Six-Month-Olds' Categorization of Natural Infant-Directed Utterances

David S. Moore
Pitzer College and Claremont Graduate University

Melanie J. Spence
University of Texas at Dallas

Gary S. Katz
University of Pittsburgh

In this study, the authors demonstrated that 6-month-old infants are able to categorize natural, 650 Hz low-pass filtered infant-directed utterances. In Experiment 1, 24 male and 24 female infants heard 7 different tokens from 1 class of utterance (comforting or approving). Then, some infants heard a novel test stimulus from the familiar class of tokens; others heard a test stimulus from the unfamiliar class. Infants categorized these tokens as evidenced by response recovery to tokens from within each category, supporting the conclusion that the results of Experiment 1 indicated categorization. The authors discuss both a mechanism that might explain the development of this ability and the mutual adaptation seen in parent–infant communication.

Interest in the physical properties of ID speech was heightened by the finding that infants prefer ID over AD speech during the first year of life (Cooper & Aslin, 1990; Fernald, 1985; Pegg, Werker, & McLeod, 1992). In addition to maintaining infants' attention more effectively than AD speech, both English ID speech (Werker & McLeod, 1989) and Cantonese ID speech (Werker, Pegg, & McLeod, 1994) produce more positive affect in 4- to 5.5-month-old English-hearing infants. Recent evidence suggests that 4-month-olds might also find ID speech more arousing than AD speech (Kaplan, Goldstein, Huckeby, & Cooper, 1995; Kaplan, Goldstein, Huckeby, Owren, & Cooper, 1995). Some authors have suggested that the attention-maintaining quality of ID speech for 4-month-olds is influenced by the frequency modulations that are characteristic of this speech register (Fernald & Kuhl, 1987; Sullivan & Horowitz, 1983; but see also Colombo & Horowitz, 1986; Cooper, 1993; Cooper & Aslin, 1989).

Studies of parents' ID speech have revealed that specific prosodic patterns are more likely to occur in certain interactional contexts than in others. Parents are more likely to use low-frequency vocalizations with falling frequency contours to soothe fussy infants, whereas they are more likely to use higher frequency, rising contours to elicit a response or attention from an infant or when interacting with infants expressing positive affect (Fernald et al., 1984; Papousek, Papousek, & Bornstein, 1985; Stern et al., 1982). These findings suggest that parents modify their ID speech as a function of the infant's behaviors and affective state, creating opportunities for infants to learn that their behavioral or emotional responses elicit different vocal responses from caregivers. These vocal patterns may communicate the caregiver's affect and intentions and thus may function as the first vehicle for the communication of meaningful information (Fernald, 1992; Papousek, 1992; Stern et al., 1982).

Support for the hypothesis that ID speech patterns may function as communicative signals has recently been provided from studies demonstrating that infants differentially respond to ID...
CATEGORIZATION OF INFANT-DIRECTED UTTERANCES

Kuhl and Hillenbrand (1979) examined infants' categorization of prosodic properties of speech stimuli. They demonstrated that 6-month-olds, presented with a set of vowels (either /a/ or /i/) synthesized to simulate male, female, and child talkers speaking with either a rising or a falling pitch contour, can categorize these tokens exclusively on the basis of their pitch contours. This result demonstrates infants' categorization of synthetic speech stimuli based on a suprasegmental dimension (frequency contour). It suggests that 6-month-old infants also might be able to categorize natural ID speech tokens. However, to date, categorization of natural ID speech tokens by infants has not been demonstrated.

It is important to examine infants' categorization of naturally produced utterances because infants respond differently to synthetic and natural stimuli. Sullivan and Horowitz (1983) reported that 2-month-olds attended more to naturally produced rising contours than to naturally produced falling contours, whereas synthetically produced contours elicited the opposite pattern of attentional preference (i.e., preference for falling contours); they concluded that "what appeared to be minor stimulus differences to the adult listener produced marked differences in infant attention patterns" (p. 531). Given these results, a demonstration of categorization of natural tokens is critically different from a similar demonstration with synthetic tokens.

The current experiments were designed to test for categorization of natural ID utterances by 6-month-old infants. In Experiment 1, our infants were asked to distinguish a variety of utterances of one class (produced by a variety of speakers) from a variety of utterances of a different class (again, produced by a variety of speakers). We chose to use natural approving ID speech tokens and natural comforting ID speech tokens as our two classes of utterance, because neither class of utterance would be aversive to infants, and because these classes of utterance have been shown to be quantitatively discriminable on several variables, including mean $F_0$, $F_0$ variability, frequency contour, and duration of vocalized $F_0$ (Fernald, 1992; Katz, Cohn, & Moore, 1996; Papousek et al., 1991). Furthermore, the use of low-pass filtered tokens prevented infants from using most segmental cues in the discrimination and categorization of ID utterances.

In a modified infant-controlled familiarization-test procedure (Ferland & Mendelson, 1989), infants heard seven different tokens, all belonging to the same class of utterance, across seven trials. Subsequently, half of these infants heard a novel instance from the familiar class of tokens; the other half heard an instance from the unfamiliar class. If, as hypothesized, infants are able to categorize natural ID speech tokens, recovery of response to tokens from the unfamiliar class but not to novel tokens from the familiar class should be seen. In Experiment 2, comparable methods and stimuli were used to establish that similar tokens within each class were, in fact, discriminable from one another, which is a necessary condition for categorization (Ferland & Mendelson, 1989; Olson & Sherman, 1983).

Experiment 1

Method

Participants

The participants were 48 healthy, full-term infants (24 boys and 24 girls) between 5 months 15 days of age and 6 months 12 days of age.
(M age = 185 days). Half of the infants were tested in a laboratory in California, whereas the other half were tested in an identical laboratory in Texas; this approach strengthened the external validity of this study while allowing us to complete data collection more rapidly. All infants were from primarily English-speaking households and were screened for a family history of epilepsy and seizure disorder (because the visual stimuli involved flashing lights). New parents were contacted by mail, and interested parents returned a postcard indicating their desire to participate in the study. Each participant was volunteered by a parent who brought the infant to the laboratory and was present with the infant throughout the procedure. Sixteen additional infants were tested but were not considered as part of the final sample because of fussiness, inattention to the visual stimuli during the session, or experimenter error.

Materials and Stimuli

Equipment. The entire procedure was controlled by custom software that was run on IBM PC-clone (386) computers that were each fitted with an 8-bit Creative Labs (San Francisco, CA) Soundblaster sound card and a Colorgraphics Communications Corporation (Atlanta, GA) VGA+ dual video card. The software assigned infants to groups, chose and presented both auditory and visual stimuli, timed trials, collected and stored looking-time data from an experimenter-operated joystick, and rejected trials deemed to have elicited insufficient looking.

Speech samples. The present investigation used speech samples examined and described in a previous study (Katz et al., 1996); these samples were maternal utterances produced in approving or comforting contexts and directed to 4-month-old infants. Utterances were obtained during a laboratory visit in which mother and infant were outfitted with wireless microphones and were videotaped during a structured set of interactions. Approving interactions required the mother to indicate approval when her infant reached for and grasped a red ring; comforting interactions required the mother to speak to her infant as if he or she were distressed. The instructions given to the mothers described the context of the expected interaction, not the content of the speech. The spoken interactions were then digitized, pending further prosodic and spectral analyses. Individual utterances were extracted from these digital recordings of the dyadic vocal interaction by two independent reliable coders using criteria based on the natural flow of the mothers' spoken words (as in Papousek, Papousek, & Haekel, 1987; see Katz et al., 1996, for more details on reliability and parsing issues).

This procedure yielded a population of approving utterances and a population of comforting utterances. For the current study, eight approving utterances and eight comforting utterances, each spoken by a different person, were drawn from these populations; these tokens were chosen using four criteria, including absence of background noise, superior auditory fidelity, microphonic (as opposed to accelerometric) recording, and duration of at least 1 s. Means for each of the chosen tokens on three of the prosodic variables measured by Katz et al. (1996) are shown in Table 1 in ascending mean F0 order; these variables include mean F0, F0 variability (as indexed by standard deviation), and duration of vocalized F0. To ensure that the chosen tokens were not a biased sample of the larger populations from which they were drawn, we examined the distribution and percentile rankings of these means and found no evidence of systematically exclusive sampling from either the tails or the central region of the population distributions of the various measures. The fundamental frequency contours of the 16 tokens used are plotted in Figures 1 and 2.

We performed Mann-Whitney U tests on each of the three measures to determine if there were a priori differences between approving and comfort tokens. This nonparametric test was used because it does not assume normality of distribution or homogeneity of variance for the sample data and was thus an appropriate test of differences between the small samples of approval (n = 8) and comfort (n = 8) tokens (Roscoe, 1969). The mean F0 of approving utterances was found to be significantly higher than the mean F0 of comforting utterances (Mann-Whitney U = 1, p < .01). In addition, the F0 variability of approving utterances was found to be significantly greater than the F0 variability of comforting utterances (Mann-Whitney U = 13, p < .05). These differences are typical of approving and comforting utterances (Fernald, 1989; Papousek et al., 1991); had we controlled for these differences, the tokens selected would not have been representative of their classes. The two classes of tokens did not differ significantly on the duration variable (p > .10).

To determine if a combination of mean F0 and F0 variability distinguished our comforting and approving tokens, all 16 tokens were rank ordered on these variables, and each token’s position in the distributions relative to the median was examined. Five of 8 comforting tokens were below the median value on both F0 and F0 variability measures (binomial p = .0231), whereas 5 of 8 approving tokens fell above the median value on both F0 variability of comforting utterances (Mann-Whitney U = 13, p < .05). These differences are typical of approving and comforting utterances (Fernald, 1989; Papousek et al., 1991); had we controlled for these differences, the tokens selected would not have been representative of their classes. The two classes of tokens did not differ significantly on the duration variable (p > .10).

During presentation, all auditory stimuli were passed through two low-pass filters placed in series. Frequencies above the nominal cutoff of 650 Hz were attenuated -96 dB/ octave. This filtering effectively removed the linguistic content of the utterances but preserved prosodic characteristics such as frequency contour and rhythmic information.1

1 To empirically assess the intelligibility of the utterances, adults’ transcriptions of the linguistic content of each of the 16 filtered tokens were obtained. Eleven adult listeners were presented the 16 low-pass filtered tokens, each repeated three times, and were asked to transcribe each. They were also instructed to leave an item blank if they could not understand the content. The responses of these 11 adults were compared with the transcriptions of a second group of 11 adults who were presented the same task and instructions but were presented unfiltered recordings of the 16 tokens. The transcriptions for each token were compiled, and the number of agreements on the linguistic content of each token was obtained for each group (filtered vs. unfiltered). Higher agree-

<p>| Table 1 Descriptive Statistics for Token Measurements |
|-------------------------------|------------|-------------|-------------------|</p>
<table>
<thead>
<tr>
<th>Token type and no.</th>
<th>M Variability (SD)</th>
<th>Duration (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comforting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO 3</td>
<td>170.00</td>
<td>29.28</td>
</tr>
<tr>
<td>CO 8</td>
<td>197.23</td>
<td>76.33</td>
</tr>
<tr>
<td>CO 7</td>
<td>217.91</td>
<td>45.56</td>
</tr>
<tr>
<td>CO 2</td>
<td>221.35</td>
<td>20.87</td>
</tr>
<tr>
<td>CO 5</td>
<td>226.02</td>
<td>17.46</td>
</tr>
<tr>
<td>CO 1</td>
<td>227.03</td>
<td>22.66</td>
</tr>
<tr>
<td>CO 6</td>
<td>228.46</td>
<td>24.59</td>
</tr>
<tr>
<td>CO 4</td>
<td>280.34</td>
<td>25.92</td>
</tr>
<tr>
<td>M</td>
<td>221.04</td>
<td>32.83</td>
</tr>
<tr>
<td>SD</td>
<td>31.20</td>
<td>19.51</td>
</tr>
<tr>
<td>Approving</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AP 1</td>
<td>253.52</td>
<td>53.79</td>
</tr>
<tr>
<td>AP 2</td>
<td>340.75</td>
<td>17.70</td>
</tr>
<tr>
<td>AP 6</td>
<td>386.00</td>
<td>279.93</td>
</tr>
<tr>
<td>AP 3</td>
<td>389.41</td>
<td>94.80</td>
</tr>
<tr>
<td>AP 4</td>
<td>417.62</td>
<td>45.46</td>
</tr>
<tr>
<td>AP 8</td>
<td>426.11</td>
<td>89.73</td>
</tr>
<tr>
<td>AP 2</td>
<td>474.44</td>
<td>75.46</td>
</tr>
<tr>
<td>AP 4</td>
<td>511.84</td>
<td>52.47</td>
</tr>
<tr>
<td>M</td>
<td>399.96</td>
<td>88.67</td>
</tr>
<tr>
<td>SD</td>
<td>79.57</td>
<td>81.26</td>
</tr>
<tr>
<td>Mann-Whitney U test</td>
<td>p &lt; .01</td>
<td>p &lt; .05</td>
</tr>
</tbody>
</table>
Thus, infants' discrimination and categorization of these tokens should be based exclusively on prosodic properties of the stimuli.

The filtered stimuli were output through a Realistic speaker with an 8.5-in. (21.59 cm) woofer and a 5.5-in. (13.97 cm) tweeter. The speaker was in the same plane as the visual stimuli at midline in front of the infant, and it produced sounds that averaged approximately 60.5 dB, as measured with a Scott (Acton, MA, Type 451) A-weighted sound level meter at the infant's ears.

Visual stimuli. The visual stimuli were presented on two high-resolution computer monitors; the monitor screens and the space between the screens (both measured horizontally) subtended 14.25 and 23.50 deg of visual angle, respectively. The images on the monitors were identical throughout the experiment; the decision to use two monitors rested entirely on the observation that infants would pay more attention to the visual stimuli when they were simultaneously presented on two monitors than when they were presented on only one monitor.

Each monitor displayed an image of three random checkerboards (Karmel, 1969) presented on a mid-gray background. Each checkerboard subtended approximately 1.5 X 2.0 deg of visual angle. Internal checks subtended approximately 0.33 X 0.26 deg of visual angle, and their initial color was either black or white, as determined by the computer's time-seeded random-number generator. In an effort to increase attention to the displays, the location of each checkerboard was changed before each trial; the screen was conceptually divided into nine sections (3 X 3), and each of the three checkerboards appeared in a randomly chosen one of the nine resulting segments in each trial. Approximately 9 times per second, the color value of internal checks was redetermined by the random number generator; thus, at random instants during a trial, black squares could become white and vice versa. As a consequence, at any given moment, a checkerboard could appear as it did initially, as a negative image of its initial appearance or as an entirely black or an entirely white square. This had the effect of making the checkerboards appear to flash and served the purpose of increasing infants' attention to the displays.

Figure 1. Fundamental frequency contours of the eight stimuli constituting the class of comforting (CO) utterances.
**Design and Procedure**

An infant-controlled familiarization procedure was used in which infants experienced seven familiarization trials followed by one test trial; trials eliciting less than 2 s of fixation were repeated. Infants heard a different speech token in each trial (unless a trial elicited less than 2 s of fixation, in which case the token heard during that trial could be repeated). With the condition that each token was to be presented in only one trial (provided the trial lasted for more than 2 s), order of presentation of the auditory stimuli was randomly determined.

Infants were randomly assigned to one of four groups. Half of the infants initially heard seven different comfort tokens, one during each familiarization trial; the other half heard seven different approval tokens during these trials. Subsequently, half of the infants within each of these two groups heard a novel token from the familiar category during a test trial; the rest heard a randomly chosen novel token from the novel category. Boys and girls were equally distributed across these experimental conditions.

Upon arrival in the laboratory, each caregiver first filled out a questionnaire requesting general information about the child (including health information). The infant was then seated in the caregiver's lap in a dimly lighted testing room, approximately 1 m from a black curtain that contained two holes at the infant's eye level that revealed the computer monitor screens. The curtain also contained one small hole (well above and centered between the monitor holes) through which the entire session could be unobtrusively observed and another hole (below and centered) through which a video camera could record the infant's behavior. Additional black curtains hung to the infant's left and right sides.

Each caregiver was given information concerning the stimuli and procedure and then asked to orient the infant toward the monitors. Caregivers were instructed to attempt to keep the infant in a quiet, alert state but to refrain from interfering with the infant's perception of the stimuli. Finally, caregivers were asked to wear headphones so that they could not systematically influence the behavior of the infants with respect to the auditory stimuli. Throughout the procedure, caregivers heard a loud, continuous audiotape recording of simultaneous presentations of the experimental stimuli. This audiotape was made by recording onto each of two channels on each of two audiotapes randomly ordered sequences of the comfort and approval tokens. These two tapes were then mixed to produce a single audiotape that continuously presented four (random) auditory stimuli simultaneously. This recording prevented caregivers from being able to ascertain which token or type of token the infant was hearing on any given trial.

Before the beginning of each trial, a string of minilights centered between the monitors was flashed on and off until the infant's attention was captured. Once the infant fixated the lights, the lights were extinguished, and the visual stimuli were presented. These stimuli were then displayed continuously until the end of the trial, which occurred either 30 s after the onset of the trial or when the infant looked away from the stimuli for 2.5 consecutive s, whichever came first. Intertrial intervals were typically less than 3 s.

*Figure 2.* Fundamental frequency contours of the eight stimuli constituting the class of approving (AP) utterances. Stimulus AP6 has the Y scale doubled because of the large frequency range.
Whenever an infant was judged to have begun fixating one of the monitors, a randomly chosen speech token from the predetermined category was played through the speaker. As long as the infant continued to fixate a monitor, the token was heard; any time the end of the utterance was reached, the token was replayed from the beginning after a 1-s pause. Whenever the infant was judged to have terminated fixation of the monitors (or at the end of 30 s), the playback of the token was terminated. If, at any point before the end of the trial, the infant refixated the visual displays, the token was replayed from the beginning.

Total duration of looking at the monitors was recorded online during each trial by coders trained to record infants' eye movements. Interrater reliability was evaluated by having the coders from the two different laboratories observe several of the same infants online. A computer program designed to assess reliability checked each observer's joystick at the conclusion of each 100 ms interval (approximately) and recorded a number (1 or 0) that reflected whether the infant was, at that moment, being coded as fixating one of the screens. This process generated two lists of zeros and ones (one list produced by each of the coders), each list a coded interval-by-interval record of the infants' fixation behavior. The agreement across these lists was assessed by computing the phi coefficient; the correlation between these two lists was greater than .95 for each infant seen by both coders.

The coders wore headphones that delivered the same auditory mask heard by the caregivers; these sounds allowed the coders to remain deaf to the auditory stimulus available to the infant in any given trial. The duration of fixation during each trial was stored in a data file, along with information regarding the infant and the stimulus heard during the trial.

**Results and Discussion**

The primary dependent variable of interest was total duration of looking (TDL) at the visual displays during specific trials. Each infant's data were first divided into three trial blocks; Trial Block 1, which consisted of averaged TDL data from Trials 1 and 2, Trial Block 2, which consisted of averaged TDL data from Trials 6 and 7 (the final two familiarization trials), and Trial Block 3, which consisted of TDL data from Trial 8 (the test trial). Consequently, each infant contributed 3 data points to subsequent analyses.

Each infant's data were first entered into a series of three preliminary three-way analyses of variance (ANOVAs; SAS PROC GLM), each designed to assess the effects of one countercoded variable and its interactions with the two variables of primary interest, namely block and experimental group. The countercoded variables examined were (a) familiarization stimulus: comfort versus approval, (b) laboratory: California versus Texas, and (c) sex: male versus female. We conducted separate analyses to examine the effects of each of these variables because our sample of 48 infants was too small to place an adequate number of infants in each cell of a single five-way ANOVA. Thus, each analysis was a 3 (block: Block 1 vs. Block 2 vs. Block 3) × 2 (group: experimental vs. control) × 2 (counterbalanced variable) mixed-design ANOVA, in which block was a repeated measures variable. Each of these analyses revealed a main effect of block across Blocks 1 and 2, \( F(1, 46) = 32.45, p < .0001 \), and a Block × Group interaction, \( F(1, 46) = 4.45, p < .015, MSE = 24.24 \) s, and a Block × Group interaction across Blocks 2 and 3, \( F(1, 46) = 8.97, p < .0045, MSE = 25.86 \) s. Thus, infants (independent of group) looked significantly longer during Trial Block 1 than during Trial Block 2, indicating some measure of habituation across familiarization trials. In addition, infants in the experimental group showed a recovery of fixation from Trial Block 2 to Trial Block 3, whereas the infants in the control group continued to habituate across these trials. This effect was not due to a few unusual infants with extreme TDL scores; 14 out of 24 experimental infants evidenced increased looking across Trial Blocks 2 and 3, in contrast to 7 out of 24 control infants who did so, \( \chi^2(1, N = 48) = 4.15, p < .05 \).

These data indicate that our infants discriminated between the comforting and approving utterances and failed to recover attention when presented with novel tokens from within the stimulus classes. However, to conclusively demonstrate that infants can categorize stimuli, one must show that although they distinguish between stimuli from different classes, they simultaneously disregard discriminable differences that exist among stimuli from within a given class (Ferland & Mendelson, 1989; Table 2: Preliminary Analyses: \( p \) Values Associated With Main Effects of, and Interactions Involving, Counterebalanced Variables

<table>
<thead>
<tr>
<th>Analysis and effect</th>
<th>( p ) value (( &gt; ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group × Block × Familiarization Stimulus</td>
<td>.70</td>
</tr>
<tr>
<td>Group × Familiarization</td>
<td>.92</td>
</tr>
<tr>
<td>Block × Familiarization</td>
<td>.32</td>
</tr>
<tr>
<td>Block × Group × Familiarization</td>
<td>.51</td>
</tr>
<tr>
<td>Group × Block × Sex</td>
<td>.50</td>
</tr>
<tr>
<td>Group × Sex</td>
<td>.86</td>
</tr>
<tr>
<td>Block × Sex</td>
<td>.41</td>
</tr>
<tr>
<td>Block × Group × Sex</td>
<td>.98</td>
</tr>
<tr>
<td>Group × Block × Laboratory</td>
<td>.63</td>
</tr>
<tr>
<td>Laboratory</td>
<td>.24</td>
</tr>
<tr>
<td>Block × Laboratory</td>
<td>.24</td>
</tr>
<tr>
<td>Block × Group × Laboratory</td>
<td>.12</td>
</tr>
</tbody>
</table>

Block 3] might have violated the ANOVA's assumption of homogeneity of variance.) No main effects of, or interactions involving, any of the countercoded variables were discovered, as revealed in Table 2. Therefore, subsequent analyses were conducted on data collapsed across all variables except block and group.

The primary analysis was a two-way (Block × Group) mixed-design ANOVA, in which block was a repeated measures variable with three levels and group was a between-subjects measure with two levels. This analysis confirmed a main effect of block, \( F(2, 92) = 21.43, p < .0001 \), and a Block × Group interaction, \( F(2, 92) = 4.45, p < .015, MSE = 13.21 \) s. Average total durations of looking and standard deviations during each block for control and experimental infants are reported in Table 3.

Because of the difficulty in interpreting an omnibus \( F \) statistic with more than 1 \( DF \) in its numerator (Rosenthal & Rosnow, 1984), planned contrast analyses were conducted as tests of our main hypotheses. These revealed a main effect of block across Blocks 1 and 2, \( F(1, 46) = 32.45, p < .0001, MSE = 24.24 \) s, and a Block × Group interaction across Blocks 2 and 3, \( F(1, 46) = 8.97, p < .0045, MSE = 25.86 \) s. Thus, infants (independent of group) looked significantly longer during Trial Block 1 than during Trial Block 2, indicating some measure of habituation across familiarization trials. In addition, infants in the experimental group showed a recovery of fixation from Trial Block 2 to Trial Block 3, whereas the infants in the control group continued to habituate across these trials. This effect was not due to a few unusual infants with extreme TDL scores; 14 out of 24 experimental infants evidenced increased looking across Trial Blocks 2 and 3, in contrast to 7 out of 24 control infants who did so, \( \chi^2(1, N = 48) = 4.15, p < .05 \).

These data indicate that our infants discriminated between the comforting and approving utterances and failed to recover attention when presented with novel tokens from within the stimulus classes. However, to conclusively demonstrate that infants can categorize stimuli, one must show that although they distinguish between stimuli from different classes, they simultaneously disregard discriminable differences that exist among stimuli from within a given class (Ferland & Mendelson, 1989; Table 2: Preliminary Analyses: \( p \) Values Associated With Main Effects of, and Interactions Involving, Counterebalanced Variables

<table>
<thead>
<tr>
<th>Analysis and effect</th>
<th>( p ) value (( &gt; ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group × Block × Familiarization Stimulus</td>
<td>.70</td>
</tr>
<tr>
<td>Group × Familiarization</td>
<td>.92</td>
</tr>
<tr>
<td>Block × Familiarization</td>
<td>.32</td>
</tr>
<tr>
<td>Block × Group × Familiarization</td>
<td>.51</td>
</tr>
<tr>
<td>Group × Block × Sex</td>
<td>.50</td>
</tr>
<tr>
<td>Group × Sex</td>
<td>.86</td>
</tr>
<tr>
<td>Block × Sex</td>
<td>.41</td>
</tr>
<tr>
<td>Block × Group × Sex</td>
<td>.98</td>
</tr>
<tr>
<td>Group × Block × Laboratory</td>
<td>.63</td>
</tr>
<tr>
<td>Laboratory</td>
<td>.24</td>
</tr>
<tr>
<td>Block × Laboratory</td>
<td>.24</td>
</tr>
<tr>
<td>Block × Group × Laboratory</td>
<td>.12</td>
</tr>
</tbody>
</table>
The values represent seconds of fixation. Mean token; the other half of the infants were familiarized and tested. To that end, we familiarized infants with a discriminating tokens within each category, we conducted a test to provide definitive evidence that the infants were capable of having demonstrated that infants can discriminate multisyllabic words that are identical except for the F0 of the final syllable. To provide definitive evidence that the infants were capable of discriminating tokens within each category, we conducted a test of this hypothesis. To that end, we familiarized infants with a single token and then observed their responses to a novel test token from within the familiar category. Half of the infants in the current study were familiarized with one of the comforting tokens used in Experiment 1 and were subsequently tested with a comforting token from Experiment 1 that most closely matched the global fundamental frequency pattern of the familiarization token; the other half of the infants were familiarized and tested with a comparable pair of approving tokens. We chose to test discrimination of only the most closely matched pair of tokens within each category because discrimination of this pair would be expected to be more difficult than would discrimination of any other pair. We expected infants in both groups to be able to discriminate the tokens constituting these pairs, as evidenced by recovery of visual fixation to the novel test tokens.

Experiment 2

Six-month-olds are known to discriminate auditory stimuli that are far more similar than those constituting each class in Experiment 1 (Bull et al., 1985; Eilers et al., 1984; Jusczyk et al., 1993; Karzon & Nicholas, 1989; Kuhl & Miller, 1975, 1982; Mehlert et al., 1988; Morse, 1972; Sullivan & Horowitz, 1983; Thorpe, 1986). For example, as already noted, Bull et al. (1985) have demonstrated that infants can discriminate multisyllabic words that are identical except for the F0 of the final syllable.

Participants

The participants were 16 healthy, full-term infants (8 boys and 8 girls) between 5 months 16 days of age and 6 months 13 days of age (M age = 186 days). Infants were recruited and screened as in Experiment 1. Twelve additional infants were tested but were not considered as part of the final sample because of fussiness (n = 3), inattention to the visual stimuli during the session (n = 2), failure to become familiarized to the stimuli (as evidenced by a decrement of attention across Blocks 1 and 2 that was less than a pre-established 5% criterion, n = 5), or experimenter error (n = 2).
tokens from within classes. The results of these studies suggest that, as predicted, 6-month-old infants are able to categorize natural low-pass filtered ID speech tokens, using prosodic cues. Insofar as these results represent the first demonstration of infants' ability to categorize naturally produced ID speech tokens, they extend the findings of Kuhl and Hillenbrand (1979) because these investigators demonstrated categorization of synthetic speech stimuli.

The fact that infants familiarized with approving tokens in Experiment 1 recovered visual attention in response to contingent presentation of a comforting token is particularly strong evidence that they were categorizing the utterances, because approving vocalizations have been shown to be able to recruit infants' attention (Papousek et al., 1990). Experimental infants familiarized with approving tokens attended more to a novel comforting utterance than control infants familiarized with approving tokens attended to an equally novel, and presumably inherently more attractive, approving utterance. Thus, it seems clear that the control infants recognized the categorical similarity between the novel approving token and the series of approving familiarization tokens inasmuch as they found the novel approving token to be relatively uncompelling.

What features of the stimuli did the infants use to categorize, and how should the categories they used be characterized? The low-pass filtering manipulation attenuated cues for phonemic distinctions that are conveyed primarily by higher frequencies, while preserving suprasegmental properties of the stimuli such as fundamental frequency characteristics, rhythm, and amplitude modulation. In addition, adults asked to identify the linguistic content of the low-passed stimuli were unable to do so. Thus, it is most likely that some subset of suprasegmental properties subserved categorization of these utterances.

Our analyses of the differences between the two subpopulations of tokens used in this study revealed the differences in mean F0 and F0 variability noted above. Accordingly, these prosodic features, which are relatively reliable cues for discriminating comforting and approving utterances (Fernald et al., 1989), were available to be used by the infants in their discrimination and categorization of these stimuli. Other prosodic features—for instance, the shapes of utterances' frequency contours—typically distinguish different types of ID utterances as well (Fernald et al., 1984; Papousek et al., 1985; Stern et al., 1982); in fact, different frequency contours characterized the populations of comforting and approving utterances from which stimuli in the present study were chosen (Katz et al., 1996). The comforting and approving stimuli in the present study did not have frequency contours that were significantly different from one another (as assessed by t tests on the curve-fit indices described by Katz et al.), but it is possible that these, and other, prosodic characteristics might also be useful in infants' categorization of natural speech and that the presence of such additional distinguishing characteristics would simply facilitate categorization.

Because mean F0 and F0 variability were the only variables that distinguished our comforting and approving tokens, it is conceivable that the infants were categorizing strictly on the basis of these prosodic characteristics. For example, it is possible that the category constructed by the infants that contained comforting utterances was not an exclusive category of only comforting utterances. Although this broader category would have included comforting utterances and excluded approving utterances, it might also have included noncomforting utterances characterized by a combination of low mean F0 and F0 variability. Therefore, our current data indicate that 6-month-olds are at least able to sort natural ID utterances into two broad categories; these categories allow them both to discriminate approving from comforting utterances and to ignore discriminable differences among various utterances within each category. In addition, the possibility remains that infants used mean F0 and F0 variability to form exclusive categories of approving and comforting ID utterances. We are currently planning additional research that will further examine the exclusivity of these categories. If these studies reveal that 6-month-olds' categories are broad, we will subsequently examine the developmental process by which such categories are ultimately refined.

The development of an infant's ability to categorize natural utterances using prosodic cues might unfold as a function of an associative mechanism of the sort proposed by Lewis (1936/1951). Babies hear utterances with nonidentical, but categorically similar, prosodic characteristics often repeated (Fernald & Simon, 1984; Papousek et al., 1985) during episodes in which specific types of interactions occur. For example, when trying to calm distressed infants, adults tend to use low frequency, continuous utterances with long durations and slowly falling frequency contours (Papousek et al., 1991) probably because these sounds most effectively achieve that end (Birns, Blank, Bridger, & Escalona, 1965). Thus, it is possible that by 6 months of age, associative correspondences are built up between classes of sounds and situations (which, for an infant, include maternal affective behavior) and that the repetitive nature of ID speech (Fernald & Simon, 1984; Papousek et al., 1985) facilitates infants' detection of invariant attributes of ID speech commonly produced within similar interactional contexts.

If a group of nonidentical but categorically similar utterances can all directly soothe an infant, and so are always heard in situations involving the reduction of infant distress, the infant might be able to learn that these diverse utterances all share the same meaning, namely, that it is the speaker's intention to soothe. The present data do not allow us to address the issue of whether infants in this study understood or categorized the meaning or affective value of the ID utterances (Fernald, 1992); however, the question of 6-month-olds' abilities in this domain has been addressed by both Cooper (1993) and Fernald (1993). Cooper pointed out that infants must perceptually differentiate F0 contours from the stream of caretaker speech before they can learn their meanings; her data (Cooper & Aslin, 1994) are consistent with the hypothesis that such differentiation has begun by 4 months of age. Fernald's (1993) data suggest further that by 6 months of age, infants may, in fact, begin to understand the communicative intent of ID utterances. Thus, it remains...
possible that infants learn to comprehend the meaning signaled by ID utterances by way of the mechanism outlined above.

Papousek et al. (1987) noted that they viewed “the simplification of parental registers . . . as a plausible adjustment to children’s constraints in perceptual and productive linguistic skills at the beginning of speech acquisition” (p. 493) and that “acoustic properties of melodic patterns in baby talk are well adapted to the infant’s auditory thresholds, discriminative capacities, preferences, and holistic processing modes” (p. 511). In addition, it is clear from the current research that by 6 months of age, an infant’s ability to categorize ID utterances is well-adapted to the variability that is characteristic of parental communication. Given this variability, categorization is a prerequisite for learning to comprehend an otherwise bewildering array of human utterances. It is therefore possible that the ability of 6-month-olds to categorize ID utterances develops as a result of exposure to adults’ categorical speech behavior.

References


Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant percep-

tion for motherese speech. *Infant Behavior and Development*, 10, 279–293.


CATEGORIZATION OF INFANT-DIRECTED UTTERANCES

Received August 1, 1996
Revision received May 27, 1997
Accepted May 27, 1997

Dannemiller Appointed Editor of

Developmental Psychology, 1999–2004

The Publications and Communications Board of the American Psychological Association announces the appointment of James L. Dannemiller, PhD, University of Wisconsin, as editor of Developmental Psychology for a 6-year term beginning in 1999.

Effective January 1, 1998, manuscripts should be directed to

James L. Dannemiller, PhD
Developmental Psychology Journal Office
Room 555 Waisman Center
University of Wisconsin—Madison
1500 Highland Avenue
Madison, WI 53706-1611
email: jldannem@facstaff.wisc.edu