Developmental Psychology
1992, Vol. 28, No. 3, 396–405

0012-1649/92/$3.00

Vowel Categorization by Very Young Infants

G. Cameron Marean, Lynne A. Werner, and Patricia K. Kuhl
Department of Speech and Hearing Sciences
Child Development and Mental Retardation Center
University of Washington

Six-month-old infants are known to categorize vowels despite variation in talker voice and pitch contour. Using the observer-based psychoacoustic procedure, this study asked whether 2- and 3-month-olds could categorize similarly. Infants were trained to respond whenever the vowel category alternated from /a/ to /i/ and to refrain from responding when the vowel category remained the same, despite variation in spectral cues associated with pitch and talker changes. Eighty percent of 2-, 3-, and 6-month-olds did not respond the first time a talker change occurred in the absence of a vowel change, suggesting that even the younger infants recognize these spectrally different sounds as perceptually equivalent. This study establishes the observer-based psychoacoustic procedure as a viable research tool in the field of infant speech perception.

A long-standing debate in speech perception is the degree to which the categories of speech sounds are learned or provided for in our genetic endowment. There is a great deal of evidence that even young infants perceive the sounds of human speech phonetically. Infants as young as 1 month of age demonstrate "categorical perception" of consonants (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971); they discriminate along phonetic dimensions in an adult-like manner both for phonetic units they have been exposed to and for those they have never heard (see Eimas, Miller, & Jusczyk, 1987; Kuhl, 1987, for a review). It has also been shown that infants can discriminate between vowels at an early age (Trehub, 1973). Furthermore, infants readily discriminate between vowel categories that are quite similar acoustically (Swoboda, Morse, & Leavitt, 1976). It is known, however, that infants eventually lose the ability to discriminate some of the phonetic contrasts not phonemic in their language (Werker & Lalonde, 1988; Werker & Tees, 1984). It appears that infants come into the world with sophisticated discrimination abilities, which are subsequently modified by experience.

A portion of these data was presented at the Seventh International Conference on Infant Studies held in Montreal, Quebec, Canada, April 19, 1990.

This article is based on a thesis by G. Cameron Marean submitted to the Graduate School of the University of Washington in partial fulfillment of the requirements for the master of science degree. This work was supported by National Institutes of Health Grant DC00396 to Lynne A. Werner.

We acknowledge Jill Y. Bargones, Jo Ann Chavira-Bash, Pat Feeney, Lisa Rickard-Mancl, and Heather Taylor for their help in data collection; Linda Ebenstein, Kerry Green, Chris Prall, and Karen Wolak for their help in stimulus acquisition and transfer; and Justin Tamblyn for software support. Finally, we thank four anonymous reviewers for their comments.

Correspondence concerning this article should be addressed to G. Cameron Marean, Department of Speech and Hearing Sciences, JG-15, University of Washington, Seattle, Washington 98195.

There is evidence that infants can categorize as well as discriminate speech sounds. This ability is critical for speech communication. Hillenbrand (1983, 1984) demonstrated that 6-month-old infants can group syllables on the basis of the initial consonant in the face of variation in vowel and talker. Kuhl (1979, 1983) showed that, by 6 months of age, infants can recognize the phonetic similarity among vowels despite spectral differences associated with the sex and age of talkers. More recent studies have shed some light on the internal structure of vowel categories in infants. Kuhl (1991) demonstrated that 6-month-old infants, like adults, show a "perceptual magnet effect" for vowels that are prototypical (best exemplars) in their language. The perceptual magnet effect is evidenced by a decrease in discriminability among vowels proximal to the prototype. Physically equidistant vowels around nonprototypical exemplars are more easily discriminated than the corresponding pairs from the prototype group. This effect appears, however, to be linked to one's language experience, because 6-month-old Swedish babies show the effect with a Swedish vowel but not with an American English vowel, and American infants demonstrate the effect with an American vowel but not with the Swedish vowel (Kuhl, 1990; Kuhl et al., 1992). In addition, monkeys do not show the perceptual magnet effect (Kuhl, 1991) even though they demonstrate categorical perception. These studies suggest that considerable "learning" is taking place between birth and 6 months of age. The initial state of phonetic categories and time course of developmental change are not known.

It has been difficult to assess age-related changes in the categorization of speech in infants younger than 6 months of age because of methodological limitations. Habituation paradigms have been successfully applied to very young infants to show whether or not infants discriminate between two speech sounds. However, the perception of similarity in the context of perceivable differences is just as important. Habituation paradigms are not well suited to demonstrating perceived similarity because the perception of similarity could be evidenced only by a negative result.

Experiments using habituation techniques suggest that very young infants may have some degree of phonetic organization of speech sounds (Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988). By 2 months of age, infants' responses to a change in syllables were proportional to the degree of phonetic difference between the familiar (preshift) and novel (postshift) syllables. Newborns detected all the changes but showed no strong tendencies to respond differentially on the basis of phonetic difference. It is unclear whether these changes reflect the emergence of phonetic organization through experience or age-related changes in habituation.

Kuhl and Miller (1982) looked for evidence of perceived similarity, or "perceptual constancy," for vowel categories in 1- to 4-month-old infants using habituation. In their study, infants detected changes in vowel identity despite a distracting variation in pitch or pitch contour. Their results also showed that when vowel identity was varied infants did not dishabituate to a change in pitch. Conversely, the infants dishabituated to pitch shifts only when the vowel identity was held constant. The results are consistent with the idea that the acoustic dimensions associated with vowel identity are highly salient for infants at these ages but do not directly assess perceived similarity.

Kuhl (1979) was able to demonstrate perceived similarity for spectrally dissimilar vowels in 6-month-olds using the modified head-turn procedure in a transfer-of-learning approach. Infants were trained to turn their head toward a mechanical toy when the background vowel category /a/ changed to /i/. During the training of the head-turn response, only vowels with male voice characteristics were used. Once trained, infants were presented trials in which the vowel changed from a male's voice to a female's voice but the vowel category did not change. Infants did not respond to the change in voice only but continued to respond when the vowel category shifted. Furthermore, Kuhl (1983) used the transfer-of-learning paradigm to show that infants could categorize the vowels /a/ and /ɔ/, which are more acoustically similar. Although the transfer-of-learning approach is sensitive to infant abilities, the head-turn response cannot be used with infants younger than 5.5 months. This makes it difficult to describe the development of vowel categorization. The sole attempt to apply the head-turn procedure to infants younger than 5.5 months obtained results consistent with the idea that infant speech categories are formed along adult phonetic dimensions (Fodor, Garrett, & Brill, 1975); however, the difficulty of eliciting the response from young infants and the variability in the results weaken this conclusion.

Studies with younger infants, then, would help to address key theoretical questions regarding developmental changes in speech perception, yet no method has been available to make direct comparisons between younger and older infants. A method has been developed that might address this need but has so far been used only in psychoacoustic tasks (Olsho, Koch, Halpin, & Carter, 1987). The method is known as the observer-based psychoacoustic procedure (OPP); it is similar to the head-turn procedure used by Kuhl (1979, 1983) but has been used successfully with infants as young as newborns (Trehub, Schneider, Thorpe, & Judge, 1991; Werner & Gillenwater, 1990). Moreover, there is reason to believe that OPP could be used for categorization tasks with young infants: Spetner and Olsho (1990) were able to obtain pulsation thresholds from 3-month-olds. Pulsation threshold estimation is a complex task requiring infants to sort tone patterns that vary in intensity or frequency into "continuous" or "intermittent" categories. Because the habituation studies (Bertoncini et al., 1988; Kuhl & Miller, 1982) make clear predictions about the performance of young infants in a vowel-categorization task, an experiment directly addressing the question is a perfect opportunity to establish OPP as appropriate for use in categorization studies. The following experiment attempts to replicate Kuhl's (1979) study with 6-month-olds and includes younger infants as well using OPP as the experimental method.

Method

Subjects

Twenty-one 2-month-olds, fourteen 3-month-olds, and eighteen 6-month-olds were tested to obtain 10 subjects from each of the three age groups for the final sample (see Table 1 for attrition information). Subjects were recruited from the infant subject pool at the University of Washington, and all subjects were tested within 2 weeks of the reported age. In addition, all subjects met the following criteria for inclusion as reported by their parents or as assessed at time of testing: (a) full-term birth, with no complications of delivery or perinatal course; (b) normal postnatal developmental course; (c) never diagnosed as having hearing loss; (d) free of colds; (e) normal middle ear function on the day of testing as assessed by impedance audiometry; (f) no more than two prior occurrences of ear infections; (g) no family history of congenital hearing loss; (h) no prior participation in psychoacoustic or speech-perception experiments.

Stimuli and Apparatus

The stimuli are a subset of those used by Kuhl (1979). Kuhl's original experiment audiotapes were obtained and redigitized at a 10-kHz sampling rate and a 5-kHz low-pass filter setting with 12-bit quantization. Stimuli were four synthetic vowel tokens of /a/ and four tokens of /i/: two male and two female, two rising and two falling. The formant frequencies and bandwidths were consistent with those of Peterson and Barney (1952) and are displayed in Table 2. The rise and fall in fundamental frequency was 20 Hz. The falling stimuli were synthesized in a piecewise linear manner, changing from the first value (112 Hz for males and 189 Hz for females) to the second (132 Hz for males and 223 Hz for females) in the first 100 ms, remaining there for 40 ms, and then falling to the third value in the remaining 360 ms. The rising stimuli were linear over the entire 500-ms duration. All tokens were 500 ms in duration and "ramped" (linearly) to avoid noise at onset.

The stimuli were played from an AT&T PC 6300 computer through a Data Translation D/A board (2801A). Stimuli were passed through a Wilsonics digital attenuator, high- and low-pass filtered using two

Table 1
Reasons for Exclusion of Subjects From the Final Sample

Age group	Did not reach training criterion	<16 test trials	Failed tympanometry
2-month-olds	5	4	2
3-month-olds	1	2	1
6-month-olds	4	2	2

Table 2
Center Frequencies and Bandwidths (in Hz) of the First Three Formants of the /a/ and /i/ Vowels for Male and Female Talkers

Vowel	Male	Female
/i/	F3 3010 (111.5)	3310 (130.7)
	F2 2290 (77.4)	2790 (98.9)
	F1 270 (52.6)	310 (79.5)
/a/	F3 2440 (83.2)	2810 (99.9)
	F2 1090 (48.4)	1220 (53.9)
	F1 730 (45.2)	850 (53.4)

Note. Bandwidths are given in parentheses.

Kemo (90 dB per octave) filters set at 50 Hz and 4900 Hz, respectively, and amplified using a QSC 1080 amplifier. The stimuli were then passed through a Coulbourn manual attenuator and presented to subjects by an Etymotic ER-1 insert earphone.[1]

The stimuli were originally synthesized with equivalent rms sound pressures and had undergone loudness matching in the Kuhl (1979) study. Additional loudness matching was conducted before data collection to ensure that any disparity created by the differing sound delivery systems was eliminated. Stimuli were presented at approximately 70 dB SPL (C scale) and the system calibrated weekly in a 2-cc coupler (which approximates an adult's ear canal volume with an insert earphone) using a Bruel and Kjaer sound-level meter (model # 2215).

Testing was conducted in a double-walled sound-attenuated booth (Industrial Acoustics Corporation). The infant sat on the parent's lap in the room, facing a window into the control room and a video camera. There was a table in front of the parent and infant. An assistant sat at the table to the infant's left. The visual reinforcer, a mechanical toy enclosed with lights in a smoked Plexiglas box, was positioned to the infant's right. The mechanical toy could not be seen until the lights inside the box were turned on under computer control.

Procedure

Infants were tested using the OPP as described by Olsho et al. (1987). An observer,[2] blind to stimulus type, used the infant's behavior to judge whether a vowel change or no-change trial was presented. In general, infants responded by making appropriate head turns to the reinforcer, decreasing their overall movement, tensing, or widening their eyes. Observer judgments were recorded on-line by the computer. The observer received feedback on a computer monitor after each trial. The infant was reinforced by the activation of the mechanical toy whenever the observer correctly identified a vowel-change trial (see later discussion). OPP incorporates features of other infant testing techniques, including conditioned head-turn procedures (Moore & Wilson, 1978) and the forced-choice preferential looking technique (Teller, 1979).

In this experiment, vowel-change trials consisted of five repetitions of an /a/ token interleaved with an /i/ token with 200 ms between tokens (total trial duration 7 s). No-change trials were identical except that the second member of the pair was also an /a/ token (i.e., the vowel identity did not change). Trials always began with a male /a/ (rising or falling pitch contour) token. The pitch contour for both tokens was chosen randomly throughout the experiment. An example of each trial type is shown in Figure 1.

An insert earphone was placed in the infant's right ear while the infant was seated on the parent's lap. Several ear tip sizes were available to fit the ear canal diameter of individual infants; we have found that infants younger than 12 months of age tolerate these earphones well. The parent and an assistant who sat in the test room wore headphones to prevent them from hearing the sounds presented to the infant. The parent listened to soft music, and the assistant monitored activity in the control room. The assistant manipulated toys to keep the infant's attention at midline before the initiation of a trial. The parent never knew when a trial was in progress; because the assistant monitored the control room where the observer was seated, the assistant knew when trials were begun. This ensured that the assistant did not interfere with trials by inadvertently changing toys during a trial.

The observer watched the infant through the window or on a video monitor. When the infant was alert, quiet, and attending at midline, the observer began a trial. Vowel-change and no-change trials were presented with equal probability, with the restriction that one trial type (vowel change or no change) could not occur more than four times in a row.[3] The mechanical toy was activated immediately when the observer correctly identified a vowel-change trial (a "hit"). Although the reinforcement contingencies varied slightly during the experiment (see later discussion), the infant never received reinforcement after a no-change trial regardless of the observer's response. The observer's response window was equal to the duration of the trial.

The sessions consisted of three phases: shaping, training, and categorization phase. The computer informed the observer of the successful completion of each phase. In the shaping phase, both types of trials contained only male tokens, and the infant was rewarded immediately each time the observer made a correct response to a vowel-change trial. In this case the reinforcer remained on for the remainder of the trial and continued for 4 s after the trial. However, in the absence of a response (a "miss"), the infant received 4 s of reinforcement immedi-

[1] The Etymotic ER-1 provides a flat response (in the free field) out to 10000 Hz. Insert earphones have several advantages over free-field sound presentation. Signals will vary in level depending on head position when presented in the free field, but head movements do not change the integrity of the stimulus reaching the eardrum with an insert earphone. By using the right ear and having the mechanical reinforcer located to the infant's right, insert earphones do not lose the orienting tendencies that head-turn methods take advantage of in infants.

[2] Observers in OPP are trained over a period of weeks or months depending on the individual. Observers are considered trained when they reach criterion in the majority of their sessions and obtain consistent data in testing. "Consistent" means that their data are stable and within the typical range of variation seen with other trained observers. In this experiment, observers trained quickly, and only those sessions conducted with trained observers are counted in the data analysis. There were a few cases in which infants were seen on their first visit by an untrained observer, were not trained, and were then seen again by a trained observer. The session with the untrained observer is not counted in the training trial data or success counts. Because the number of these instances is small, it does not compromise the analysis. When untrained observers completed the training phases, the session was not counted and the infant did not return for subsequent testing.

[3] There is a concern that observers could use this knowledge to improve performance. In the present study, change trials after 4 consecutive no-change trials never occurred on the first occurrence of those trial types in the "categorization phase." Overall, this happened five times in 240 trials. No-change trials occurred on 3 of 240 trials after 4 consecutive change trials and never on the most relevant trials: the first occurrence of those trial types. Thus, the constraint was invoked so rarely that it could not have influenced the first occurrence data (see results). The reader is referred to Spetner and Olsho (1990) for further discussion of these issues in infant testing.

No-change Trial

Talker	male	male	male	male	male	male	male	male	male	male
Vowel	/a/	/a/	/a/	/a/	/a/	/a/	/a/	/a/	/a/	/a/
Pitch Contour	rising	falling	rising	falling	rising	falling	rising	falling	rising	falling

Vowel Change Trial

Talker	male	male	male	male	male	male	male	male	male	male
Vowel	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/
Pitch Contour	rising	rising	rising	rising	rising	rising	rising	rising	rising	rising

Voice Change Trial

Talker	male	female	male	female	male	female	male	female	male	female
Vowel	/a/	/a/	/a/	/a/	/a/	/a/	/a/	/a/	/a/	/a/
Pitch Contour	falling	rising	falling	rising	falling	rising	falling	rising	falling	rising

Voice + Vowel Change Trial

Talker	male	female	male	female	male	female	male	female	male	female
Vowel	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/
Pitch Contour	rising	falling	rising	falling	rising	falling	rising	falling	rising	falling

Figure 1. One example of each of the four trial types. (Note that all trials begin with male /a/ tokens. Training trials are restricted to no-change and vowel-change trials [8 possible combinations]; in the categorization phase, all four trial types are presented [16 possible combinations]. "Rising" and "falling" refer to the pitch contour of the token. Tokens are 500 ms in duration; interstimulus interval is 200 ms, and the duration of the entire trial was 7 seconds [intertrial interval varied with infant state as determined by the observer].)

ately at the end of the trial. This phase continued until the observer responded correctly to four of the previous five trials (vowel change or no change).

In the training phase, all conditions remained the same except that the infant received no reinforcement at the end of a vowel-change trial if he or she failed to provide a noticeable response. This was to teach the infant that reinforcement was contingent not only on the stimulus type but also on the infant's response. The training phase continued until the observer was correct on four of the previous five vowel-change trials and four of the previous five no-change trials.

In the categorization phase, conditions remained essentially as they were except that the second member of the stimulus pair could be a female token—rising or falling /a/ or /i/. Thus, the male /a/ followed by the female /a/ was a voice-change trial, and the male /a/ followed by a female /i/ was a vowel plus voice-change trial. As in the preceding phases, only responses to those trials containing a vowel change were reinforced. Trials that the infant had received in training were presented with the same probability as the novel tokens in the categorization phase. This phase continued until at least 16 test trials had been completed. Trial selection was constrained to ensure one presentation of each possible token permutation during this interval of the phase. After 16 trials in the categorization phase, the session was stopped if the infant had reached criterion (4 of the last 5 vowel-change or vowel plus voice-change trials correct and 4 of the last 5 voice-change or no-change trials correct). The session was continued if criterion had not been reached (up to a maximum of 40 trials in the categorization phase). Sessions typically lasted less than 20 min.

Results

The primary question addressed in this study was whether or not young infants recognize the phonetic similarity among /a/ tokens that are spectrally different. These differences include average fundamental frequency and the location (in frequency) of salient formants. In addition, the task must be performed with random rise or fall in fundamental frequency direction.[4]

[4] Initial analysis and pilot data showed only minor effects of intonation contour. The effects are discussed in detail later. However, there is

There are several ways of addressing that question. One analysis assumes that the best way to determine how infants categorize the novel tokens is to see how they respond the first time they hear each trial type in the categorization phase. This was the primary approach taken in the Kuhl (1979) study. If performance on trials containing male voices only is the same as the trials with both male and female voices, then one can infer that infants are treating male and female tokens of the same phoneme as perceptually similar. We thus examined group performance on the first occurrence of each trial type in the categorization phase.

Another approach is to ask how well infants perform overall in the categorization phase. If infants perform as well throughout the categorization phase as they do in the last 10 trials of the training phase, this would be further evidence that the novel stimuli presented in the categorization phase are similar to the training stimuli. A signal-detection approach was used to obtain bias-free measures of sensitivity during the categorization phase.

In the third approach, the number of trials to criterion in training and categorization phases was compared. If the infants reach criterion performance (i.e., 80% correct) in the categorization phase faster than they did in training, this is an indication that infants do not have to relearn the task; infants should not have to relearn the task if they are categorizing by phonetic identity.

Of course, although our primary interest is the performance of 2- and 3-month-old infants, it is important to demonstrate that the results of Kuhl (1979) with 6-month-olds are replicated in the present experiment. Unless 6-month-olds can be shown to categorize, it would be impossible to interpret a failure to categorize at younger ages. This section therefore begins by examining the evidence for 6-month-old categorization.

A secondary question is "How well suited for speech-perception testing is OPP?" Attrition rates and other measures are examined to evaluate the success rate with OPP.

Vowel Categorization

6-month-olds: First occurrence of each trial type. In the Kuhl (1979) study, the strongest evidence that infants categorized novel tokens phonetically was group performance on the first occurrence of each trial type in the categorization phase, voice-change trials in particular. Kuhl (1979) found that seven of eight 6-month-olds correctly ignored the first occurrence of a voice-change trial and that an equally high proportion of infants responded correctly on the first occurrence of both vowel and vowel plus voice-change trials. Table 3 shows the proportion correct on the first occurrence of each trial type for 6-month-olds in Kuhl's (1979) study and in the present study. The proportion correct on vowel-change trials and vowel plus voice-change was .86 or better. Performance on no-change and voice-change trials on the first occurrence was good as well as is demonstrated by a proportion correct of .80 or better.

It is clear from these results that infants are not responding randomly to the novel voice-change and voice plus vowel-change trials. The binomial probability of obtaining 8 correct in 10 attempts given chance performance (response probability of 50%) is very low (p = .05). These results demonstrate that infants continued to respond to a change in category while ignoring the irrelevant dimension of talker. Good performance on the first occurrence of no-change or voice-change trials in the categorization phase was not the product of low overall responsiveness because the proportion correct for both vowel and vowel plus voice-change trials was high. The first occurrence data provide no evidence that infants respond any differently to voice-change trials in the categorization phase than to no-change trials in training.

These results replicate Kuhl's (1979) result that 6-month-olds demonstrate good categorization for spectrally dissimilar vowel tokens. Other details of 6-month-olds' performance are discussed relative to 2- and 3-month-olds later.

All age groups: First occurrence of each trial type. Responses to the first occurrence of each trial type in the categorization phase for 2-, 3-, and 6-month-olds are displayed in Table 4. The proportion correct was the same for all three groups on the first occurrence of a voice-change trial. Performance on vowel plus voice-change trials for the two younger age groups was only slightly worse than training criterion, and no-change trials remained near or above training criterion on the first occurrence in the categorization phase.

It is difficult to assess the statistical significance of differences in the number of infants responding correctly in different conditions and at different ages because of the small number of subjects. However, assume that the true probability of a correct response is .80, which seems reasonable given the training criterion and performance on no-change and vowel-change trials in the categorization phase. Then, assuming a Bernoulli process, fewer than 5 of 10 infants in any age group would have to respond correctly before we would say with 95% confidence that the proportion was significantly below .80. Because 7 or more infants responded correctly in each condition, we have no evi-

---

no evidence that infants in any age group only categorized certain intonation contours correctly or that infant responses to changes in intonation contour added significantly to the variability of response.

The degree to which the intonation contour of the token pairs on each trial affected performance was investigated separately for trials with and without voice changes. Trial types were identified as having tokens of the same intonation contour pattern (rise/rise or fall/fall) or different intonation contour pattern (rise/fall or fall/rise). Multivariate analyses of variance (MANOVAs; Age [3] × Trial Type [2]) indicated that on both kinds of trials neither age nor trial type effects were significant. For trials without a voice change, the Trial Type × Intonation interaction was significant (p = .029). Performance on no-change trials was significantly better when intonation contours differed across tokens as indicated by post hoc t tests. One explanation might be that, when contours are identical, there are "jumps" in fundamental frequency from one stimulus to the next, which elicit false alarms. This is not the case when token pairs have different contours. In this situation, there is a smooth flow in terms of pitch contour between the end of one stimulus and the beginning of the next (except in male–female token pairings).

The Age × Intonation interaction for trials without voice changes approached significance (p = .062). Two-month-olds tended to obtain fewer correct responses when intonation patterns were the same across token pairs. No other significant interactions were found.

Table 3
Proportion Correct of Four Trial Types on the First Occurrence in Testing for Kuhl's Original (1979) Study and for 6-Month-Olds in the Present Experiment

Study	Voice change	Voice + vowel change	No change	Vowel change
Kuhl (1979)	0.86 (0.131)	0.86 (0.131)	0.93 (0.096)	0.86 (0.131)
Present study	0.80 (0.131)	1.00	0.80 (0.131)	1.00

Note. Standard errors of the binomial are in parentheses.

dence of a statistical difference in performance in any age group in any condition. Furthermore, the probability of 14 of 20 two- and three-month-old infants responding correctly by chance in the vowel plus voice-change condition (the only questionable case) is still low ($p = .05$).

These results show, then, that 2- and 3-month-olds categorize these vowel tokens as well as 6-month-olds. The findings of Kuhl (1979) are replicated and extended to include younger ages.

Overall performance in categorization: Bias-free measures of sensitivity. Proportion correct is limited as a measure of performance because it reflects both sensitivity to the vowel change and response bias. Response bias, or the subject's criterion for responding "yes," may change during the course of a session while the underlying sensitivity remains the same. A shift toward a more lax criterion, for example, might increase responses to no-change trials and lower the overall proportion correct.

A nonparametric bias-free measure known as p(A) can be estimated from the hit rate (proportion correct to vowel-change or voice plus vowel-change trials) and the corresponding false-alarm rate (proportion of incorrect responses or saying "yes" to no-change or voice-change trials) for each subject. p(A) is the proportion of area under the "receiver operating characteristic" curve (Green & Swets, 1966) and is equivalent to percentage correct in a two-interval forced-choice task. Tabled values can be found in McNicol (1972). The transform, $2*\arcsin\sqrt{p(A)}$, is applied to obtain a nonparametric approximation of d', a more commonly used bias-free measure of sensitivity. By way of orientation, d' is essentially a z score. Thus, a d' of 2.50 would be equivalent to 90% correct in a single-interval, yes–no experiment if the observer is unbiased, whereas a d' of 0.00 would indicate chance performance. A d' of 1.00 is often taken as "threshold" because it is equivalent to 68% correct under these conditions.

In this study, d' was typically above 2.0 over the first 16 categorization trials for all ages. A MANOVA (Age [3] × Voice [2]) was conducted on d' based on the first 16 categorization trials. No significant main effects or interactions were found. Thus, there is no evidence that any one age group was more sensitive overall than another.

Figure 2 shows group d's,[5] based on group hit and false-alarm rates, for trials with and without voice changes on the first four occurrences in the categorization phase. For 2- and 6-month-olds, it appears that after the first occurrence introducing female tokens only slightly affects sensitivity. Two-month-olds become slightly less sensitive to trials without voice changes on the second occurrence but quickly recover. Their performance on trials with voice changes remains uniformly good throughout the session; they respond only on vowel plus voice-change trials. Six-month-olds show a gradual (and slight) degradation in performance from start to finish for trials with and without voice changes. Three-month-olds appear to behave differently, with performance uniformly degraded on the second occurrence regardless of talker. This condition is momentary because they recover to first-trial performance thereafter.

Overall, these data provide no evidence of differences between 6-, 3-, and 2-month-olds in ability to categorize these stimuli. It might be an interesting question to replicate and investigate why 3-month-olds, unlike other infants, seem less sensitive to the vowel change on the second occurrence of each trial type in the categorization phase.

Trials to criterion: Training. The number of trials to criterion in training is displayed in Table 5 and is collapsed over shaping and training phases. A one-way analysis of variance (ANOVA; age [3]) showed no significant differences among groups in the number of training trials to criterion. This shows that infants quickly learned to categorize the training stimuli and to ignore the irrelevant dimension of pitch contour.

Trials to criterion: Categorization phase. The number of trials to criterion in the categorization phase is displayed in Table 5 as is the number of infants reaching criterion. For the purposes of this analysis, all categorization trials were considered because more than 16 trials were required in some cases. A two-way ANOVA (Age [3] × Phase [2]) showed no significant effect of age on the number of trials to criterion in the categorization phase. One conclusion from this result is that there are no significant age-related changes between 2 and 6 months in the ability

[5] Because of the limited number of trials that are collected in such experiments, measures of sensitivity on individual subjects at different trial type occurrences could not be computed. It has been argued that a detection theory analysis with group data can be used if certain conditions are met (Macmillan & Kaplan, 1985). Among these conditions are that d's be of moderate magnitude and that subjects within a group will not differ greatly in response bias. Both of these conditions were met in the present experiment. Whenever the number of trials is small or the population variance is unknown, the main worry is in violating underlying assumptions of the analysis. Nonparametric indices of d' were used in the current study. Transforming the proportion correct to equate variance for proportions of differing magnitude was used here to avoid violating assumptions of between-groups homogeneity.

Table 4
Proportion Correct of Four Trial Types on the First Occurrence in Testing for Three Age Groups in the Present Experiment

Age group	Voice change	Voice + vowel change	No change	Vowel change
2-month-olds	0.80 (0.131)	0.70 (0.151)	0.70 (0.151)	0.80 (0.131)
3-month-olds	0.80 (0.131)	0.70 (0.151)	0.90 (0.10)	0.90 (0.10)
6-month-olds	0.80 (0.131)	1.00	0.80 (0.131)	1.00

Note. Standard errors of the binomial are in parentheses.

to maintain performance on this task. Although trials to criterion in the categorization phase was not reported in the Kuhl (1979) study, these results are encouraging when compared with Clarkson and Clifton's (1985) study on perceptual constancy for pitch with a "missing fundamental" in 7-month-olds. Even with a more lax criterion (five of six consecutive responses), their infants typically required 23 trials to reach criterion in their corresponding perceptual constancy phase. It is also true, however, that in their perceptual constancy phase, only novel tokens were presented, unlike our infants who received both training and novel stimuli in the categorization phase. The average for all infants in the present study to reach criterion in the categorization phase was about 15.

Clarkson and Clifton (1985) compared the number of trials to criterion in each phase as a measure of the strength of categorization. They reasoned that if infants reach criterion in the perceptual constancy phase in fewer trials than in training, it would be strong evidence that infants were not having to relearn how to categorize the novel stimuli. Clarkson and Clifton (1985) did not, however, find a statistically significant effect of experimental phase. In the present study, there is a consistent trend for infants to reach criterion in the categorization phase in fewer trials than in training, but the effect of phase was not significant. Variability among infants may account for the lack of a significant effect. The fact that infants do not reach criterion in fewer trials in the categorization phase than they do in training may also be evidence that the novelty of the female stimuli is salient and does affect performance beyond the first occurrence.

Success Rate With OPP

The results described in the previous section are evidence that OPP is a sensitive tool for speech-perception research with infants. The Kuhl (1979) study with 6-month-olds was replicated by the first occurrence data. Three-month-olds proved no more difficult to train than older infants (see Tables 1 and 5). They were, in fact, easier in terms of success rate, and 3-month-olds performed like older infants on the first occurrence of each trial type. In addition, this study represents the first attempt to use OPP with 2-month-old infants. The data for 2-month-olds show performance comparable to 6-month-olds in almost every respect. It is clear in the present experiment that OPP worked well.

To obtain the final sample, more infants had to be tested in the present experiment than in Kuhl's (1979) study. However, infants were able to complete all phases of this experiment in a single sitting. In Kuhl (1979), infants typically required four visits to the laboratory to proceed through all of the experimental stages. Thus, although more infants were tested with OPP, the total amount of time spent in data collection was not greater.

Discussion

The major finding of this study is that infants as young as 2 months of age can categorize spectrally dissimilar vowels phonetically. Two results strongly support this conclusion. Two-month-olds respond to a novel exemplar of a vowel based on its phonetic identity on the first presentation of the novel exemplar. Second, infants maintain a high level of performance in vowel categorization in the context of increased nonphonetic spectral variation. That the infants perceive the increased variation is suggested by the finding that infants do not reach criterion performance in the categorization phase faster than they do in the training phase.

These data suggest that there is little developmental change in vowel categorization between 2 and 6 months of age. Strong conclusions regarding the similarity of vowel categorization in 2- and 6-month-olds must be tempered, however, by the relative simplicity of the task. By increasing the complexity of the task by either including a third talker (Kuhl, 1979) or using more similar vowels (Kuhl, 1983), the limit of 2-month-olds' ability to perform such a "sorting" task might be reached. For example, age-related changes in the ability to categorize novel visual patterns has been shown to occur between 2 and 7 months for complex but not for simple patterns (Younger & Gottlieb, 1988). Increasing the complexity of the task would help to address these issues.

It must be recognized that infants may not categorize vowels on the same physical bases that adults do. Vowels may be categorized on the basis of overall spectral shape, absolute values of formants, or ratio relationships among formants (Nearey, 1989; Strange, 1989). The gross spectral differences between the vowel categories used in this and other studies may allow infants to sort vowels solely on the basis of psychophysical distance rather than phonetic identity. At the same time, it is undoubtedly important for the development of speech perception

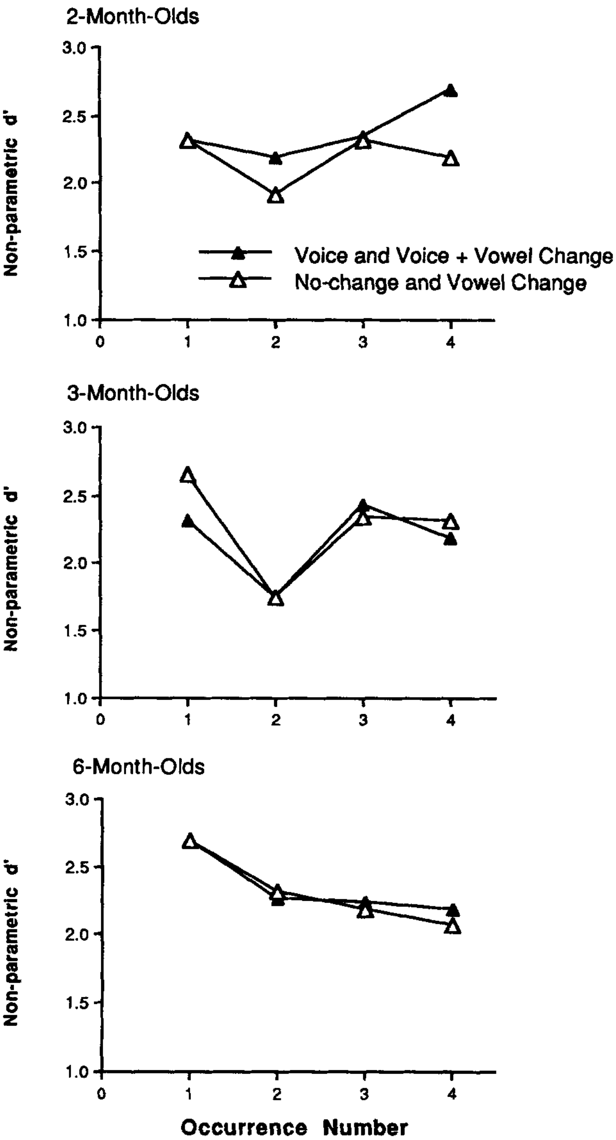

Figure 2. Nonparametric *d'* of group data as a function of occurrence number for no-change/vowel-change and voice-change/voice plus vowel-change trials in the categorization phase.

that infants are able to categorize speech sounds on any basis. Further work in both adults and infants is necessary to resolve this question.

Another important finding is the usefulness of OPP for testing infants on speech-categorization tasks. Results obtained with head-turn procedures were replicated using OPP. In addition, success rates at all ages tested were acceptable. With OPP, the same experimental method can be applied to infants throughout the first year of life (Olsho et al., 1987). Recent modifications to OPP have made the testing of newborns possible (Werner & Gillenwater, 1990). Thus, OPP may allow us to finally answer questions regarding the initial state of infant speech perception.

Signal-detection theory (Green & Swets, 1966) proved to be a useful tool in interpreting the current results. Signal-detection theory has a long history in speech-perception research (e.g., Wood, 1976), but it is not a common form of analysis in the development literature. Because developmental data must often be collected in a single-interval paradigm, the ability to measure bias over time is critical. In the present experiment, a signal-detection approach provided several important insights, suggesting that future research should be designed to optimize a signal-detection analysis. Both OPP and head-turn procedures allow for the application of this approach.

Given that OPP is an appropriate tool for the testing of developmental speech perception, there are several questions that can now be directly tested using younger infants. First, developmental changes in speech perception, as well as the effects of specific language experience on speech perception, can now be assessed using this new technique. For example, in Kuhl's (1990; Kuhl et al., 1992) cross-language study, 6-month-old Swedish infants do not demonstrate a perceptual magnet effect for the American English vowel, and American infants do not demonstrate the effect for the Swedish vowel. Both groups show the effect only for their native-language vowel prototype. The only way to follow the developmental time course of the effect is to test younger ages.

Second, some models of speech perception claim that speech categories are fundamentally based on experience. For example, Durlach and Braida's (1969) model of intensity perception has been applied to the perception of speech (Macmillan, Goldberg, & Braida, 1988). The model posits that in a discrimination or identification task listeners compare the stimulus with a stored representation of a sound. Performance is limited by "context noise." Along a continuum, context noise is low when there is a good representation of the stimulus and high when the stored representation is poor. Listeners adopt reference points in memory, or "perceptual anchors," where context noise is low to make decisions. For speech sounds, listeners come into an experiment with considerable experience and robust perceptual anchors, thus accounting for the noncontinuous nature of speech-sound discrimination. Very young infants, however, with much less experience with speech sounds, should be more susceptible than adults to paradigmatic variations that affect context noise if the model is correct. Thus, developmental data can provide a critical test of the model. Future research could be directed toward investigating the emergence of perceptual anchors after birth with OPP.

Finally, there are several important issues in complex sound processing that could also be investigated with OPP. For example, although relatively mature, complex pitch perception has been demonstrated for 7-month-old infants using head-turn procedures (Clarkson & Clifton, 1985), in very young infants dishabituation to changes in static or dynamic fundamental frequency patterns is not always evident in the context of other spectral variations (Bundy, Colombo, & Singer, 1982; Karzon & Nicholas, 1989; Kuhl & Miller, 1982). These findings may reflect an immaturity in complex sound processing or the limitations of habituation methods. OPP might provide a more sensitive measure on such a task because, with reinforcement, the

Table 5
Number of Training Trials to Criterion and Number of Test Trials to Criterion With Standard Deviations

Age group	Average training trials to criterion	*SD*	Average test trials to criterion	*SD*
2-month-olds	19	6.67	13 (7)	3.25
3-month-olds	24	8.64	17 (9)	5.80
6-month-olds	23	9.38	14 (8)	6.53

Note. Number of infants reaching criterion in testing is given in parentheses.

infant's attention could be directed to the appropriate stimulus dimension; infants might be trained to respond to changes in intonation contour or pitch category regardless of other acoustic variation.

In summary, this study shows that infants as young as 2 months of age can categorize spectrally dissimilar vowels. OPP appears to be a useful tool in infant speech-perception research. Because OPP has been used successfully with infants as young as newborns and as old as 12 months of age, OPP may provide a way to investigate in greater detail the emergence of auditory perceptual processes throughout the first year of life.

References

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' representations of speech sounds. *Journal of Experimental Psychology: General, 117,* 21–33.

Bundy, R. S., Colombo, J., & Singer, J. (1982). Pitch perception in young infants. *Developmental Psychology, 18,* 10–14.

Clarkson, M. G., & Clifton, R. K. (1985). Infant pitch perception: Evidence for responding to pitch categories and the missing fundamental. *Journal of the Acoustical Society of America, 77,* 1521–1528.

Durlach, N. I., & Braida, L. D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America, 46,* 372–383.

Eimas, P. D., Miller, J. L., & Jusczyk, P. (1987). On infant speech perception and the acquisition of language. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 161–195). New York: Cambridge University Press.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171,* 303–306.

Fodor, J. A., Garrett, M. F., & Brill, S. L. (1975). Pi ka pu: The perception of speech sounds by pre-linguistic infants. *Perception and Psychophysics, 18,* 74–78.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

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.

Karzon, R. G., & Nicholas, J. G. (1989). Syllabic pitch perception in 2- to 3-month-old infants. *Perception & Psychophysics, 45,* 10–14.

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

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In P. Salapatek & L. Cohen (Eds.), *Handbook of infant perception, Vol. 2. From perception to cognition* (pp. 275–381). New York: Academic Press.

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

Kuhl, P. K. (1991). Human adults and infants exhibit a "perceptual magnet effect" for speech sounds, monkeys do not. *Perception and Psychophysics, 50,* 93–107.

Kuhl, P. K., & Miller, J. L. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception and Psychophysics, 31,* 279–292.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255,* 606–608.

Macmillan, N. A., Goldberg, R. F., & Braida, L. D. (1988). Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua. *Journal of the Acoustical Society of America, 84,* 1262–1280.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin, 98,* 185–199.

McNicol, D. (1972). *A primer of signal detection theory.* Sydney, New South Wales, Australia: Australasian.

Moore, J. M., & Wilson, W. R. (1978). Visual reinforcement audiometry (VRA) with infants. In S. E. Gerber & G. T. Mencher (Eds.), *Early diagnosis of hearing loss.* (pp. 177–214). New York: Grune & Stratton.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America, 85,* 2088–2113.

Olsho, L. W., Koch, E. G., Halpin, C. F., & Carter, E. A. (1987). An observer-based psychoacoustic procedure for use with young infants. *Developmental Psychology, 23,* 627–640.

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

Spetner, N. B., & Olsho, L. W. (1990). Auditory frequency resolution in infancy. *Child Development, 61,* 632–652.

Strange, W. (1989). Dynamic specification of coarticulated vowels spo-

ken in sentence context. *Journal of the Acoustical Society of America, 85,* 2135–2153.
Swoboda, P. J., Morse, P. A., & Leavitt, L. A. (1976). Continuous vowel discrimination in normal and high risk infants. *Child Development, 47,* 459–465.
Teller, D. Y. (1979). The forced-choice preferential looking procedure: A psychophysical technique for use with human infants. *Infant Behavior and Development, 2,* 135–153.
Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology, 9,* 91–96.
Trehub, S. E., Schneider, B. A., Thorpe, L. A., & Judge, P. (1991). Observational measures of auditory sensitivity in early infancy. *Developmental Psychology, 27,* 40–49.
Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24,* 672–683.
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.
Werner, L. A., & Gillenwater, J. M. (1990). Pure-tone sensitivity of 2- to 5-week-old infants. *Infant Behavior and Development, 13,* 355–375.
Wood, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America, 60,* 1381–1389.
Younger, B., & Gottlieb, S. (1988). Development of categorization skills: Changes in the nature or structure of infant form categories? *Developmental Psychology, 24,* 611–619.

Received November 5, 1990
Revision received September 7, 1991
Accepted November 11, 1991 ■

Neuropsychology to Be an APA Journal

In January 1993, *Neuropsychology*, which has been published by the Educational Publishing Foundation (a subsidiary publishing program of the American Psychological Association), will be published by the American Psychological Association. The Publications and Communications Board of the APA has appointed Nelson Butters as editor of *Neuropsychology*. As of January 1, 1992, manuscripts should be submitted to

Nelson Butters
Chief, Psychology Service (116B)
Department of Veterans Affairs Medical Center
3350 La Jolla Village Drive
La Jolla, CA 92161

Manuscripts considered by the incoming editor will be published beginning in the January 1993 issue. Submitted manuscripts should fall within the following new editorial policy statement:

> The mission of *Neuropsychology* is to foster (a) basic research, (b) the integration of basic and applied research, and (c) improved practice in the field of neuropsychology, broadly conceived. The primary function of *Neuropsychology* is to publish original, empirical papers in the field. Occasionally, scholarly reviews and theoretical papers will also be published—all with the goal of promoting empirical research on the relation between brain and human cognitive, emotional, and behavioral function. Sought are submissions of human experimental, cognitive, and behavioral research with implications for neuropsychological theory and practice. Papers that increase our understanding of neuropsychological functions in both normal and disordered states and across the lifespan are encouraged. Applied, clinical research that will stimulate systematic experimental, cognitive, and behavioral investigations as well as improve the effectiveness, range, and depth of application is germane. *Neuropsychology* seeks to be the vehicle for the best research and ideas in the field.