INFANT BEHAVIOR AND DEVELOPMENT 6, 313–330 (1983)

Developmental Changes in Male/Female Voice Classification by Infants*

CYNTHIA L. MILLER
University of Western Ontario

The ability of infants to classify human voices according to speaker gender (i.e., male vs. female) is important both in its relationship to emerging cognitive abilities during early infancy, as well as in its possible relationship to the perception of speech and language. The present study was designed to investigate the development of voice gender categorization in 2- and 6-month-old infants. Following habituation to several tokens of one voice category (male or female) in a standard habituation-dishabituation procedure, infants were presented with one of the following stimulus changes for discrimination: (a) between-category (e.g., female to male); (b) within-category (e.g., female to new female); or (c) control (no voice change). The results indicated that 2-month olds discriminated (i.e., dishabituated to) *all* stimulus changes, while 6-month olds demonstrated discrimination *only* in the between-category condition. This evidence of male/female voice categorization in 6-month olds thus adds another item to our growing inventory of stimulus categories to which these infants are responsive. In addition, the difference in the discrimination of voice categories by 2- and 6-month olds may also suggest fundamental differences in the way in which these two groups of infants process speech and language stimuli.

voice perception—infancy   infant categorization   categorization development   voice gender discrimination—infancy   infant auditory perception/discrimination

INTRODUCTION

The investigation of categories and category formation is a relatively recent development in the study of infant behavior, most likely due in large part to the increasing methodological ingenuity and sophistication of recent years. Thus far, most of this work has been conducted in the visual mode, and has demonstrated infant categorization for a wide variety of visual/object categories (e.g., Caron, Caron, & Carlson, 1979; Cohen & Strauss, 1979; Cohen & Younger, 1981; Cornell, 1974; Ruff, 1977; Younger & Cohen, 1981). Relative to these studies of visual categori-

* This research was conducted while the author was a BEH Postdoctoral Trainee in the Department of Human Development, University of Kansas, and was supported by grant #300–77–0308 awarded by the Bureau for the Education of the Handicapped to Frances D. Horowitz and the Early Childhood Institute (Judith LeBlanc and Edward Meyen, Principal Investigators), University of Kansas. The author gratefully acknowledges Frances Degen Horowitz for her input and her generosity in providing support facilities for the research, and the entire staff of the Kansas University Infant Research Laboratory for their support and assistance in testing subjects. Correspondence and requests for reprints should be addressed to Cynthia L. Miller, Department of Psychology, University of Western Ontario, London, Canada, N6A 5C2.

zation, investigations of auditory category formation are considerably more limited, both in number and in content. In fact, for the most part, studies of categorization of auditory stimuli are restricted to studies of infant speech perception, and even more specifically to the investigation of phonetic categorization.

Historically, investigators of infant speech perception have always been interested in the question of infant categorization, the beginnings of which can be seen in the very early studies of infant "categorical perception." Categorical perception, as viewed from the model of adult speech perception, has been defined as the ability to discriminate between-category contrasts better than within-category contrasts (e.g., Liberman, 1970). Thus, in a prototypical infant study of categorical perception (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971), infants are presented with a between-category (e.g., [ba] vs [pa]) and a within-category contrast (e.g., within–[bal]), both of which vary equally along one stimulus dimension, for example, voice onset time (VOT). To the extent that discrimination of the between-category contrast is superior to that for the within-category contrast, the two are said to be discriminated categorically. With this type of approach, many investigators have observed that, at this level of phonetic categorization, even very young infants discriminate many stop consonant contrasts in a manner resembling adult categorical perception (e.g., Eimas, 1974; Eimas et al., 1971; J. Miller, 1974; Miller & Morse, 1976).

Still within the domain of infant speech perception, several more recent studies have gone beyond this traditional conception of "categorical perception" to the investigation of "perceptual constancy" in phonetic categorication (cf. Kuhl, 1980, for a review). In these studies, perceptual constancy is defined by the extent to which infants can (a) treat *several* (i.e., more than two) different members of the same phonetic category as similar and as distinct from members of another phonetic category; and (b) tolerate within-category variability along several acoustic dimensions simultaneously. For example, one study of this sort presented many different tokens of the vowel /i/ and the vowel /a/ to infants, each varying along such dimensions as speaker, pitch, and intonation contour (Kuhl, 1979). The task of the infant was to ignore, or to generalize across, these irrelevant dimensions and to respond solely to the relevant dimension, i.e., the phonetic category. Successful performance on this type of task indicates the infant's ability to recognize phonetic similarity among acoustically dissimilar speech stimuli. Thus, the observation of perceptual constancy in phonetic categorization appears to be a considerably more general phenomenon than the demonstration of categorical perception. Although this type of ability has not been explored developmentally, it does seem to be available to infants by at least 7 months of age (Kuhl, 1980).

Thus far, studies of infant categorization for auditory stimuli have been restricted to these questions of phonetic categorization, i.e., categorical discrimination of stop consonants and phonetic perceptual constancy. Although these are important questions, phonetic categories are only one of many auditory/speech categories to which young infants are exposed and to which mature listeners are responsive. One such categorical distinction, for example, is speaker gender, i.e., male vs. female voices. Evidence collected from adult speakers and listeners suggests speaker gender to be an important and systematic source of acoustic variation in both the production and the perception of speech and language. In production studies, Fant (1973) and Peterson and Barney (1952) have mapped in extensive detail the specific spectral characteristics of many phonemes, and have found clear differences in such features as fundamental frequency (F_0) and formant structure as a function of the class of speaker (e.g., man, woman, child). These studies have shown, for example, not only that male and female speakers produce different acoustic output for the same phoneme, but they often may produce *similar* spectral patterns for *different* phonemes (e.g., the formant structure of a male /a/ resembles that of female /ɔ/).

These output differences in production seem to have an important influence on listeners' perceptual experiences as well. While Fant (1973) and Peterson and Barney (1952) have described structural differences between male and female speakers (i.e., in F_0 and formant structure), Coleman (1973) has demonstrated their functional importance in listeners' identification of voices. By independently manipulating these two features (e.g., combining a female fundamental frequency with male formant values), Coleman found both to contribute to the perception of speaker gender. Although fundamental frequency accounted for the largest part of the variance in subjects' ability to identify male and female voices, the resonance characteristics of the formants contributed in large part to this identification as well. That the formant structure of particular phonemes varies as a function of speaker gender appears to have an effect upon listeners' phonetic perception as well. For example, Rand (1971) has shown that speaker gender has systematic effects upon the specific identification and category boundaries that adult listeners assign to a [bae-dae-gae] continuum. Thus, it would appear that speaker differences present to the speech perceiver more than simply random noise, and in fact, accurate speech perception for the mature listener may, at some level, involve more fundamental knowledge of the characteristics of male and female voices.

These studies have two important implications for the study of voice perception and voice classification in young infants. First, if we assume that infants are presented daily with many new and different voices, then categorizing these voices with respect to speaker gender is one way in which infants can impose some organization and regularity into this array of acoustic input, thus allowing them to process new information (i.e., a new voice) more efficiently. In addition, moreover, since speaker gender does appear to produce systematic effects on the perception and production of speech and language for adults, voice classification might also have important implications for the infant's perception of language. We have suggested elsewhere, for example, that the ability to ignore speaker differences in perceptual constancy studies may be related to this more fundamental knowledge of male and female voice categories (Miller, Younger, & Morse, 1982). That is, the ability to categorize different voices with respect to gender may be one process by which infants can "normalize" the extensive (and probably very salient) acoustic variability produced by different speakers. Thus, examining the development of voice categorization skills in infants will allow us to explore not only how infants learn to organize voices and voice information in general, but may also help us to better understand the relationship between voice perception and other features of linguistic perception, e.g., phonetic categorization.

In previous work, we have demonstrated that by at least 7 months of age, infants do appear able to categorize male and female voice tokens (Miller et al., 1982). In these studies, 6–7-month-old infants were tested in a modified version of the operant head-turning paradigm developed by Wilson, Moore, and Thompson (1976) and Eilers, Wilson, and Moore (1977), using a variety of different male and female voices as discriminative stimuli. The procedure was similar to that employed by Kuhl and her colleagues (Kuhl, 1980) in their studies of phonetic categorization. The results of these studies indicated that infants could learn to respond discriminatively to these groups of male and female voices only when they were organized into the appropriate gender categories. That is, when male and female voices were ordered randomly or ordered according to a single dimension (i.e., fundamental frequency), irrespective of gender category, infants did not learn the discrimination (Miller et al., 1982).

Having demonstrated that infants are sensitive to male vs. female voices by the age of 6–7 months, the purpose of the present study was to begin to explore the development of this classification skill. Since the operant head-turning paradigm used in previous studies of phonetic and voice classification is not well-suited for use with very young infants (and thus does not lend itself well to studies of development), the present study instead employed an extension of a conventional infant discrimination paradigm. Similar versions of this general technique have been employed quite successfully in several studies of infant visual categorization (e.g., Cohen & Strauss, 1979; Cornell, 1974; Ruff, 1977). The procedure employed was the infant-control visual-habituation paradigm, described by Horowitz (1975), in which the infant is presented with auditory stimuli (e.g., female voices) contingent upon the fixation of a visual stimulus (in this case, a checkerboard slide). Following several trials of the familiarization stimuli (i.e., the female voices), the infant's looking behavior typically habituates, whereupon the auditory, but not the visual, stimulus is changed (e.g., to male voices). Recovery (dishabituation) of the infant's visual fixation following this change in auditory stimulation indicates discrimination of the auditory stimulus contrast.

The major assumption underlying the use of this habituation-dishabituation technique in studies of infant discrimination is that the observation of habituation to a particular stimulus (or set of stimuli) suggests that the infant recognizes that stimulus (or stimuli) as familiar, i.e., has formed a memory for it. When the stimulus is changed, dishabituation implies that the infant has noticed that change, whereas a failure to dishabituate suggests that at some level the infant views the new stimulus as the same as, or similar to, the old (now familiar) stimulus. Thus, by comparing the discriminability of various stimulus contrasts one can make inferences regarding the types of stimulus features to which infants are most responsive. For example, in the categorical perception studies cited earlier, it is the observation that dishabituation occurs to between-category but not to within-category changes that allows the inference of "categorical perception."

A similar approach was employed in the present study, in which infants were presented with three types of stimulus contrasts for discrimination: (a) a *between-category* discrimination, in which the dishabituation stimuli were members of a *different* gender category from the familiarization stimuli (e.g., a change from female to male voices); (b) a *within-category* discrimination, in which the dishabituation stimuli were from the *same* category as the familiarization stimuli (e.g., a change from one group of female voices to another group of female voices); and (c) a *control* condition, in which the dishabituation stimuli were the same voices as those presented during familiarization. The stimuli were the same as those employed by Miller (1979) and included a total of 12 female and 12 male voices. Thus, in the present study, *categorical* discrimination of voices will be inferred by the extent to which between-category discrimination is superior to within-category discrimination. In addition, however, the reader should note that the use of several different tokens of each category quite closely resembles the perceptual constancy paradigm (cf. Kuhl, 1980), and thus, will allow us to assess the generality of this categorical discrimination.

As indicated, the major purpose of the study was to begin to explore the development of voice classification over the first six months. Accordingly, both 2-month-old and 6-month-old infants were tested with these three categorization tasks. If the previous results of Miller et al. (1982) are generalizable across tasks, then we can expect 6-month olds to demonstrate voice classification in this situation as well, i.e., to exhibit superior between-category to within-category discrimination. Moreover, assuming that voice classification is a skill that develops as a function of increasing experience and/or cognitive abilities, we should not expect 2-month olds to be capable yet of this type of organizational ability. Thus in the case of these younger infants, it was expected that the between- and within-category contrasts would be equally discriminable.

METHOD

Subjects

Twenty-four 2- to 3-month olds (range = 8.5–13.5 weeks, $\overline{X}$ = 10 weeks) and 24 6- to 7-month olds (range = 23–29.5 weeks, $\overline{X}$ = 27.25 weeks) were used as subjects in the present experiment. There were 12 male and 12 female 2-month olds, and 19 male and 5 female 6-month olds. A total of 94 infants was tested to achieve the final sample of 48. The remaining 46 infants (28 2-month olds, 18 6-month olds) were eliminated due to inappropriate state (18 2-month olds, 14 6-month olds), low interobserver reliability (9 2-month olds, 3 6-month olds), and equipment failure/experimenter error (1 2-month old, 1 6-month old). Criteria for acceptance into the study are described below.

Stimuli

The stimuli were the same as those employed by Miller (1979) in her studies of voice categorization in 7-month olds. The initial sample of stimuli was collected by recording the voices of mothers and fathers saying "hi" to their 7-month old infants (Uher 4400 stereo tape deck, microphone M517). Several tokens of each voice were collected and all intelligible tokens were digitized through a 5KHz low-pass filter, matched for intensity, and stored on a Harris computer using the VOCAL program (Gillman & Wilson, 1979) of the Waisman Computing Facility of the University of Wisconsin. The final sample of stimuli consisted of 12 male and 12 female voices saying "hi." All 24 of these stimuli were presented to adult subjects for identification, and all were reliably classified into male and female categories by these adult listeners (Miller, 1979). The 12 tokens of each category were divided into 2 sets of 6 voices each (Sets F1 and F2, Sets MI and M2).

The stimuli in sets M1 and F1 were the same as those used by Miller et al. (1982) in the original study of voice classification. The stimuli constituting sets M2 and F2 were used by Miller (1979) and were selected primarily according to durational criteria such that the mean duration of all category tokens was similar for the four sets of voices. Statistical analyses (independent t-tests) showed no significant differences in mean duration among the four sets of voices (across the 24 voices employed, stimulus duration ranged from 380 msec to 660 msec). No attempt was made either to minimize or to maximize the qualitative (e.g., prosodic) differences among stimulus sets; however, any stimuli appearing to the investigator as "extreme" in those prosodic features were eliminated from consideration. It should also be noted, however, that all stimuli were the voices of parents talking to their infants; consequently, the exaggerated intonational contours characteristic of "motherese" (Garnica, 1977; Newport, 1977; Stern, Spieker, Barnett, & MacKain, in press) were evident in both the male and the female voice tokens. Figure 1 illustrates representative samples of these contours.

Acoustic analyses for each stimulus were performed using the VOCAL program, which estimates (using a linear prediction model), for successive 25.6 msec intervals, the fundamental frequency (F_0), and the frequency, intensity, and bandwidth of each of the first four formants. T-tests for each stimulus contrast (i.e., M1 vs. F1, M1 vs. M2, and F1 vs. F2) were performed on each of these measures. Significant differences were observed in F_0 as well as in the mean frequency of the first and second formant for the M1/F1 comparison, whereas no statistically significant differences were found for either within-cateogry contrasts (i.e., M1 vs. M2 and F1 vs. F2). Finally, as indicated above, the mean durations of category tokens were equated across the four sets of voices. Table 1 summarizes both the peak and the mean F_0 for each of the four sets of voices.

From these four sets of voices, six stimulus conditions (3 shift conditions × 2

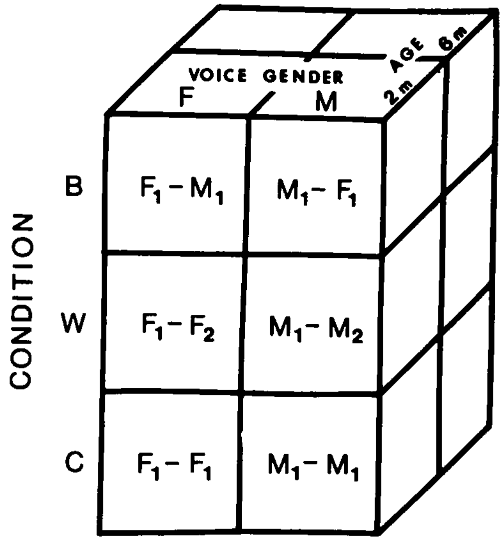

Figure 1. Design of experiment, illustrating distribution of shift conditions (n = 8/cell). Legend: B = Between-category;
W = Within-category;
C = control;
F1, F2 = 2 groups of female voices (6 voices each);
M1, M2 = 2 groups of male voices (6 voices each).

TABLE 1
Peak Fundamental Frequency (F_O) and Mean Fundamental Frequency (Computed Across Entire Stimulus) in Hertz for All Stimuli.

Peak F_O	*F1*	*F2*	*M1*	*M2*
	226	275	186	180
	355	335	165	234
	285	360	331	203
	333	220	260	137
	337	250	222	280
	330	319	117	193
$\bar{X}$	331	293.2	213.5	204.5
Mean F_O	*F1*	*F2*	*M1*	*M2*
	180.7	*217.5*	*123.4*	*121.3*
	256.2	*264.7*	*141.7*	*212.4*
	230.3	*247.5*	*256.0*	*157.7*
	221.9	*185.8*	*217.1*	*118.0*
	246.5	*193.5*	*163.2*	*261.1*
	262.2	*201.2*	*109.2*	*137.2*
$\bar{X}$	233.0	218.4	168.4	168.0

orders) were generated: (a) Between-category = M1/F1 and F1/M1; (b) Within-category = M1/M2 and F1/F2; and (c) Control = M1/M1 and F1/F1. Each set of voices was ordered into a 1/2-hour sequence of randomly-ordered repetitions of the six voices (interstimulus interval = 500 msec). The stimulus tapes were generated such that the two sets of voices constituting a given condition were recorded on separate channels of the same audio tape using the VOCAL program (Gillman & Wilson, 1979). Thus, during the testing session, presenting the stimulus shift condition was simply a matter of changing the output channel of the tape.

Design

The design of the present study was a factorial combination of three factors: (a) age (2 months, 6 months); (b) shift condition (between-category, within-category, control); and (c) gender of familiarization stimuli (male, female). Within each age group of infants, 8 infants received a between-category shift (B), 8 received a

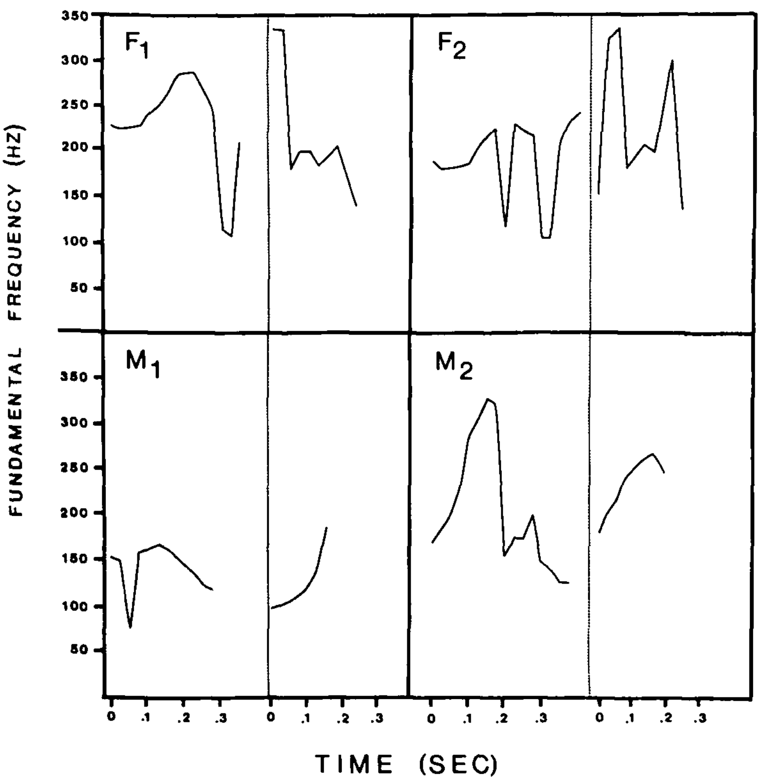

Figure 2. Schematic illustrations of the fundamental frequency contours for representative category tokens. Each category depicted (F1, F2, M1, M2) is represented by two of the six category tokens employed.

within-category shift (W), and 8 received a control condition (C). The gender of the familiarization stimuli was counterbalanced across subjects within each shift condition, i.e., half of the infants were familiarized with male voices and half with female voices. This design is depicted in schematic form in Figure 2.

Apparatus

Infants were tested in a room adjacent (and connected via one-way windows) to the control room, which housed all of the testing apparatus and in which the experimenter and parents viewed the infant via a closed-circuit monitor system. The testing room was painted black. The infant was positioned in an infant seat placed inside a portable crib and surrounded on two sides by royal blue curtains. Directly in front of the infant was a brown masonite panel in the center of which (directly within the infant's field of vision) was a 16 cm × 22 cm opaque screen. The visual stimulus, a blue and red 4 × 4 checkerboard slide with a broken yellow circle in the center, was rear-projected from the control room onto this screen. On each side of the screen was a small peephole covered with screening. The two observers sat behind the panel and viewed the infant through these peepholes. The infant sat approximately 45 cm from the viewing screen.

The auditory stimuli were played to the infant from a Pioneer RT707 stereo tape deck, through a Realistic amplifier, and into an Advent II loudspeaker, which was mounted in the center panel just above (20 cm) the screen on which the infant viewed the slide. The stimuli were played to the infant at 68 ± 1 dBA SPL against an ambient background of 50 dBA, measured by a Bruel and Kjaer sound level meter.

Both the slide projector and the output of the amplifier were coupled to a system of logic that controlled the presentation of the visual and auditory stimuli as a function of the infant's looking behavior. Both observers held a panel of keys that were connected to an Esterline Angus event recorder. Additionally, the keys of one of the observers were coupled to the logic system. One of the keys was depressed when it was judged (by corneal reflection) that the infant was looking at the slide. The depression of this key produced a deflection of the pens on the event recorder and also provided a signal to the logic system which opened a gate to release the auditory stimuli to the speaker. The key was depressed throughout the fixation and the auditory stimuli continued until the key was released, signalling the end of a fixation. When the key was released for a period of two seconds, the slide was automatically turned off, signalling the end of a trial, and then immediately returned to the screen (this process took approximately one second). A new trial began when the infant returned his or her gaze to the slide and the looking key was again depressed.

The keys of the second observer were connected only to the event recorder and were used to assess interobserver reliability (see below). Both observers wore headphones, which delivered music at a comfortable listening level, to mask the presentation of the auditory stimuli delivered to the infant, and a panel placed between the two observers prevented them from seeing the key depressions of the other. Interobserver reliability was computed across the session by counting the number of agreements and disagreements for each .5 sec interval (r = number of agreements/number of disagreements + number of agreements). Only subjects for whom interobserver reliability was at least 80% were accepted for inclusion in the study (mean reliability across all subjects = 87.9%). Two additional keys on the observers' key panels were employed to signal state changes, i.e., sleeping and crying. A video camera was located above and in front of the infant and was connected to a monitor in the control room for viewing by the experimenter and parent(s).

Procedure

Upon arrival at the laboratory, the purposes and procedures of the study were explained to the infant's parents, after which they were requested to sign a consent form giving permission for their infant to participate in the study. Infant and parents were then escorted into the testing room. After the infant was situated in the infant seat, the parents and experimenter left the testing room and the session began.

Each testing session consisted of a familiarization and a shift phase. During the familiarization phase, infants received several trials of auditory stimulation (i.e., male or female voice stimuli) according to the infant-control visual-habituation paradigm described above. A trial (presentation of auditory stimuli) began when the infant fixated a checkerboard slide, and ended when the infant looked away from the slide for a period of 2 seconds. Each trial consisted of randomly-ordered repetitions of six male or six female voices, which continued throughout the infants' fixation. The auditory stimuli were presented *only* during the infant's fixation. If the infant looked away from the slide for a period of less than 2 seconds, the auditory stimuli were discontinued during the look-away period, although the trial was continued. At the end of the trial, the slide was briefly turned off and then returned to the screen. The next trial began when the infant returned gaze to the slide.

During the familiarization phase, the experimenter continuously monitored the infant's progress by computing the duration of each fixation as it occurred. The familiarization phase continued until the infant's looking times habituated to criterion, which was defined as a decrement in looking time on two consecutive trials of at least 50% of the mean of the two peak looking times produced across the familiarization trials.

The shift phase began immediately following habituation to the familiarization stimuli. During the shift phase, the visual stimulus remained the same, while the auditory stimuli were changed to produce either a between-category shift (M1/F1 or F1/M1), a within-category shift (M1/M2 or F1/F2), or a control condition (M1/M1 or F1/F1). The shift phase of the session continued until the infant achieved habituation to the shift (or control) stimuli, at which point the session was terminated. The session was terminated prematurely if the infant fell asleep or cried excessively; the data from these infants were not included in the analyses. In addition, if subjects cried only periodically, but the total amount of crying time exceeded 30 seconds over the course of the session, the data from those infants were also not employed in analyses.

RESULTS

Familiarization

The performance data during the familiarization phase for both age groups of subjects are summarized in Table 2. Presented in Table 2 as a function of age group and shift condition are: (a) number of trials to reach habituation criterion; (b) mean (per trial) looking over the familiarization phase; (c) the mean of the two peak looking times during familiarization; and (d) the mean of the two criterion habituation looks. An analysis of variance, with factors of age (2-month, 6-month), shift condition (B, W, C), and familiarization stimulus (F, M) was performed separately on each of these four measures. Analyses of number of trials, total looking times, and habituation looking times yielded no main effects or interactions. The analyses performed on peak looking times yielded a significant main effect of age, $F(1,42) = 6.78$, $p < .025$, but no other main effects or interactions. As can be seen in Table 2, this main effect is characterized by significantly longer peak looking times by 2-month olds than by 6-month olds.

This effect of age in peak looking times was also apparent in an analysis designed to examine the degree of habituation. In this analysis, both peak looking times and habituation looking times were included in an analysis of variance with factors of age, shift condition, familiarization stimulus, and trials (mean of peak looks vs. mean of habituation looks). As expected, this analysis revealed a highly

TABLE 2
Familiarization Data as a Function of Age and Shift Condition (B = Between, W = Within, C = Control). Values Represent the mean (n = 8) of (1) Total Number of Familiarization Trials, (2) Mean Looking Times Across the Familiarization Phase, (3) Mean of Two Peak Looking Times, and (4) Mean of the Two Habituation Looks (Final Two Familiarization Trials).

	2-Month				*6-Month*			
	Number of Trials	*Mean Looking (sec)*	*Mean Peak Looking (sec)*	*Mean Habit. Looking (sec)*	*Number of Trials*	*Mean Looking (sec)*	*Mean Peak Looking (sec)*	*Mean Habit. Looking (sec)*
B	6.5	25.5	50.0	10.2	7.8	11.1	21.4	5.0
W	7.5	16.9	41.3	5.7	8.1	18.0	42.9	9.1
C	7.1	34.5	89.0	21.2	11.2	16.0	33.8	6.2
$\bar{X}$	7.0	25.6	60.1	12.4	9.0	15.0	32.7	6.8

significant effect of trials, $F\ (1,42) = 79.58$, $p < .001$, which merely confirms the fact of habituation for both age groups and all shift conditions. In addition, however, the main effect of age, $F\ (1,42) = 6.43$, $p < .025$, and the age × trials interaction, $F\ (1,42) = 6.94$, $p < .025$, were also significant. Both of these effects can be attributed to the longer peak looking times of the 2-month olds (cf. Table 2). As this analysis and Table 2 suggest, 2-month-old infants showed a greater decrement of response across familiarization than 6-month olds. Although this analysis provided some evidence of differential habituation between the two age groups, since the age factor did not interact with *shift condition* in any analyses, we may assume that any effects due to shift condition in analyses of discrimination data cannot be attributed to differential performance during familiarization. In addition, there were no reliable effects of familiarization stimulus (i.e., male vs. female) in any analyses, suggesting that male and female voices elicited similar amounts of attention from these infants.

Discrimination

Discrimination was assessed by the extent to which recovery of visual fixation (following habituation) was greater in the shift conditions (B and W) than in the control condition (C). All analyses of discrimination performance employed difference scores computed between the mean of the two habituation looks (i.e., final two familiarization trials) and (a) the mean of the first two post-shift looks (Trial block 1) and (b) each of the first five post-shift trials (because subjects received different numbers of trials, five post-shift trials were chosen for analysis as this was the minimum number of trials received by any infant). These discrimination data are presented as a function of age and shift condition in Figures 3 and 4. Figure 3 displays the difference between the mean of the last two habituation looks and the mean of the first post-shift trial block, and Figure 4 presents the post-shift difference scores for the first five post-shift trials.

An analysis of variance (age × shift condition × familiarization stimulus) performed on the difference scores for post-shift trial block 1 yielded a main effect of shift condition, $F\ (2,42) = 5.01$, $p < .025$, and an age × shift condition interaction, $F\ (2,42) = 3.5$, $p < .025$. Planned comparisons for simple effects suggested by this interaction revealed: (a) for the 2-month olds, significant differences emerged for the between-control (B-C) comparison, $t\ (14) = 3.05$, $p < .01$, and for the within-control (W-C) comparison, $t\ (14) = 2.49$, $p < .05$, but not for the between-within (B-W) comparison, $t < 1$; and (b) for the 6-month olds, significant differences were observed for the B-W comparison, $t\ (14) = 2.20$, $p < .05$, and for the B-C comparison, $t\ (14) = 2.17$, $p < .05$, but not for the W-C comparison, $t < 1$. There were no reliable effects due to the gender of the familiarization stimuli.

Identical effects were observed when the difference score data for the first five post-shift trials were analyzed in a 4-factor (between-subjects factors = age, shift condition, familiarization stimulus; within-subject factor = trials) analysis of vari-

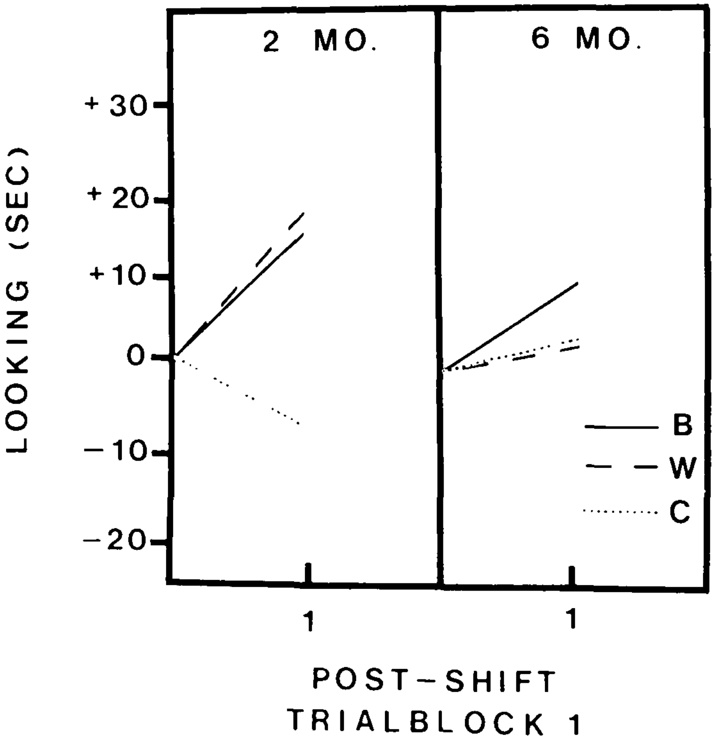

Figure 3. Difference scores computed between the mean looking times (in sec) of the two habituation (final familiarization) trials and the mean of the first two post-shift trials (post-shift trial block 1). Legend: B = Between-category; W = Within-category; C = Control.

ance. Once again, a main effect for shift condition, $F(2,42) = 5.57, p < .01$, and a shift condition × age interaction, $F(2,42) = 4.125, p < .025$, emerged as significant effects. Analyses for simple effects for the age × shift condition interaction revealed the same pattern of reliable effects that was observed in analyses of the first post-shift trial block: (a) 2-month olds: B-C, $t(14) = 2.41, p < .025$; W-C, $t(14) = 2.52, p < .025$: B-W, $t < 1$; and (b) 6-month olds: B-W, $t(14) = 2.79, p < .02$; B-C, $t(14) = 2.15, p < .05$; W-C, $t < 1$. In addition, this analysis yielded a significant shift condition × trials interaction, $F(8,160) = 9.13, p < .001$, and an age × shift condition × trials interaction, $F(16,160) = 2.31, p < .025$. Once again, the gender of the familiarization stimuli did not influence these data.

To summarize, these analyses have shown that looking times of 2-month olds, in both the between-category and the within-category conditions, significantly exceeded responding in the control condition, yet did not differ from each other. For the 6-month olds, in contrast, between-category responding differed significantly

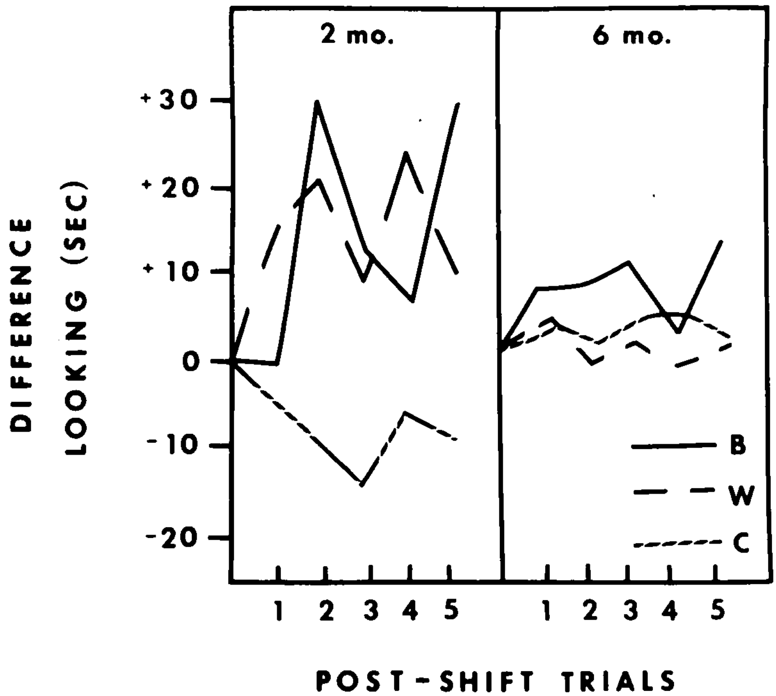

Figure 4. Difference scores computed between the mean looking times (in sec) of the two habituation (final familiarization) trials and each of the first five post-shift trials. Legend: B = Between-category;
W = Within-category;
C = Control.

from both control responding *and* within-category responding. There was no difference between within-category and control post-shift looking times for the 6-month olds. Thus, these data confirm the observation in Figures 3 and 4 that, whereas 2-month olds discriminated both within-category and between-category changes, 6-month olds evidenced discrimination only when the change denoted a category change (i.e., between-category discrimination, but not within-category discrimination).

DISCUSSION

The purpose of the present study was to investigate age changes in the discrimination of male and female voice categories. In conventional speech perception terms, categorical discrimination is inferred if subjects demonstrate better between-category than within-category discrimination. By this criterion we can say that in the present study 6-month-old infants, but not 2-month olds, clearly demonstrated categorical discrimination for male and female voices. Whereas 6-month olds exhibited between-category (but *not* within-category) discrimination, 2-month olds exhibited equal amounts of response recovery (i.e., discrimination) to both between-category and within-category changes. Moreover, insofar as the present study represents a conceptual extension of the conventional categorical discrimination paradigm, we can also argue that 6-month-old infants have demonstrated a form of perceptual constancy across these voice gender categories. That is, categorical responding in this situation (i.e., the use of multiple category tokens during familiarization) seems to require that the infant recognize invariant relationships among male and female category tokens.

The use of a conventional habituation-dishabituation paradigm to examine the question of perceptual constancy also provides indirect evidence concerning the infant's processing of these voice stimuli. The fundamental assumption underlying the use of this paradigm is that the observation of habituation implies that the infant has learned and formed a memory for certain aspect(s) of the familiarization stimuli. An examination of the stimulus conditions that elicit dishabituation then reveals which of these stimulus features the infant has remembered. According to this view, the present results suggest that 6-month olds were extracting and/or storing voice *category* information during familiarization. These results thus replicate and extend the findings of Miller et al. (1982) of voice gender categorization in 6- to 7-month olds. Whereas generalization to novel category tokens was not tested in the Miller et al. (1982) study, the present study did not reveal such generalization, in that when novel voices of the familiar category were presented, habituation continued. Similar evidence of stimulus generalization was not exhibited by the younger infants.

Although these results suggest clear differences between 2- and 6-month olds in their responses to these stimuli, inferences regarding differential *abilities* must not be made too hastily. The present results do not suggest that 2-month olds are insensitive to category information, nor that 6-month olds cannot discriminate within-category tokens.[1] Rather, they suggest that, insofar as the present paradigm addresses differences in attentional qualities of stimulus contrasts, 2-month olds appear to regard both within- and between-category changes as equally salient, whereas 6-month olds regard between-category differences as more important and/or salient than within-category differences. Beyond this, we can only speculate. Despite the moderate nature of this interpretation, however, the present results do bear on the larger question of these infants' processing of speech and language stimuli. The observation that 2-month olds respond equally to within- and between-category speaker differences might also suggest that these young infants are either unable or simply not inclined to ignore irrelevant speaker variations when listening to speech and language. If it is true that forming categories is one way in which we can cognitively "dispose of" irrelevant within-category variation, thereby processing relevant information more efficiently, we can then speculate that the "undue"

[1] Although discrimination of individually presented within-category tokens was not tested directly in this study, it is reasonable to suspect that, under different conditions, these 6-month olds would readily exhibit such discrimination. Since 2-month olds in the present study, and adult listeners and 7-month olds in a previous study (Miller, 1979), discriminated these within-category tokens, there is no reason to believe that these 6-month olds would be any less competent.

attention of 2-month olds to within-category speaker differences might interfere with the more relevant task of attending to and extracting phonetic categories. This speculation would predict superior phonetic constancy for 6-month olds to that of 2-month olds, a prediction that has yet to be tested directly. Perhaps an approach similar to the one employed here might make possible such a direct comparison.

In addition to this possible relationship to the development of speech perception, the results of the present study bear directly upon our growing knowledge of cognitive development in infancy. The study of infant categorization, although relatively recent historically, has rapidly become one of the most active areas of infant research. In the visual domain, numerous studies have documented the infant's increasing awareness of category invariance (e.g., Caron et al., 1979; Cohen & Strauss, 1979; Cohen & Younger, 1981; Cornell, 1974, Ruff, 1977; Younger & Cohen, 1981). Perhaps the most interesting observation arising from these studies is that this type of ability seems to be most apparent between 5 and 7 months of age. The consistency with which this age range is cited across a wide diversity of categories suggests an important and fundamental change in the infant's cognitive approach to the world. In both the auditory and the visual modality, the 6-month old infant's world appears to be a significantly more organized and "abstract" one than the world of the 2-month old.

Considering that many stimulus categories have been formed by 6 months of age, the examination of development in concept formation becomes an even more compelling question. In the present study, we have begun this process by demonstrating that 2-month-old and 6-month-old infants appear to process male and female voice stimuli quite differently. However, the mere demonstration of age differences does not explain how this development happens. We still do not know, for example, to what specific features of voices 2-month-old and 6-month-old infants might be differentially responsive, nor do we know upon what feature(s) 6-month olds base their voice categorization. Male and female voices differ along a variety of acoustic dimensions (Fant, 1973; Peterson & Barney, 1952), most specifically in fundamental and formant frequency characteristics. More systematic examination of the infant's response to these differences, perhaps through the use of synthetic and/or selectively-filtered natural speech tokens, would help elucidate these developmental changes in voice perception.

In addition, although the present paradigm is useful for a variety of reasons, it does not disclose the specific nature of the infant's processing of these stimuli. For example, although the dishabituation data suggest that 2-month olds and 6-month olds recognized different stimulus information about voices, it is not clear whether this information was extracted differently, stored differently, or retrieved differently. In addition, these data cannot tell us whether the lack of within-category discrimination by 6-month olds was a failure to discriminate or was a more active process of equivalence/similarity judgments.

Finally, our assumptions throughout have been that voice perception is intimately related to the perception of speech and language and that age differences in voice perception imply age differences in speech perception. However, there is no direct evidence on this issue, and it certainly is one that requires more direct testing. In examining this relationship, we will need to explore more extensively the development of the psychoacoustic foundation of both voice perception and phonetic categorization, the social/cognitive correlates of both, and the cognitive bases underlying both.

To summarize, the results of the present study have shown that 6-month-old infants respond in a categorical way to male and female voices, whereas 2-month olds do not seem functionally to organize voices in this way. These results thus suggest qualitatively different modes of processing speaker information in 2-month old and 6-month-old infants, which might also suggest that these infants are hearing and processing *language* in fundamentally different ways. The relationship between speaker classification and speech perception in adults (e.g., Rand, 1971) suggest that 6-month olds may be experiencing linguistic input from a more orderly and qualitatively more mature perspective than 2-month olds. The specific nature of these different modes, i.e., to what features and/or patterns of features of male and female voices (and of language in general) infants are responsive, remains to be explored. More importantly, it remains for future research to specify the nature of the developmental changes that are occurring during this time with respect to both the infant's *experiences* with different male and female speakers, as well as the changes that are occurring in the infant's perceptual/cognitive apparatus for processing speaker and other language-related information.

REFERENCES

Caron, A. J., Caron, R. F., & Carlson, V. R. Infant perception of the invariant shape of objects varying in slant. *Child Development*, 1979, *50*, 716–721.

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

Cohen, L. B., & Younger, B. A. *Perceptual categorization in the infant*. Paper presented at the meeting of the Eleventh Annual Jean Piaget Symposium, Philadelphia, May 1981.

Coleman, R. O. A comparison of the contributions of two vocal characteristics to the perception of maleness and femaleness in the voice. *Speech Transmission Laboratory, Quarterly Progress Status Report 2–3*, Royal Institute of Technology, Stockholm, Sweden, 1973.

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

Eilers, R. W., Wilson, W. R., & Moore, J. M. Developmental changes in speech discrimination in three-, six- and twelve-month-old infants. *Journal of Speech and Hearing Research*, 1977, *20*, 766–780.

Eimas, P. D. Auditory and linguistic processing of cues for place of articulation by infants. *Perception and Psychophysics*, 1974, *16*, 513–521.

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

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

Garnica, O. Some prosodic and paralinguistic features of speech to young children. In C. Snow & C. Ferguson (Eds.), *Talking to children: Language input and acquisition*. London: Cambridge University Press, 1977.

Gillman, C., & Wilson, D. *Unified system for the synthesis, presentation, and analysis of speech.* Paper presented at the meeting of the National Conference on the Use of On-Line Computers in Psychology, November 1979.

Horowitz, F. D. Visual attention, auditory stimulation, and language discrimination in young infants. *Monographs of the Society for Research in Child Development*, 1975, *39*(5–6, Serial No. 158).

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

Kuhl, P. K. Perceptual constancy for speech-sound categories. In G. Yeni-Komshian, J. Kavanaugh, & C. Ferguson (Eds.), *Child phonology: Perception and production*. NY: Academic Press, 1980.

Liberman, A. The grammars of speech and language. *Cognitive Psychology*, 1970, *1*, 301–323.

Miller, C. L. Voice recognition and categorization in infants. *Research Status Report, Infant Development Laboratory, University of Wisconsin-Madison*, 1979, *3*, 95–178.

Miller, C. L., & Morse, P. A. The "heart" of categorical speech discrimination in young infants. *Journal of Speech and Hearing Research*, 1976, *19*, 578–589.

Miller, C. L., Younger, B. A., & Morse, P. A. The categorization of male and female voices in infancy. *Infant Behavior and Development*, 1982, *5*, 143–159.

Miller, J. *Phonetic determination of infant speech perception*. Unpublished doctoral dissertation, University of Minnesota, Minneapolis, 1974.

Newport, E. L. Motherese: The speech of mothers to young children. In N. J. Castellan, D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive theory* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

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

Rand, T. C. *Vocal tract size normalization in the perception of stop consonants*. Paper presented at the meeting of the Acoustical Society of America, Washington, D.C., 1971.

Ruff, H. A. *The role of stimulus variability in infant perceptual learning*. Paper presented at the meeting of the Society for Research in Child Development, New Orleans, March 1977.

Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. The prosody of maternal speech: Infant age and context-related changes. *Journal of Child Language*, in press.

Younger, B. A., & Cohen, L. B. Infant perception of correlations among attributes. Unpublished manuscript, University of Texas-Austin, 1981.

Wilson, W. R., Moore, J. M., & Thompson, J. *Sound-field auditory thresholds of infants utilizing Visual Reinforcement Audiometry (VRA)*. Paper presented at the meeting of the American Speech and Hearing Association, Houston, 1976.

28 May 1981; Revised 30 June 1982 ■