatively low ratings, with an increase in the perception of goodness as they neared the region of the vowel space shared with stimuli around the good /i/.

These ratings were subjected to a two-way ANOVA with repeated measures on both the stimulus factor (prototype vs. nonprototype) and on the distance factor to determine whether stimulus quality for the prototype and nonprototype stimuli differed significantly. The results showed that the stimuli around the good /i/ received significantly higher ratings than did the stimuli located around the poor /i/, $F(1, 7) = 163.8$, $p < 0.001$. Distance from the background stimulus also affected average goodness ratings. For both groups combined, the goodness rating tended to decrease as distance from the background stimulus increased, $F(3, 21) = 65.9$, $p < 0.001$. However, when the prototype and nonprototype conditions were compared for the effect of stimulus distance on the goodness rating, a Condition × Distance interaction was observed, $F(3, 21) = 139.4$, $p < 0.001$. The stimuli surrounding the prototype showed systematic decreases in goodness ratings with increases in distance of the stimuli from the background stimulus, whereas the nonprototype condition showed less uniform ratings as a function of stimulus distance.

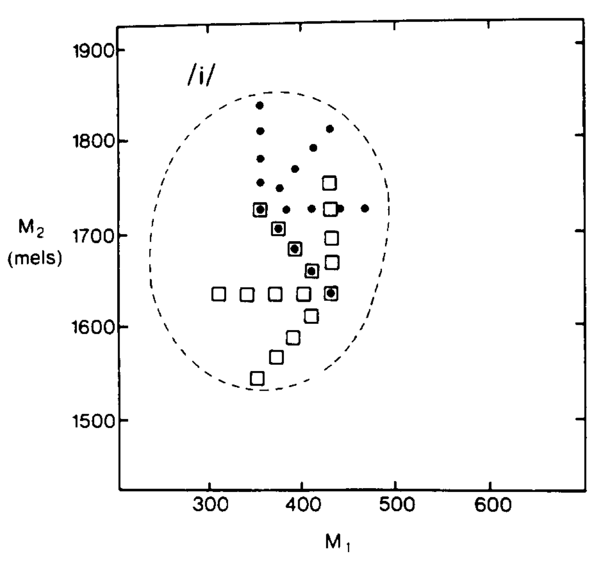

Figure 3. The good /i/ vowel and variants on four vectors emanating from it (dots) and the poor /i/ vowel and variants on four vectors emanating from it (squares). (The stimuli on one vector were common to both sets.)

Procedure

Design. The visually reinforced head-turn procedure described for Experiment 1 was used to test the infant subjects in Experiment 2. The experimental design differed from that of Experiment 1. Rather than being required to "sort" multiple exemplars of two vowel categories, Experiment 2 involved only exemplars of a single vowel category. We wanted to know whether infants who were familiarized with a good as opposed to a poor /i/ would show equivalent generalization to other members of the same category. Thus, two groups of infants were compared. For one group, the good /i/ served as the background stimulus, and all 16 variants around the good /i/ (the prototype stimuli) served as the test stimuli. For the second group, the poor /i/ served as the background stimulus, and all its 16 variants were presented as test stimuli.

The hypothesis was that the prototypical good /i/, being more "redundant," would resemble a greater number of novel category variants. There are two measures that test for this pattern of results, one general and the other more specific. The general measure is the overall percentage-correct discrimination score. This measure reflects the degree to which the good or poor /i/ is discriminated from its surrounding members. The prototype hypothesis argues that the good /i/ resembles its surrounding stimuli more than the poor /i/ resembles its surrounding stimuli, even though the physical distances between each of the target vowels and their surrounding variants were equated. Thus, the hypothesis is that discriminability between the good /i/ and its variants will be poorer than discriminability between the poor /i/ and its variants and, therefore, that the prototype group will show lower overall-percentage-correct discrimination scores.

A more precise comparison between the prototype and nonprototype groups involves the specific pattern of generalization that occurs for the two groups at each distance (30, 60, 90, and 120 mels) from the good or poor /i/. The prototype hypothesis predicts that the good /i/ will resemble stimuli over a greater distance than will the poor /i/. The response measure used in this comparison is infants' "misses." A miss response indicates that the infant did not detect a change from the background vowel (good or poor /i/) to the variant stimulus. We will refer to these as "generalization responses" and calculate a "generalization score" that is the percentage of test trials in which the infant failed to indicate the detection of a stimulus change. The prototype hypothesis predicts that the generalization score will be higher for the prototypical good /i/ vowel than for the nonprototypical poor /i/ vowel.

Training phase. As in Experiment 1, the training phase involved two stages: conditioning and discrimination. Both groups of infants were trained to discriminate the same two stimuli: good /i/ and poor /i/. For the prototype group, good /i/ was the background vowel (the referent stimulus), and poor /i/ was the test stimulus. For the nonprototype group, the reverse situation was true. During conditioning, only test trials were presented until the infant produced three consecutive antici-

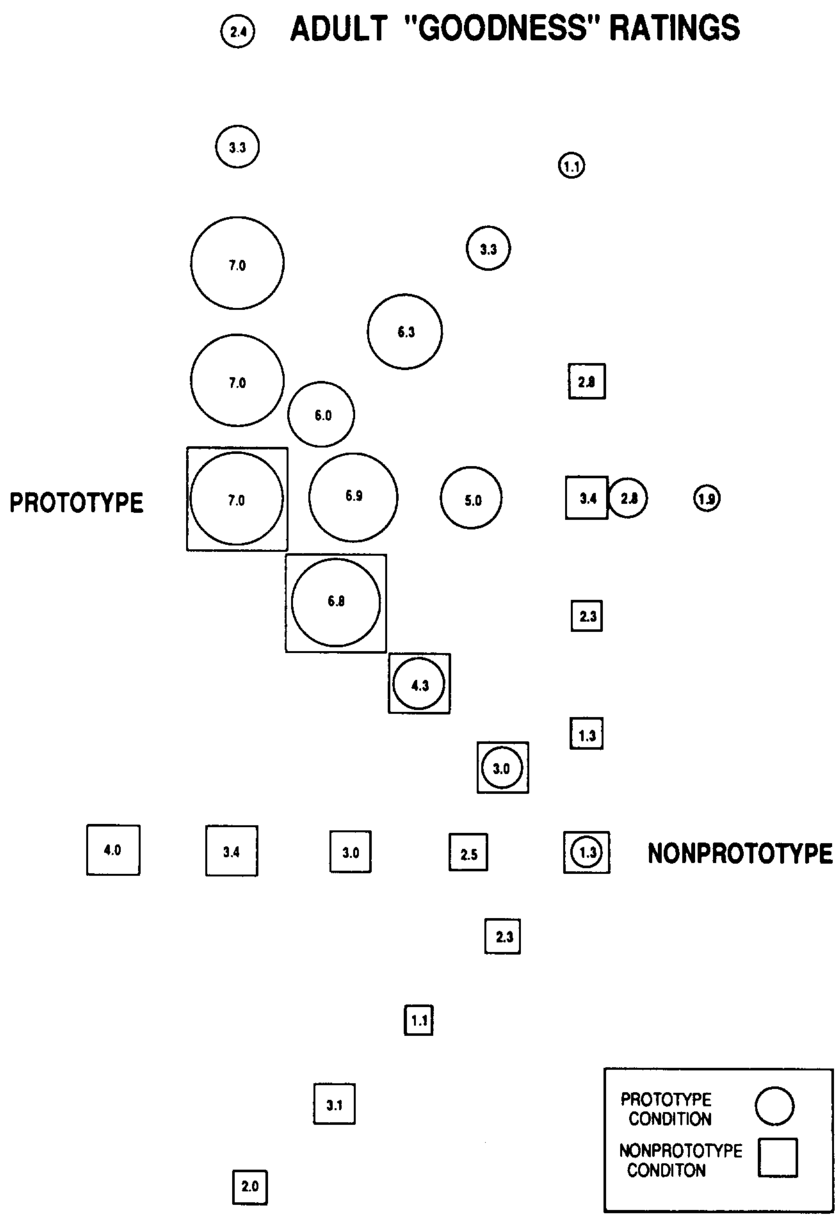

Figure 4. Goodness ratings for the good /i/ vowel (the prototype), the poor /i/ vowel (the nonprototype), and variants on four vectors emanating from each. (Goodness was judged by adults using a scale from 1, *a poor exemplar,* to 7, *a good exemplar.* The size of the circles and squares correlates with the degree of "goodness," with larger forms indicating better exemplars.)

patory head-turns. This criterion had to be met within 35 trials. When the criterion was met, infants progressed to the discrimination stage, wherein both test trials and control trials were run with a 50% probability of occurrence. Criterion performance during this phase was seven out of eight consecutive correct responses within the first two sessions. Infants failing to meet this criterion were not tested further.

Generalization test. In the next phase, generalization, infants' perception of the 16 novel /i/s that surrounded the good (or poor) vowel was tested. Each test stimulus was presented to each infant on 4 different trials (16 stimuli × 4 = 64 test trials), using a randomized-block design, during the course of the generalization stage. On test trials, a single variant was repeated four times during the 4.5-s trial so that infants' head-turn responses could be attributed to a single stimulus. In addition, 64 control trials, during which the background stimulus was presented, were randomly interspersed. No more than 3 test or control trials could occur in a row. Typically, about 40 trials were run each day in a session lasting approximately 30 min. At the beginning of each new session, 3 test trials using the original training stimuli were given. These served as "refresher" trials and were not analyzed.

Results

Training Phase

The mean number of trials-to-criterion during conditioning was 14.0 for the prototype group and 18.5 for the nonprototype group, a difference that was nonsignificant, $t(30) = -1.25$, $p > 0.10$. In the discrimination stage, however, the prototype group reached the criterion of 7 out of 8 in an average of 19.7 trials, whereas the nonprototype group averaged 40.0 trials. This difference was highly significant, $t(30) = 3.96$, $p < 0.001$. This difference may be due to the fact that poor exemplars are initially more difficult to remember, as has been reported for visual stimuli (Pomerantz, Sager, & Stoever, 1977).

Overall Percentage Correct During Generalization

The overall-percentage correct measure—(% hits + % correct rejections)/2—supported the prototype hypothesis. Across both groups, infants scored 71.1% correct, a score significantly different from the 50% chance level, $t(31) = 7.1$, $p < 0.001$, indicating that, in general, discriminating among variants within the same /i/ vowel category is fairly easy for infants. As predicted, however, the nonprototype group scored higher (77.5%) than the prototype group (64.8%), indicating better discrimination of the test stimuli from the background vowel stimulus. A *t* test comparing overall percentage correct for these two groups was highly significant, $t(30) = 9.4$, $p < 0.001$.

Generalization Scores: First-Trial Data

The prototype hypothesis was strongly supported by generalization on the first trial. The generalization score reflects the percentage of test trials on which infants did not produce a head-turn response the first time a novel variant was presented. The data were submitted to a two-way ANOVA, with repeated measures on the second factor, that examined the group (prototype or nonprototype) and distance (30, 60, 90, or 120 mels) effects (see Footnote 1). As predicted, higher generalization scores were obtained by infants in the prototype group (54.7%) than by those in the nonprototype group (46.1%); this difference is shown by the significant effect of group, $F(1, 30) = 5.2$, $p < 0.03$. There were also significant differences in the generalization scores depending on distance of the stimuli from the center vowel stimulus, as shown by the highly significant effect of distance, $F(3, 90) = 36.9$, $p < 0.001$. As expected, both groups tended to show more generalization to stimuli nearest the background stimulus (at 30 mels, 71.9%; at 60 mels, 63.3%; at 90 mels, 43.8%; and at 120 mels, 22.7%). This overall pattern was more pronounced, however, for the prototype group (at 30 mels, 82.8%; at 60 mels, 68.8%; at 90 mels, 46.9%; and at 120 mels, 20.3%) than for the nonprototype group (at 30 mels, 61.0%; at 60 mels, 57.8%; at 90 mels, 40.6%; and at 120 mels, 25.0%). The Group × Distance interaction approached significance, $F(3, 90) = 2.3$, $p < 0.08$, and was due to the greater generalization shown by the prototype group for the 30-, 60-, and 90-mel stimuli.

Generalization Scores: Total-Trial Data

This same pattern of results obtained for the total-trial data, once again providing strong support for the prototype hypothesis. A two-way ANOVA, with repeated measures on the second factor comparing the two groups (prototype and nonprototype) at each distance (30, 60, 90, and 120 mels), showed that the prototype group produced significantly more generalization ($M = 42.2\%$) than the nonprototype group ($M = 35.1\%$), $F(1, 30) = 29.0$, $p < 0.001$. Generalization was also affected by distance from the background stimulus. Across both groups, more generalization occurred for stimuli nearest the background stimulus (at 30 mels, 73.3%; at 60 mels, 40.1%; at 90 mels, 20.6%; and at 120 mels, 22.6%). Thus, the distance main effect was highly significant, $F(3, 90) = 485$, $p = 0.001$. However, the prototype group's generalization scores were much larger, particularly at the 30- and 60-mel distances (at 30 mels, 81.2%; at 60 mels, 50.9%; at 90 mels, 19.0%; and at 120 mels, 19.0%), than those obtained by the nonprototype group (at 30 mels, 65.3%; at 60 mels, 29.3%; at 90 mels, 22.3%; and at 120 mels, 24.2%). Thus, in line with predictions from prototype theory, the Group × Distance interaction was highly significant, $F(3, 90) = 1157.8$, $p < 0.001$.

Because the two groups of infants (prototype and nonprototype) had both been tested on stimuli located on a vector shared by the two groups (see Figure 3), we were very interested in examining the profiles of response for stimuli on the shared vector for the two groups. A polynomial trend analysis was conducted on the Group × Distance interaction. This involved partitioning the three degrees of freedom assigned to the interaction into orthogonal polynomial contrasts, each with one degree of freedom (see Bock, 1975, Chapter 7, for a full description and computational procedures). The results revealed that the profiles of the two groups differed significantly. The generalization function decelerated at a faster rate for infants in the nonprototype group, and this produced a significant difference in the quadratic trend, $F(1, 30) = 4.34$, $p < 0.05$. This finding shows that the two groups of infants responded differently even when they were tested on the exact same stimuli.

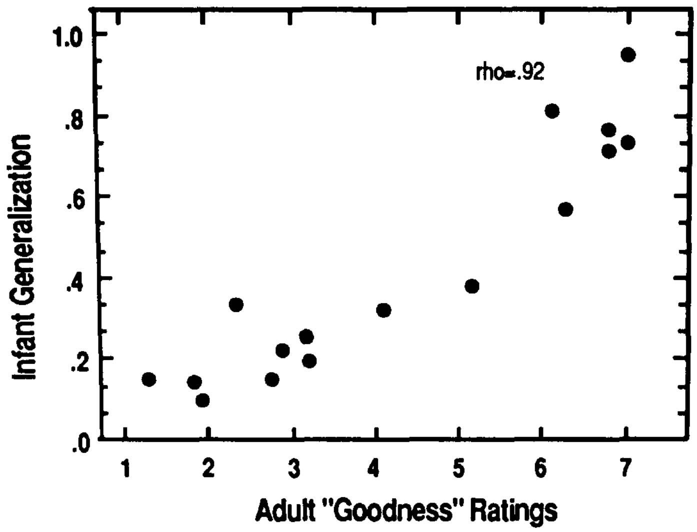

Figure 5. Plot showing the correlation between infants' generalization scores and adults' goodness ratings to stimuli surrounding the good /i/ vowel. (The Spearman ranked correlation is 0.92.)

Comparison of Adult and Infant Data

In order to test whether the gradient of generalization around the good /i/ for infants bore any resemblance to the gradient underlying adults' goodness judgments, a correlation was computed relating the degree of infant generalization to variants around the good /i/ to adult goodness ratings of the same stimuli. The infants' generalization scores and the adults' goodness ratings were assigned ranks and subjected to correlational analysis to determine whether the numerical rankings changed similarly for both measures as a function of distance from the good /i/. As shown in Figure 5, Spearman ranked correlation coefficients showed a high positive correlation (r_s = 0.92) between infant generalization and adult goodness. Both changed systematically as a function of distance and direction from the center stimulus.

Discussion

The results of Experiment 2 confirmed two hypotheses. First, infants' success in categorizing vowel stimuli in Experiment 1 cannot be attributed to an inability to discriminate among the members of a single vowel category. Infants in both the prototype and nonprototype groups were quite successful in discriminating their respective /i/ vowels from other vowels in the category. Both groups scored significantly above chance. In other words, vowel stimuli from a single category are not perceptually indistinguishable by infants.

Second, and more important, the results of Experiment 2 demonstrate an effect of stimulus goodness and thereby support prototype theory. Stimuli defined by adults as better exemplars of the category resulted in greater generalization to other members of the same vowel category by infants. In fact, a high correlation (0.92) was observed between adult judgments of the relative goodness of particular /i/ vowels and infants' patterns of head-turn responses to these same stimuli. This suggests that whatever causes the gradual decay in quality as you move from the good stimulus toward the poor one is similar for adults and infants.

General Discussion

In the two experiments reported here using speech sounds, human infants have demonstrated the two distinct phenomena that embody categorization: equivalence classification and prototype effects. Experiment 1 demonstrated equivalence classification. The phenomenon of equivalence classification requires that discriminably different stimuli belonging to a single category are treated as equivalent (Bornstein, 1981; Kuhl, 1983). Infants in Experiment 1 demonstrated equivalence classification by categorizing 32 novel /i/s and 32 novel /ɛ/s into their respective vowel categories after being trained on a single good exemplar from each of the two vowel categories. Infants classified the novel vowels correctly on the first trial over 90% of the time.

This result replicated previous work demonstrating infants' abilities to categorize novel vowels from two vowel categories immediately after being trained on a single exemplar from each of the two categories (Kuhl, 1979, 1983). However, this result extended the previous findings by showing that equivalence classification takes place when the novel variants from the two vowel categories vary in relative goodness. In previous studies, the novel variants were all good exemplars. In the current study, the goodness of the variants was systematically varied by manipulating the formant frequencies of the vowels. Adult listeners rated each vowel and confirmed that goodness systematically varied. Yet infants treated all stimuli as the same, showing no significant differences in head-turn accuracy nor in response latency to stimuli that were less good.

Equivalence classification requires attention to the criterial differences between categories and disregard for the differences among stimuli that belong to a single category. During equivalence classification, perceivers focus on the fundamental equivalence among members of the category. Apparently, the design of Experiment 1 focused infant attention on category differences, and this served to minimize within-category differences. Two aspects of the design are probably critical. First, stimuli from two different categories are presented to infants, rather than stimuli from just a single one, and this probably helps to focus infant attention on the criterial differences that separate the two categories. Second, infants in Experiment 1 were trained on good exemplars of the two categories, and this supports the formation of equivalence classes; the results of Experiment 2 suggest that a good exemplar may enhance infants' categorization abilities by providing a kind of perceptual anchor for each category.

The second phenomenon embodied in adult categorization studies, and demonstrated in Experiment 2, is the effect of a prototype. Prototypicality requires perceivers to focus on qualitative differences among members of the category. It is reflected in the tendency to regard some stimuli as "better" instances of the category than others and in the tendency for these prototypical stimuli to resemble a greater number of members of the category. Prototype effects are best seen when members of the same category are compared, as in Experiment 2. Here, infants were encouraged to demonstrate their abilities to distinguish among members of the same category. This, in turn, allowed us to test whether the relative goodness of a particular stimulus affects categorization. The results showed that generalization to other members of a vowel category was significantly altered by the goodness of the stimulus on which infants were trained. Infants trained on the good vowel stimulus showed generalization to a larger number of novel members of the vowel category than infants trained on the poor vowel stimulus. Thus, infants in Experiment 2 provided data supporting the idea that they organize speech sounds around prototypes.

The next important issue raised by the data is a developmental one. What is the developmental origin of vowel prototypes?

There are two possible explanations. The first is that vowel goodness is inherently defined by the auditory perceptual system. If this were the case, it would mirror the situation in color vision, where it has been shown that regardless of culture, adults and infants prefer the same focal colors (Bornstein, 1981). Thus, preference for focal colors reflects an inherent prototype. Animals' demonstration of this same effect reinforces the view that primate color vision is physiologically determined (Sandell, Gross, & Bornstein, 1979). Speech-sound prototypes might also be inherently determined. (See Kuhl, 1987b, 1988, and 1989 for discussion.)

An alternative hypothesis is that these effects are the result of experience in listening to a specific language. Six-month-olds have had considerable experience in listening to the sound patterns of English and they may already have begun to assimilate that experience, recognizing the speech-sound patterns that are typical in their particular language environment. Werker and Tees (1984) have shown an effect of linguistic experience in 10- to 12-month-olds. If the prototype effect observed here is attributable to linguistic experience, it would push the period during which infants are sensitive to linguistic experience much earlier. The critical test of this language-experience hypothesis will consist of testing American infants on non-English vowel contrasts to see whether equivalence classification and the prototype effect depend on linguistic experience. These tests are now under way.

It has been argued that the existence of equivalence classes and of prototypes to represent them reflects an ability to organize and summarize a stimulus set, and that this ability increases perceptual and cognitive efficiency (Rosch, 1977). This increased efficiency is particularly important for speech perception in young infants because they have to contend with a broad array of stimuli varying in many different dimensions. We have shown here that infants have both the ability to recognize category equality (equivalence classification) and category gradience (the prototype effect) for phonemic categories. The latter finding demonstrates that 6-month-olds recognize a good vowel stimulus and generalize more completely to others in the category after exposure to a good vowel. This is the first finding in support of speech-sound prototypes in human infants.

References

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

Bomba, P. C., & Siqueland, E. R. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology, 35,* 294–328.

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* (pp. 39–91). Hillsdale, NJ: Erlbaum.

Cohen, L. B., & Strauss, M. S. (1979). Concept acquisition in the human infant. *Child Development, 50,* 419–424.

Cornell, E. (1974). Infants' discrimination of photographs of faces following redundant presentations. *Journal of Experimental Child Psychology, 18,* 98–106.

Eimas, P. D., & Miller, J. L. (1978). Effects of selective adaptation on the perception of speech and visual patterns: Evidence for feature detectors. In R. D. Walk & H. L. Pick (Eds.), *Perception and experience* (pp. 307–345). New York: Plenum Press.

Fagan, J. F. (1976). Infants' recognition of invariant features of faces. *Child Development, 47,* 627–638.

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

Garner, W. (1974). *The processing of information and structure.* New York: Halsted Press.

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 and Memory, 3,* 375–385.

Gopnik, A., & Meltzoff, A. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. *Child Development, 58,* 1523–1531.

Hertzog, C., & Rovine, M. (1985). Repeated-measures analysis of variance in developmental research: Selected issues. *Child Development, 56,* 787–809.

Hillenbrand, J. (1983). Perceptual organization of speech sounds by infants. *Journal of Speech and Hearing Research, 26,* 268–282.

Hillenbrand, J. (1984). Speech perception by infants: Categorization based on nasal consonant place of articulation. *Journal of the Acoustical Society of America, 75,* 1613–1622.

Huynh, H., & Feldt, L. A. (1970). Conditions under which mean square ratios in repeated measures designs have exact *F* distributions. *Journal of the American Statistical Association, 65,* 1582–1589.

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

Klatt, D. H. (1986). Problem of variability in speech recognition and in models of speech perception. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 300-324). Hillsdale, NJ: Erlbaum.

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. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 2, Perception* (pp. 41–66). New York: Academic Press.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development, 6,* 263–285.

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. A. 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. (1987a). Perception of speech and sound in early infancy. In P. Salapatek & L. B. Cohen (Eds.), *Handbook of infant perception. Vol. 2. From perception to cognition* (pp. 275–382). Orlando, FL: Academic Press.

Kuhl, P. K. (1987b). The special-mechanisms debate in speech: Categorization tests on animals and infants. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 355–386). New York: Cambridge University Press.

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

Kuhl, P. K. (1989). *Human adults and human infants exhibit a prototype effect for phoneme categories: Monkeys do not.* Manuscript submitted for publication.

Lehiste, I., & Peterson, G. E. (1959). Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America, 31,* 428–435.

Massaro, D. W., & Cohen, M. M. (1976). The contribution of fundamental frequency and voice onset time to the /zi/–/si/ distinction. *Journal of the Acoustical Society of America, 60,* 704–717.

Mervis, C., & 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. L. (1983). A possible auditory basis for the internal structure of phonetic categories. *Journal of the Acoustical Society of America, 73,* 2124–2133.

Oden, G. C., & Massaro, D. W. (1978). Integration of feature 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.

Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 422–435.

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

Repp, B. H. (1976). Identification of dichotic fusions. *Haskins Status Report on Speech Research, 45/46,* 123–129.

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.* London: Academic Press.

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

Sandell, J. H., Gross, C. G., & Bornstein, M. H. (1979). Color categories in macaques. *Journal of Comparative and Physiological Psychology, 93,* 626–635.

Sawusch, J. R., & Pisoni, D. B. (1974). On the identification of place and voicing features in synthetic stop consonants. *Journal of Phonetics, 2,* 181–194.

Sherman, T. (1985). Categorization skills in infants. *Child Development, 56,* 1561–1573.

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.

Strauss, M. S. (1979). Abstraction of prototypical information by adults and 10-month-old infants. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 618–632.

Sugarman, S. (1983). *Children's early thought: Developments in classification.* Cambridge, England: Cambridge University Press.

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

Received February 10, 1988
Revision received November 2, 1988
Accepted November 2, 1988 ■