Journal of Speech and Hearing Research, HILLENBRAND, *Volume 26,* 268-282, June 1983

PERCEPTUAL ORGANIZATION OF SPEECH SOUNDS BY INFANTS

JAMES HILLENBRAND
Northwestern University, Evanston, Illinois

An operant head-turn procedure was used to test whether 6-month-old infants recognize the auditory similarity of speech sounds sharing a value on a phonetic-feature dimension. One group of infants was reinforced for head turns when a change occurred from a series of repeating background stimuli containing nasal consonants ([m, n, ŋ]) to repetitions from a category of syllables containing voiced stop consonants ([b, d, g]), or to a change from stops to nasals. The stimuli were naturally produced by both male and female talkers. The performance of infants in this "phonetic" group was compared to that of infants in a "nonphonetic" control group. Using the same procedures, these infants were reinforced for head turns to a group of phonetically unrelated speech sounds. Results indicated that the performance of infants in the group trained on phonetically related speech sounds was far superior to that of infants in the nonphonetic control group. These findings suggest that prelinguistic infants can perceptually organize speech sounds on the basis of auditory properties related to feature similarity.

A major focus of speech-perception research over the past several decades has been an attempt to define phonetic categories in terms of acoustic properties—for example, to specify the acoustic attributes that define or "cue" the segment [g], or the feature [velar], in all the contexts in which it occurs. Much of the literature in this area has suggested that the critical cues to phonetic categories are often highly variable with changes in "context." The physical cues to speech-sound categories have been found to vary with changes in noncritical dimensions such as the phonetic environment in which the segment appears, the position that the segment occupies within the syllable, and the talker who produces the utterance. These results, combined with a variety of other findings, have led some investigators to theorize that the cues to phonetic categories are not derived from the physical signal in a direct way. Specifically, the suggestion has been made that the perception of speech is mediated in some way by knowledge of how speech is produced. According to this view, the speech waveform is assumed to be interpreted in terms of the articulatory gestures that were used to produce the signal (Liberman, 1970; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Stevens & House, 1972).

Other investigators have argued that attempts to relate phonetic categories to the acoustic signal have failed to account seriously for the psychophysical processes involved in the coding of complex auditory signals. According to this point of view, invariant acoustic cues to phonetic categories can, in fact, be derived from the physical signal without appealing to articulatory knowledge (Fant, 1967; Kuhl, 1979a; Miller, Engebretson, Spenner, & Cox, 1977; Searle, Jacobson, & Rayment, 1979; Stevens & Blumstein, 1978).

Speech-perception research with infants can provide specific kinds of evidence on the contention that articulatory knowledge is a necessary condition for the categorization of speech sounds. The reasoning is relatively simple: Since prelinguistic infants are not assumed to possess sophisticated knowledge about the production of speech, demonstrations of phonetic categorization by infants will indicate the limits of the type of articulatory knowledge likely to be involved in this process. In a recent series of experiments, Kuhl and her associates (Kuhl & Miller, 1982; Kuhl, 1977; 1979b; Holmberg, Morgan, & Kuhl, 1977; Kuhl & Hillenbrand, Note 1), attempted to determine the extent to which young infants recognize similarities among speech sounds when variations are introduced in noncritical dimensions. For example, an experiment by Kuhl (1979b) demonstrated that 6-month-old infants could detect a change from one category of vowels to another when the tokens varied randomly in talker and pitch contour. Infants in this experiment were initially trained to make a head turn for a visual reward when a change occurred from repetitions of a single token of [a], synthesized to simulate a male voice with a falling pitch contour, to repetitions of a single token of [i], produced by the same male "talker" with the same pitch contour. The infants were then gradually exposed to a number of novel tokens synthesized to simulate female and child talkers with either falling or rising pitch contours. The results showed that infants readily transferred learning from the tokens produced by the male talker to the novel tokens produced by female and child talkers.

Similar experiments have tested the perception of an [a]-[ɔ] contrast across variations in talker and pitch contour (Kuhl, 1977), fricative contrasts across variations in vowel context and talker (Holmberg et al., 1977), a nasal-consonant place contrast across variations in vowel context and talker (Hillenbrand, 1980, Note 2), and, using a different version of the operant head-turn procedure, a stop-consonant place contrast across variations in vowel context (Fodor, Garrett, & Brill, 1975).

To date, infant research on phonetic categories has focused exclusively on the infant's ability to recognize

© 1983, American Speech-Language-Hearing Association

0022-4685/83/2602-0268$01.00/0

phonetic similarity at the level of the phone, or phonetic segment. The purpose of the present study was to extend these findings and to test infants on their ability to organize speech sounds at the more abstract level of the phonetic feature. The feature contrast was a stop/nasal distinction: [b, d, ɡ] versus [m, n, ŋ]. This contrast seemed like a logical starting point for testing feature perception in infancy for two reasons. First, a good deal is known about the physical correlates of this distinction. During the occlusion portion of nasal consonants, a nasal murmur is produced that is characterized by (a) a low-frequency first resonance at 200–300 Hz, well separated from higher formants; (b) relatively high damping factors (large formant bandwidths and low formant levels); and (c) an antiformant that varies in frequency with place of articulation (Fant, 1960; Fujimura, 1962). Voiced stop consonants, on the other hand, (a) do not show a nasal murmur (although a low-frequency "voice bar" may be present during the occlusion), (b) are characterized by aperiodic release bursts, and (c) typically show more rapid changes in amplitude following release than nasal consonants (Fant, 1960). A second reason for studying the stop/nasal contrast is that information is available on infants' discrimination of stop and nasal consonants. Evidence is available to show that infants can discriminate individual pairs of speech sounds differing in stop-consonant place of articulation (Eimas, 1974; Morse, 1972), nasal-consonant place of articulation (Hillenbrand, Note 2), and a stop-nasal manner-class contrast (Eimas & Miller, 1980).

The present study examined the ability of infants to categorize speech sounds according to the stop-nasal distinction. In other words, the study was designed to determine whether infants recognize that the stops [b, d, ɡ] are similar to one another and distinct from a class consisting of the nasals [m, n, ŋ].

METHODS

The general approach of the study was similar to the transfer-of-learning experiments by Kuhl and her colleagues (Kuhl, 1977; 1979b; Holmberg et al., 1977; Kuhl & Hillenbrand, Note 1). One group of 6-month-old infants was visually reinforced for head-turn responses when a change occurred from a background category of syllables containing nasal consonants ([m, n, ŋ]) to a comparison category of syllables containing voiced stop consonants ([b, d, ɡ]), or to a change from stops to nasals. The speech sounds were produced by both male and female talkers. The performance of infants in this "phonetic" group was compared to the performance of a separate group of infants run in a procedurally identical "nonphonetic" condition. These infants were tested using the same pool of stimuli used in the phonetic condition, but the stimuli were assigned to reinforced and unreinforced categories in such a way that the categories could not be organized according to phonetic attributes or talker.

The procedure, which is described in detail below, used a visual reward to train an infant to make a head-turn response when a change occurred from a class of repeating background stimuli to repetitions from a comparison category. The experimental stages for the phonetic condition are shown in Table 1. The first stage contrasted a single token of [ma] with a single token of [ba]. Both syllables were naturally produced by the same male voice. In the second stage, postdental consonants were added to each class; that is, [ma] and [na] were contrasted with [ba] and [da]. In the third stage, labial and postdental consonants produced by a female voice were added to each category. In the fourth and final stage, velar consonants were added to each class, resulting in a contrast between male and female [m, n, ŋ] and male and female [b, d, ɡ]. Half of the infants were trained with the stop consonants as the comparison category, and half were trained with the nasal consonants as the comparison category.

TABLE 1. Experimental stages for the phonetic condition.

Stage	*Category 1*	*Category 2*
1: Initial training	ba (M)	ma (M)
2: Place variation	ba (M)	ma (M)
	da (M)	na (M)
3: Talker × Place	ba (M)	ma (M)
	da (M)	na (M)
	ba (F)	ma (F)
	da (F)	na (F)
4: Transfer of learning	ba (M)	ma (M)
	da (M)	na (M)
	ga (M)	ŋa (M)
	ba (F)	ma (F)
	da (F)	na (F)
	ga (F)	ŋa (F)

In the final stage of the experiment, the infant's task was to make a head-turn response whenever a change occurred from a category of nasal consonants to a category of voiced stop consonants—or from stop consonants to nasal consonants—independent of random variation in place of articulation and talker. If subjects in this task succeeded in responding to the stimuli in the comparison category, it would be tempting to conclude that the infants recognized the similarity of speech sounds sharing a phonetic-feature value. It is possible, however, that infants might simply memorize which tokens were reinforced and which ones were not. Memorizing tokens, of course, would not necessarily require a perceptual grouping of the stimuli. To test for this possibility, the performance of infants run in the phonetic task described above was compared to the performance of a separate group of infants run in a nonphonetic condition. In the nonphonetic condition categories were arranged in such a way that the six stimuli in each class could not be organized according to phonetic or acoustic characteristics. Subjects were tested using the same procedures and equipment, plus the same pool of 12 stimuli as in

TABLE 2. Experimental stages for the nonphonetic condition.

Stage	*Category 1*	*Category 2*
1	ba (F)	na (M)
2	ba (F) ŋa (M)	na (M) ga (F)
3	ba (F) ŋa (M) da (F) ma (M)	na (M) ga (F) ma (F) ba (M)
4	ba (F) ŋa (M) da (F) ma (M) ga (M) na (F)	na (M) ga (F) ma (F) ba (M) da (M) ŋa (F)

the phonetic condition. The experimental stages for the nonphonetic condition are shown in Table 2. Subjects were initially trained on a relatively gross contrast between a male [na] and a female [ba]. The subsequent stages were analogous to those of the phonetic condition in terms of the number of tokens added in each stage. However, sounds were added in such a way that, by the final stage, it was not possible to organize the stimuli along any simple dimension: Each class included an equal number of stops and nasals, male voices and female voices, labials, postdentals, and velars. As in the phonetic condition, half of the subjects were trained with category 1 as the comparison class and the other half with category 2. It was reasoned that the only way an infant could succeed on this task was to memorize which individual stimuli were reinforced and which ones were not. If the performance of infants in the phonetic group proved to be superior to that of the nonphonetic group, the effect could be attributed to perceptual categorization of the speech sounds by infants in the phonetic group.

Stimuli

The stimuli were naturally produced tokens of [m, n, ŋ, b, d, g] in prevocalic position with the vowel [a]. One adult male and one adult female produced several tokens of each syllable. Audio recordings were made in a sound-treated booth with a cardioid microphone (Sennheiser MKH 415T-U) and a high-quality full-track recorder (Nagra 4.2). The talkers were instructed to produce all stimuli with approximately equal durations, intensities, and slightly falling pitch contours. A VU meter was used to monitor intensity. The recorded stimuli were digitized and stored in the disk memory of a digital computer (DEC PDP 11/10). A sample rate of 20 kHz was used with a maximum amplitude resolution of eight bits within a ±4-V dynamic range. All signals were low-pass filtered at 8 kHz and conditioned with an autocorrelator noise-reduction device (Phase Linear 1000). One token of each syllable produced by the two talkers was selected for use in the discrimination tests. The tokens were chosen by selecting those stimuli that showed the closest match on computer-derived measurements of fundamental frequency contour, intensity contour, and duration. In the final set of stimuli there were no systematic differences between the stop and nasal categories in fundamental frequency, overall RMS intensity, or duration. (Measurements of these stimuli are given in Table A of the Appendix.) Formal listening tests showed that all stimuli were identified reliably by a panel of five adult listeners.

Audiotapes for discrimination testing were prepared by recording stimuli from the two categories on separate channels of tape. At the output of the D/A converter, the stimuli were low-pass filtered at 8 kHz, conditioned with an autocorrelator noise-reduction device (Phase Linear 1000), and recorded with a constant 1.7-sec onset-to-onset interstimulus interval. The onsets of the stimuli on the two channels of each tape were synchronized using a cueing procedure described by Hillenbrand, Minifie, and Edwards (1979). Gain settings at the input to the tape deck (TEAC 3340-S) were adjusted so that the two stimuli that had contrasted in the initial-training stage balanced for loudness.

Calibration

Signals were calibrated by a combination of sound-level measurements and a loudness-balance procedure. The gain setting at the output of the tape deck was adjusted so that the peak intensity of one syllable in the initial-training pair measured 65 dBA, using the fast-response setting of a sound-level meter (Bruel & Kjaer, Model 2209). A loudness-balance procedure was used to adjust the output gain of the channel carrying the contrasting syllables. An experimenter used an electronic switch to alternate between the two channels. The output gain of the channel carrying the contrasting syllables was adjusted until one adult listener judged that the two signals were equally loud. These same gain settings were used for the experimental conditions involving multiple tokens of the two categories. The loudness balance was checked as part of the daily calibration procedure.

Procedures

1. General. A schematic of the experimental site is shown in Figure 1. The infant was held on the parent's lap facing an assistant. An experimenter in an adjacent room controlled the equipment and was able to observe the infant on a video monitor. A loudspeaker (Electro-Voice SP-12) was positioned at a 90° angle to the assistant. In front of the speaker was an electrically operated stuffed toy bear in a smoked plexiglass box. When activated, the box was illuminated and the bear tapped on a drum.

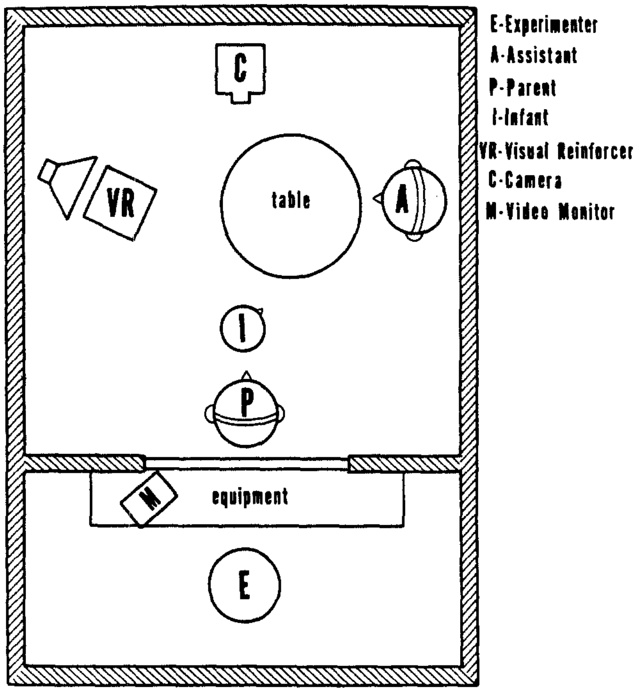


FIGURE 1. Experimental site for the visually reinforced head-turn procedure (from Kuhl, 1979b).

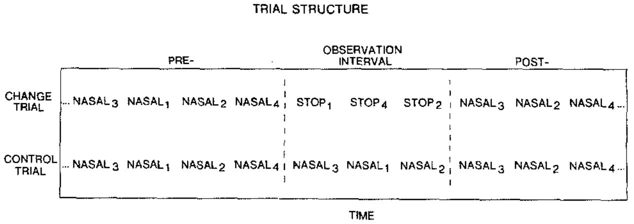


FIGURE 2. Trial structure for the phonetic condition. The figure shows stimuli being presented before, during, and after change and control trials. The subscripts refer to the individual stimuli in the background and comparison categories. The example shown here is for stage 3 of the phonetic condition in which the stop category was reinforced (after Kuhl, 1979b).

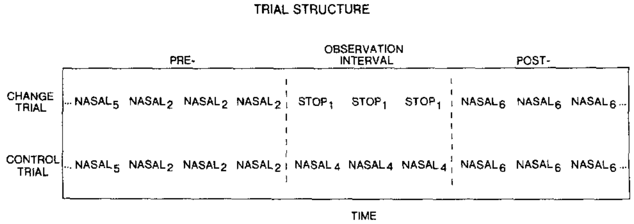


FIGURE 3. Trial structure for the final stage of testing (stage 4). The stimuli are presented in random order, but each stimulus in the order is repeated three times (see Kuhl, 1979b).

The experiment was run with a tape deck (TEAC 3340-S) and a logic device. Throughout the entire experiment, tape-recorded stimuli were continuously presented at onset-to-onset intervals of 1.7 sec. The assistant's task was to keep the infant's attention by manipulating silent toys. When the assistant judged the infant to be in a "ready state," that is, quiet and attending to the toys, he pressed a button signaling the experimenter to initiate a 5-sec observation interval. Two kinds of trials could occur during the interval: change trials or control trials. Figure 2 shows stimuli being presented before, during, and after change and control trials for the phonetic condition. During a change trial, a silent switch initiated a change in tape-recorder channels from the repeating background category to three presentations from the comparison category. A hand-held vibrotactile device signaled the start of a 5-sec observation interval to the assistant; a small light mounted on the monitor signaled the start of the interval to the experimenter. If both the experimenter and the assistant judged that a head turn occurred during the observation interval, they independently pressed buttons that activated the visual reinforcer for 3 sec. And-gate circuitry ensured that the reinforcer would be activated only on change trials in which both judges voted during the 5-sec observation interval. During a control interval, the infant continued to hear stimuli from the background category. On control trials, both the experimenter and the assistant made a judgment about the occurrence of a head turn, but reinforcement was not provided, regardless of the infant's response. For the final stage of testing (stage 4), stimuli were presented using a special three-repetition trial structure described by Kuhl (1979b). As shown in Figure 3, the stimuli were presented in random order, but each stimulus in the order was repeated three times. Since a single token was presented on any given trial, this format made it possible to assign the infant's response to a particular stimulus. On both change and control trials the experimenter recorded the stimulus that was presented and the infant's response.

For all stages of the experiment an infant's performance was measured by comparing the proportion of head turns on change trials to the proportion of head turns on control trials. To reduce the possibility that the parent or assistant might cue the infant's response, and to control for bias in judging head turns, music was presented over earphones to both adults in the test room at a level sufficient to mask a change from one stimulus to another. The experimenter was able to hear the stimuli over an audio monitor in the control room and therefore could have been biased in his judgment of head turns. Experimenter bias in this task would be revealed by his failure to agree with the assistant, who was unbiased. Interjudge agreement for all trials was 98%, indicating that experimenter bias did not play a large role in the judgment of head turns. When the two judges did fail to agree, the trials were always scored as errors. As a further effort to reduce the possibility of bias, an electronic probability generator, set at 50%, was used to determine whether a given observation interval would be a change or control trial. Since previous work with the head-turn procedure suggested that long strings of change and control trials increased the probability of infant errors, the experimenter was instructed to override the probability generator for a single trial after three consecutive change or control trials (see Kuhl, 1979b).

2. *Conditioning the head-turn response.* The head-turn response was conditioned by initiating a change trial and, after a few presentations of the comparison stimulus, activating the visual reinforcer. After a variable number of these trials, most infants began to make head turns that anticipated the activation of the visual reinforcer. To be included in the experiment, an infant was required to make three consecutive anticipatory head turns. Subjects were allowed a maximum 25 trials to meet the conditioning criterion. Testing on the initial-training stage was not begun until the infant met the conditioning criterion. Experience with the head-turn procedure has shown that infants who meet the conditioning criterion very quickly will sometimes perform poorly on the initial-training stage. For that reason, all infants were given a minimum of 15 conditioning trials.

3. *Progressing subjects through the experiment.* An infant advanced from one stage of the experiment to the next when he/she met an accuracy criterion of 9 correct responses in 10 consecutive trials, half being change trials and half being control trials. If an infant did not meet this 9-out-of-10 criterion in 20 trials, he/she was automatically progressed to the next stage of the experiment. When an infant reached the final stage of the experiment, he/she was given as close to 75 trials as possible. A variety of problems prevented this in some cases, including scheduling difficulties, experimenter error, and infants who had become fussy after prolonged testing. The number of trials run on the final stage ranged from 63 to 75, with an average of 68.9 trials.

4. *Retraining.* It was often the case that infants at various stages of testing would show a marked drop in performance. In many cases the infant appeared to have forgotten the experimental contingencies or seemed to lose interest in the task. Infants were retrained by the presentation of conditioning trials—change trials in which the visual reinforcer was manually activated if the infant did not respond within about 4 sec of the stimulus change. Two rules controlled the presentation of these retraining trials:

1. A single retraining trial was presented after three consecutive misses on change trials.
2. If after the first 15 trials of a session an infant had missed more than half of the change trials, the next five trials were retraining trials. Regardless of the stage of testing that the infant was in, these retraining trials used the pair of stimuli from the initial-training stage.

5. *Testing sessions.* A test session was terminated when either the experimenter or the assistant judged that the baby was becoming tired or fussy or at the end of 30 trials. Testing sessions lasted about 10–15 minutes, with an average of 20 trials per session. Infants were usually given all of the trials for a particular experimental stage within the same session. However, if a session was terminated before an infant completed testing on a given stage, testing on the next session would resume where the infant left off. Seven or eight sessions were generally required to complete the experiment.

Subjects

The subjects were normal 5½- to 6½-month-old infants selected by mail solicitation to parents in the Seattle area. A parent questionnaire was used to screen out infants who (a) had been treated for middle-ear problems, (b) had a family history of congenital hearing loss, or (c) were born more than 2 weeks premature or 2 weeks late. Subjects were assigned randomly to either the phonetic or the nonphonetic group. A total of 23 subjects began testing. Subjects were run until eight infants completed testing in each group. To be included in the study, an infant had to pass the conditioning criterion of three consecutive anticipatory head-turn responses in the first 25 trials of testing. Six subjects failed to pass the conditioning criterion on the [ma]-[ba] contrast for the phonetic study. One additional subject in the phonetic group was eliminated due to an experimenter error, leaving seven subjects in this group instead of eight. The nonphonetic condition offered subjects a much grosser, multidimensional contrast, consequently, only one subject in the nonphonetic condition failed to pass conditioning in the allotted 25 trials.

RESULTS

The most interesting results of this study come from an analysis of the babies' responses on the final stage of each condition. These analyses are discussed first, followed by a description of the infants' performance on the preliminary stages. Figure 4 displays the percentages of head turns on change trials and on control trials for infants in the phonetic and nonphonetic groups for the final stage of testing. The graph shows that more head turns were observed on change as opposed to control trials for both groups of infants. The trial-type effect, however, was much more pronounced for the phonetic group. Infants in the two groups responded about equally often on control trials, but the phonetic infants responded much more often on change trials than the nonphonetic infants. A two-way analysis of variance for trial type and group, with repeated measures on the trial-type variable, revealed significant main effects for both trial type ($F = 17.4$; $df = 1, 13$; $p < .001$) and group ($F = 8.0$; $df = 1, 13$; $p < .01$). There was also a significant group × trial-type interaction ($F = 7.2$; $df = 1, 13$; $p < .05$), indicating that the trial-type effect was significantly larger for the phonetic group. Post hoc analysis showed that the trial-type effect was statistically reliable for both the phonetic group ($F = 11.9$; $df = 1, 6$; $p < .01$) and the nonphonetic group ($F = 8.0$; $df = 1, 7$; $p < .05$). These comparisons indicate that infants in both groups performed significantly above chance on the final stage of testing, but that infants in the phonetic group performed with greater accuracy than those in the nonphonetic group.

It was also of interest to determine specifically how the subjects distributed their responses among the individual sounds in the reinforced and unreinforced

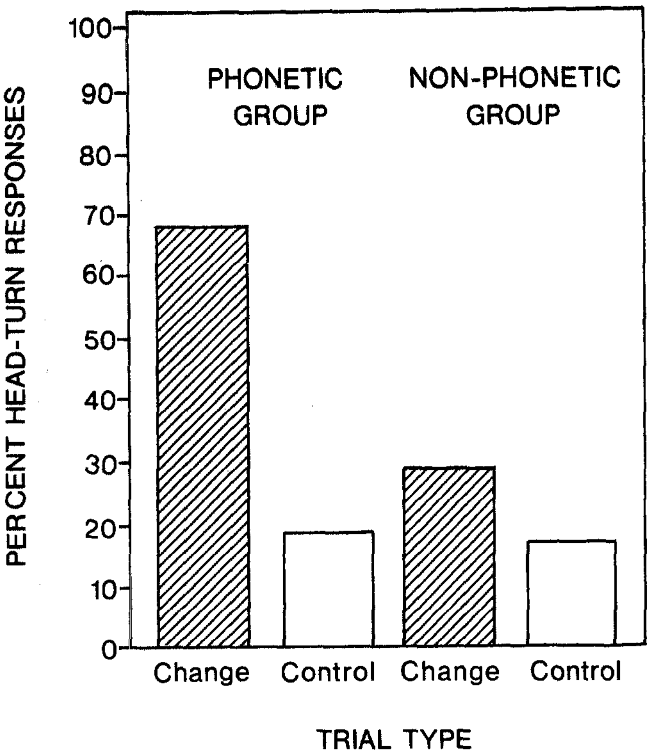


FIGURE 4. Percent head-turn responses on change and control trials for infants in the phonetic (n = 7) and nonphonetic (n = 8) groups. The data in this figure and in Figures 5–11 are from the final stage of testing (stage 4).

categories. Figure 5 presents these data for the three infants in the phonetic group who were trained to turn to the stop category. The six shaded columns to the left show the percentage of head turns to each of the six stop consonants presented on change trials; the six unshaded columns to the right show the same data for the six nasal consonants presented during control intervals. The stimulus is given on the horizontal axis. Since the stimuli were arranged in random order on the audiotape, the experimenter had no control over what stimulus would be presented on a given trial. As a consequence, the number of presentations of the stimuli varied somewhat.

The most obvious feature of Figure 5 is that, as noted previously, many more head turns were observed on change trials as compared to control trials. More specifically, however, infants seemed to turn in roughly equal proportions in response to each of the six sounds in the two categories; that is, they did not show any prominent, consistent preference for a particular talker or place-of-articulation value. This was also true for the subgroup of four infants reinforced for head turns in response to the nasal consonants (see Figure 6). Again, the general picture is one of a relatively even distribution of responses among the stimuli. It is especially interesting that the infants did not show a preference for the stimulus used in the initial-training stage, shown at the extreme left of each graph. In fact, Figure 6 shows a slight tendency to avoid the training token, although this effect is not particularly prominent.

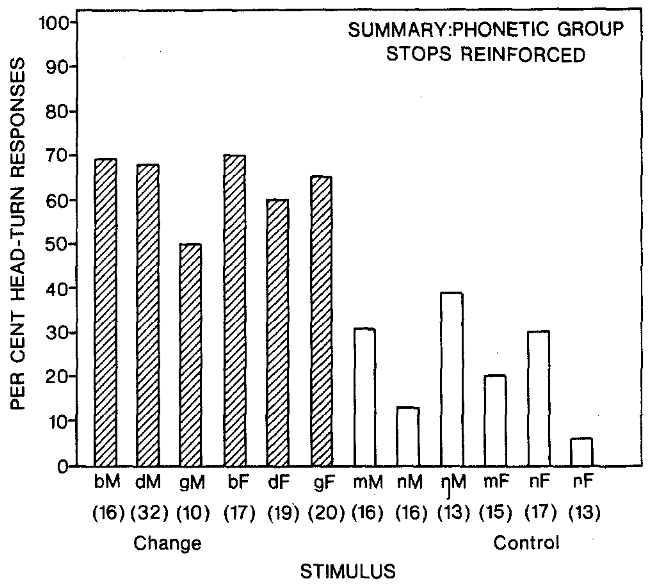


FIGURE 5. Percent head-turn responses to each of the stimuli presented during change trials (shaded columns) and control trials (unshaded columns) for the phonetic subgroup in which the stop category was reinforced (n = 3). The figures in parentheses indicate the number of times each stimulus was presented. M = male voice; F = female voice.

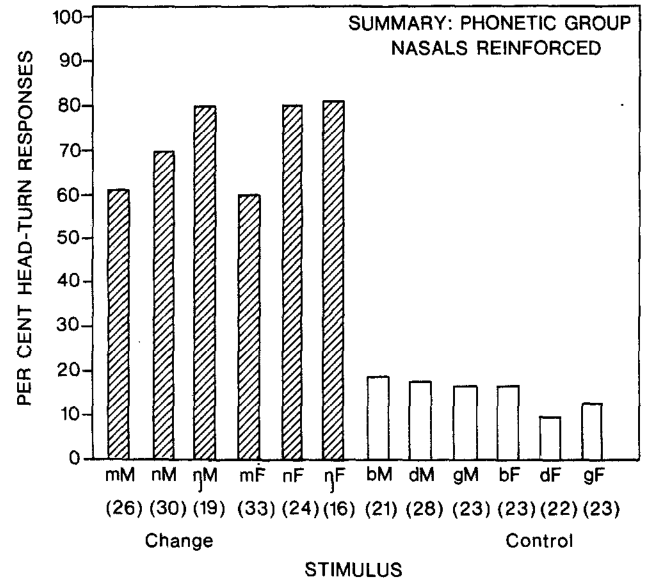


FIGURE 6. Percent head-turn responses to each of the stimuli presented during change trials and control trials for the phonetic subgroup in which the nasal category was reinforced (n = 4).

A clearer picture of these results can be obtained by combining the data for all seven infants in the phonetic group. This can be done by contrasting reinforced versus unreinforced stimuli and collapsing the data into broader categories such as "labial, male," "dental, male," and so on. A graph combining the data from all subjects in the phonetic group is shown in Figure 7. The impression of an even distribution of responding to the stimuli is even

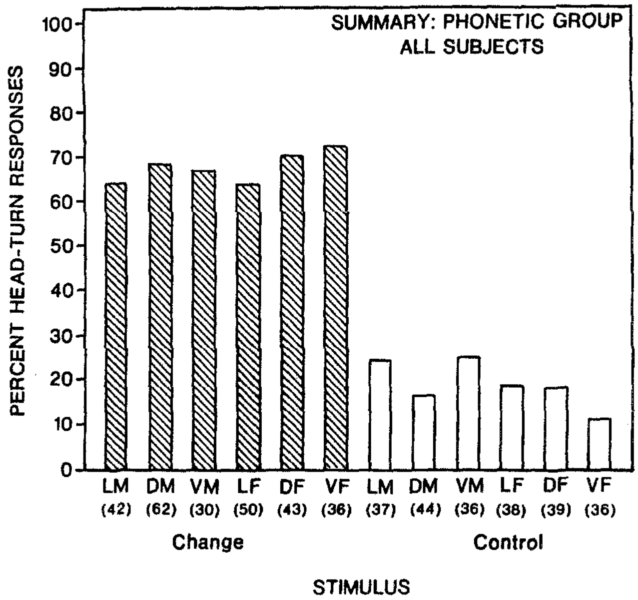


FIGURE 7. Percent head-turn responses to each of the stimuli presented during change trials and control trials for all subjects in the phonetic group ($n = 7$). L = labial; D = postdental; V = velar.

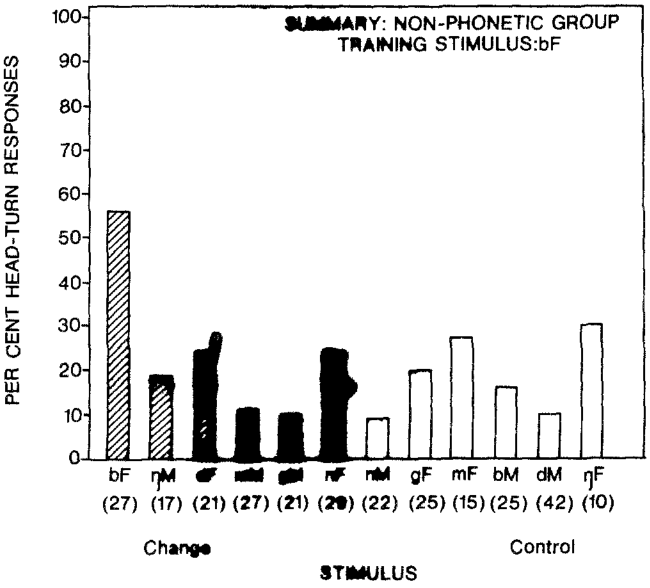


FIGURE 8. Percent head-turn responses to each of the stimuli presented during change trials and control trials for the nonphonetic subgroup in which [ba] (female) served as the training stimulus ($n = 4$).

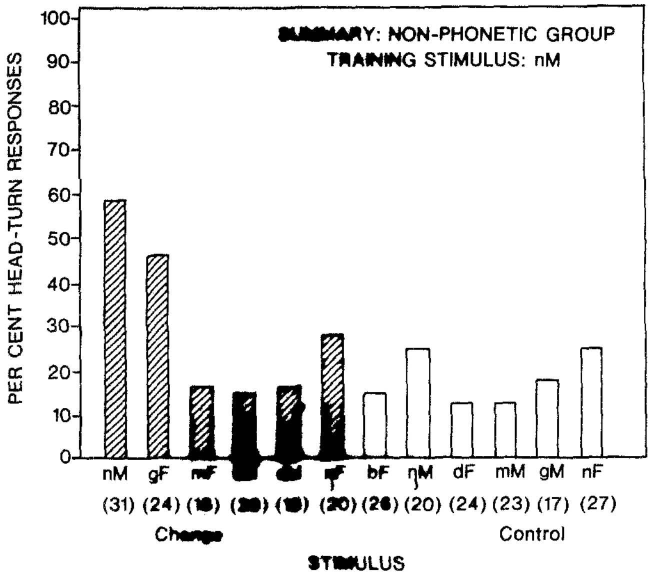


FIGURE 9. Percent head-turn responses to each of the stimuli presented during change trials and control trials for the nonphonetic subgroup in which [na] (male) served as the training stimulus ($n = 4$).

stronger in this graph. The mean response percentage to the reinforced stimuli was 67.5%, with a range of only 8% and a standard deviation of 2.9%. A three-way analysis of variance for talker (male vs. female), place of articulation (labial vs. postdental vs. velar), and trial type (change vs. control) revealed a significant main effect for the trial-type factor only ($F = 13.4$; $df = 1, 6$; $p < .01$). There were no effects for talker ($F = 1.1$; $df = 1, 6$; p NS) or place of articulation ($F = 1.4$; $df = 2, 12$; p NS), and none of the interactions approached significance.

The pattern of responding in the nonphonetic group was quite different from that of the phonetic group. Figure 8 shows the percentage of head turns to each of the stimuli presented to the group of four nonphonetic infants who were reinforced in initial training for head turns to [ba] (female). As a group, these infants tended to turn more often to the six stimuli in the reinforced class than to those in the unreinforced class (25% vs. 16%). But, unlike the pattern observed for the phonetic infants, the responses were distributed very unevenly among the six reinforced stimuli. Specifically, many more responses were cued by the [ba] (female) stimulus, which served as the reinforced token in the initial-training contrast. A very similar pattern can be seen in Figure 9 for the subgroup of four infants trained with the categories reversed, that is, the infants for whom [na] (male) served as the reinforced stimulus in initial training. Again, the infants were responding most often to the stimulus used in the initial-training contrast, with relatively low levels of responding to the other stimuli. As a group, the eight subjects in the nonphonetic condition responded to 29% of the change trials, compared to 19% of the control trials. However, when data are removed from trials on which training stimuli were presented, the rate of responding on change trials is only 18%, almost identical to the response rate on control trials. This suggests that the significant trial-type effect found for this group was due almost exclusively to responses to the training stimulus.

It was not possible to combine the data from the two subgroups in the nonphonetic condition. For the phonetic condition this was accomplished by combining the responses to reinforced stimuli which shared values on all dimensions except the stop/nasal dimension. This perfect symmetry, of course, did not exist for the nonphonetic categories. Consequently, it was not possible to line up each stimulus in one category with a stimulus in the other category that differed on a single feature value.

Profiles of Individual Subjects

The data presented thus far are the results of averages from groups of subjects. Results from the seven individual infants in the phonetic group are presented in Figure 10. Three measures are given to the right of each graph: (a) the percentage of head turns on change trials (CH), (b) the percentage of head turns on control trials (CL), and (c) the overall percent correct on both change and control trials (%C). These graphs should be examined with some caution because of the variation in the number of presentations of the stimuli, given in parentheses on the horizontal axis. Since the experimenter had no control over which stimulus was presented on a given trial, some of the data points in these graphs are based on very few responses. Examination of these data clearly shows that the infants do not form a homogeneous group. Two of the infants, Subjects 3 and 7, appeared to be responding randomly to the stimuli, while the remaining five infants performed with relatively high accuracy.

Figure 11 shows the response patterns of the eight infants tested in the nonphonetic group. Intersubject variability in the performance of these subjects is also evident. Some of the infants, particularly Subjects 1, 2, 4, and 6, apparently found the task very difficult and produced what seemed to be essentially random head-turn responses to the 12 stimuli. Other infants, however, responded with some consistency to the stimulus used in the initial-training contrast, shown at the extreme left of each graph. Subject 8, in fact, appeared to have memorized a second stimulus. It is interesting that this second stimulus ([ga], female) has little in common with the training stimulus ([na], male). On the other hand, Subject 3, who was initially trained to [ga] (female), responded almost exclusively to the tokens produced by the female talker. The pattern shown by this infant is more typical of other infants who have been run using this type of procedure—that is, some attempt by the infant to formulate a general rule to organize the stimulus categories (Kuhl, Holmberg, Morgan, Hillenbrand, & Cameron, Note 3).

Results from Preliminary Stages

The data described to this point were derived from analyses of the infants' responses on the final stage of the experiment. This section provides a brief description of the results from the preliminary stages of the experiment; a more detailed account of these results can be found in Hillenbrand (Note 4). Table 3 shows the results from the first three experimental stages and from the conditioning phase for infants in the phonetic and nonphonetic groups. For the conditioning phase the criterion was three consecutive anticipatory head turns; for the three experimental stages the criterion was nine correct responses in 10 consecutive trials.

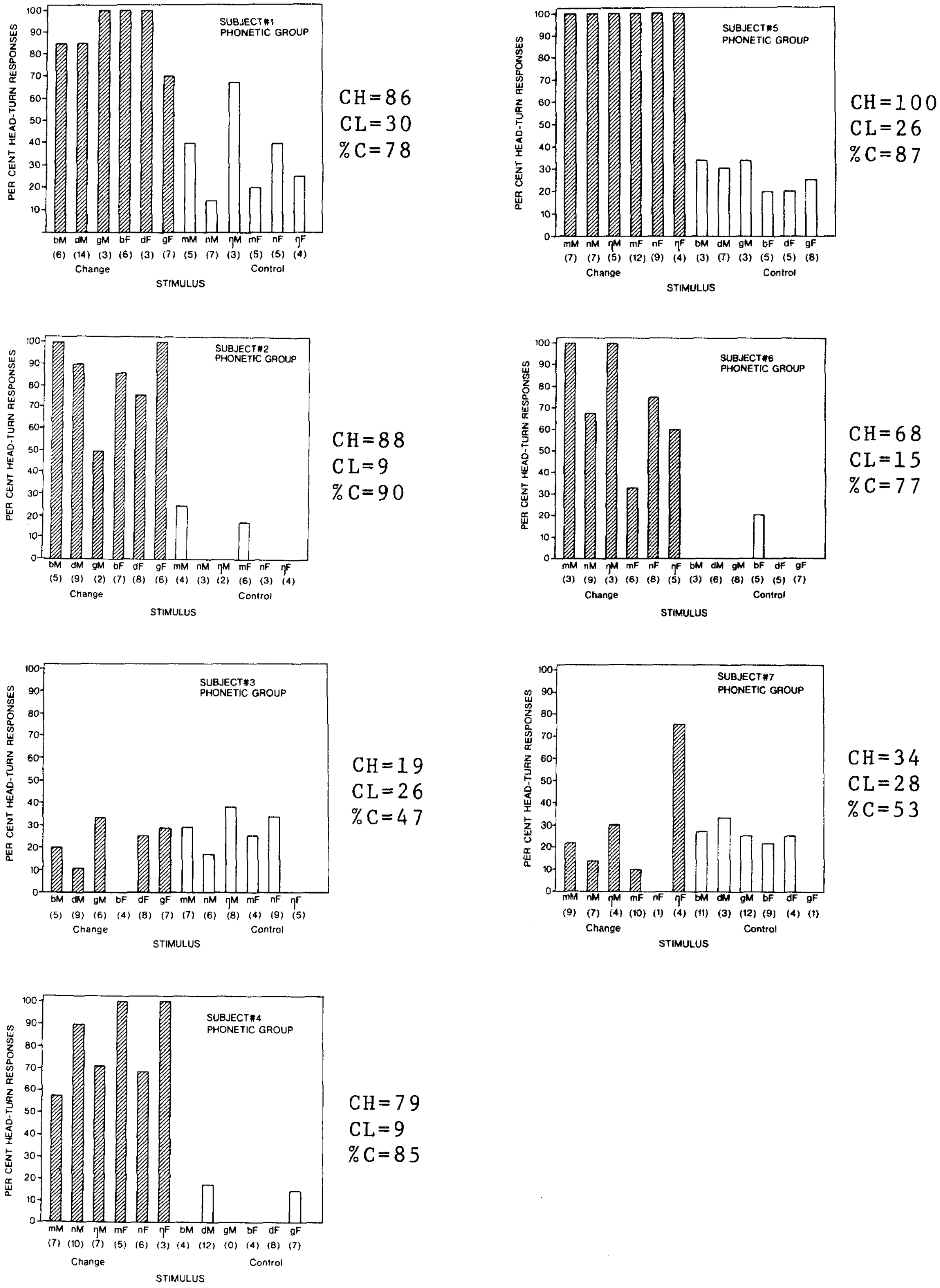
TABLE 3. Number of trials required to reach criterion for subjects in the phonetic and nonphonetic groups.

		Experimental stage		
Subject	*Condition*	*1*	*2*	*3*
Phonetic group				
1	10	20	16	–
2	13	–[a]	–	–
3	11	–	–	–
4	10	–	12	–
5	20	–	15	17
6	21	10	–	14
7	20	–	–	–
Nonphonetic group				
1	8	–	–	–
2	6	10	–	–
3	7	10	–	–
4	9	–	–	–
5	20	–	–	–
6	9	10	–	–
7	5	10	–	–
8	3	–	–	–

[a]Subject failed to meet criterion (indicated by dashes).

One fairly prominent finding from these tables is that, on the average, infants in the phonetic group required more trials to reach the conditioning criterion ($\bar{x} = 15.0$) than did infants in the nonphonetic group ($\bar{x} = 8.7$). This difference was predictable since the nonphonetic infants were trained on a contrast involving differences in several acoustic dimensions, while the phonetic infants were trained on a minimal pair. A second feature of interest in these tables is that in the majority of cases infants did not meet the 9-out-of-10 accuracy criterion and, consequently, were progressed to the next experimental stage after 20 trials. This was true for both groups and for all three stages. This was not particularly surprising since previous work has shown that infants typically require more than 20 trials to reach criterion on consonant contrasts (Holmberg et al., 1977).

A more revealing picture of the infants' performance throughout the experiment can be seen by examining the overall percentage of correct responses as a function of the experimental stage. Mean and standard deviation percent correct for each experimental stage are plotted in Figure 12 for the phonetic group and in Figure 13 for the nonphonetic group. Figure 12 shows that there was no tendency for the performance of the phonetic infants to decline as the experiment became more complex. In fact, these data show a slight trend in the opposite direction. In contrast, the performance of the nonphonetic infants dropped rather sharply from stage 1 to stage 2 and remained at a relatively low level. These results suggest that the nonphonetic infants were able to learn the head-turn task but were unable to memorize the unrelated tokens that were added as the experiment progressed.



CH=86
CL=30
%C=78

CH=100
CL=26
%C=87

CH=88
CL=9
%C=90

CH=68
CL=15
%C=77

CH=19
CL=26
%C=47

CH=34
CL=28
%C=53

CH=79
CL=9
%C=85

FIGURE 10. Individual response profiles for subjects in the phonetic group. The figures to the right of each graph indicate the percentage of responses on change trials (CH), the percentage of responses on control trials (CL), and the overall percent correct (%C).

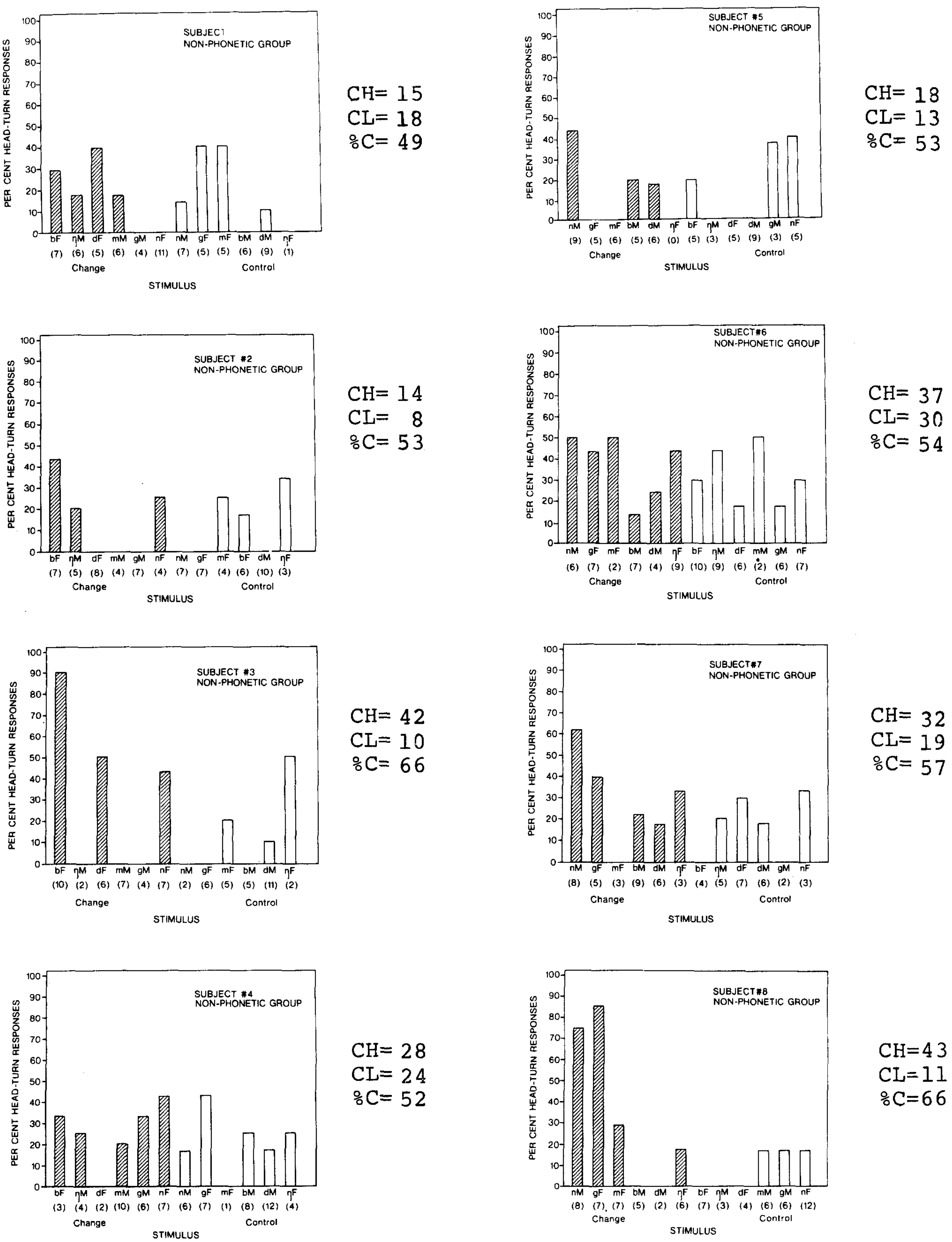


FIGURE 11. Individual response profiles for subjects in the nonphonetic group. The figures to the right of each graph indicate the percentage of responses on change trials (CH), the percentage of responses on control trials (CL), and the overall percent correct (%C).

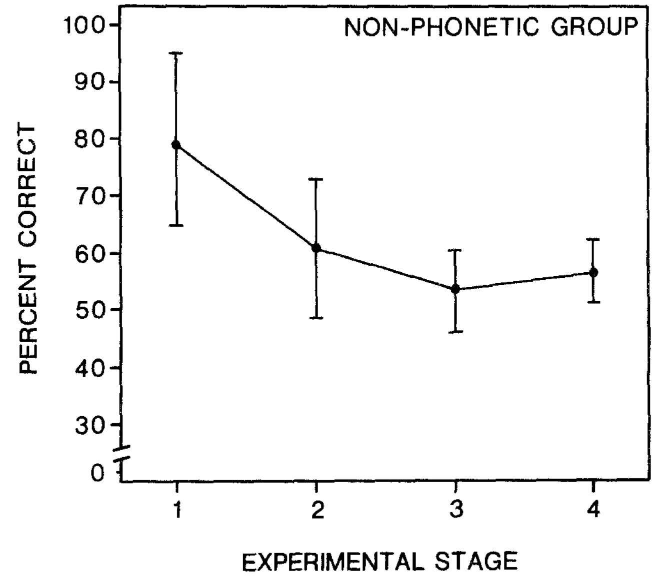


FIGURE 12. Overall percent correct for each experimental stage for the phonetic group. The error bars indicate one standard deviation.

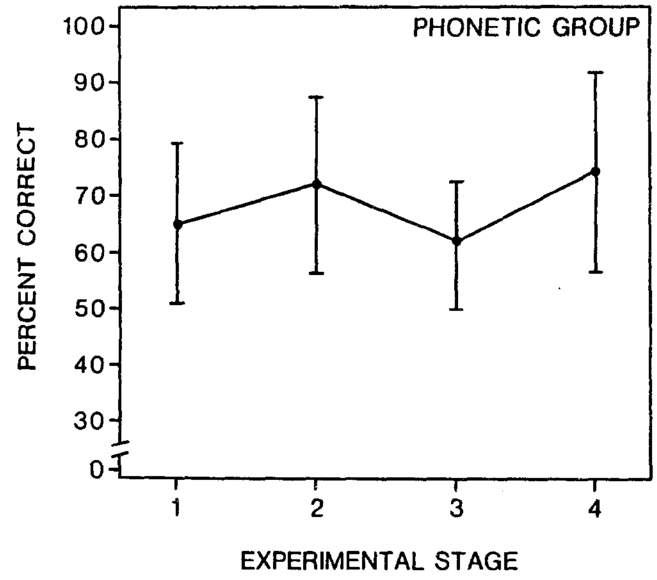


FIGURE 13. Overall percent correct for each experimental stage for the nonphonetic group. The error bars indicate one standard deviation.

An Additional Control Condition

As was discussed previously, the nonphonetic condition was designed to test infants on a set of stimuli comparable to that used in the phonetic condition but which could not be grouped on the basis of auditory similarity. The relatively good performance of infants in the phonetic group led to the conclusion that these subjects recognized similarities among sounds in the stimulus categories. However, infants in the phonetic group were initially trained on a more difficult contrast than were infants in the nonphonetic group. As a consequence, more infants in the phonetic group failed to meet the conditioning criterion. For this reason, it could be argued that the phonetic/nonphonetic difference was the result of bias in subject selection. It is possible that the more difficult initial-training contrast in the phonetic condition resulted in the selection of better subjects than those in the nonphonetic condition.

To test for this possibility, an additional control condition was run using a nonphonetic task in which the initial-training contrast was the same as that for the phonetic group—[ma] versus [ba]. The experimental stages for this condition are shown in Table 4. As in the phonetic condition, the initial-training stage contrasted [ma] (male) with [ba] (male). However, as in the nonphonetic condition described previously, stimuli were added in subsequent stages in such a way that the categories could not be organized by talker or by place or manner of production. Testing procedures were identical to those described previously except that the tape deck and modular programming logic were replaced by a digital computer (DEC PDP 11/34). A computer program presented stimuli and controlled experimental contingencies according to the same rules and with the same timing parameters as were used to design the programming logic described previously. Six 5½- to 6½-month-old infants began testing; two of these subjects failed to pass the conditioning criterion.

TABLE 4. Experimental stages for an additional nonphonetic control condition.

Stage	*Category 1*	*Category 2*
1	ba (M)	ma (M)
2	ba (M)	ma (M)
	ga (F)	ŋa (M)
3	ba (M)	ma (M)
	ga (F)	ŋa (M)
	ma (F)	da (F)
	na (M)	ba (F)
4	ba (M)	ma (M)
	ga (F)	ŋa (M)
	ma (F)	da (F)
	na (M)	ba (F)
	da (M)	ga (M)
	ŋa (F)	na (F)

The results of this control experiment do not support the possibility that the phonetic/nonphonetic difference was due exclusively to bias in subject selection. Average performance on the initial-training stage was 68% correct, comparable to that of the phonetic group. However, unlike the performance of the phonetic group, these subjects' performance fell very close to chance and stayed there for the remaining stages. Average performance for the final stage was 58% correct. These findings support the conclusion that infants in the phonetic condition performed well because they recognized the perceptual similarity of syllables sharing a value on a feature dimension.

DISCUSSION

The principal findings of this study were:

1. The overall performance of infants in the phonetic group was significantly better than that of the nonphonetic group.
2. The phonetic infants tended to distribute their responses more or less evenly among the stimuli in the reinforced category, while infants in the nonphonetic group tended to favor the stimulus that was used in the initial-training contrast.
3. There was no evidence of a systematic decline in the performance of phonetic infants as the experiment became more complex, whereas the performance of infants in the nonphonetic group tended to drop as tokens were added to the two categories.

These results suggest that infants do recognize the similarity of speech sounds that share a value on a phonetic-feature dimension. The alternate possibility that simple rote memorization was responsible for these results seems unlikely in light of the relatively poor overall performance of infants in the nonphonetic group. This same phonetic/nonphonetic difference was also found in a similar study examining categorization of fricatives (Kuhl et al., Note 3) and in a study examining categorization of nasal consonants (Hillenbrand, Note 2). It is important to point out, however, that the nonphonetic results do not prove that memorization was not involved in any form in the phonetic condition. It is a well-established finding that memorization is most efficient when the items to be recalled can be organized in some fashion (e.g., see Bartlett, 1932; Bransford & Franks, 1974; Tulving & Donaldson, 1972). The phonetic/nonphonetic effect suggests that if memorization was involved, the process was aided by the perceptual similarity of the speech sounds. Whatever the exact role of memory in these experiments, it appears that recognition of perceptual similarity is a necessary condition for good performance on this kind of task.

One additional issue that needs to be addressed in interpreting these findings concerns the discriminability of tokens within the stop and nasal categories. To qualify as categorization, it must be demonstrated that the tokens in the particular class are being treated as equivalent but *different*. That is, it would not be interesting to demonstrate common responses to the class [b, d, g] if infants could not discriminate stop-consonant place of articulation. The literature provides ample evidence that infants can discriminate among voiced stop consonants (Eimas, 1974; Morse, 1972). In addition, a recent experiment using procedures very similar to those described in this report provides evidence for the discrimination of nasal-consonant place of articulation by young infants (Hillenbrand, Note 2). These discrimination results suggest that infants in the present study demonstrated what Bornstein (1981) has called "equivalence classification," or "the equivalent treatment of discriminably different stimuli based on their perceptual similarity" (p. 40).

Perceptual Development and Theories of Speech Perception

The present results extend the findings of previous research on infants in which speech-sound categorization was tested at the level of the phonetic segment (Fodor et al., 1975; Holmberg et al., 1977; Kuhl, 1977; 1979b; Kuhl & Miller, 1982; Hillenbrand, Note 2). Taken as a group, these studies suggest that young infants have relatively sophisticated abilities to focus on the critical acoustic dimensions that "define" speech-sound categories while ignoring prominent variation in noncritical dimensions. These findings are analogous to the more extensive developmental literature on perceptual constancies in vision. The work of Bower (1964), for example, suggests that young infants perceive the true size of an object despite the substantial variations in retinal-image size that result when object-observer distance is changed.

The exact role of experience is not clear in these vision experiments, nor is it a simple issue in relation to the infant studies on speech-sound categories. Since the subjects in these studies were not newborns, it is not possible to rule out learning or simply the effects of exposure to speech in accounting for these results. Two conclusions seem reasonable, however. First, if these abilities are learned, they are learned very quickly and apparently without any specific training. Second, and perhaps more important than the specific question of innateness, these abilities predate the acquisition of detailed knowledge of speech production and the acquisition of sophisticated speech-comprehension abilities. This observation bears directly on specific theoretical debates in speech-perception research. An important contention of "motor theories" of speech perception is that the invariance problem is resolved by processes that involve the mediation of articulatory knowledge. The results of the present study, and other demonstrations of perceptual constancy for speech by infants, suggest that sophisticated articulatory knowledge is not a necessary condition for the demonstration of these abilities. It appears that prelinguistic infants are capable of extracting the acoustic properties that form the basis of phonetic categories. If this general finding is corroborated by further research, it would seem to support the auditory-based theories proposed by Fant and others (Fant, 1967; Miller, 1977; Miller et al., 1977; Searle et al., 1979). However, it is possible to formulate a version of an articulation-based theory consistent with the infant findings. It is necessary only to assume that the articulatory knowledge which mediates the perception of speech is phylogenetically rather than ontogenically acquired; that is, that part of human genetic endowment is a species-specific mechanism for speech perception. In fact, this sort of approach has been successful in explaining the perception of biologically relevant signals in other species (Hailman, 1969; Marler, 1970; 1976). However, recent experiments on speech perception by nonhuman listeners are not consistent with this view. Research on the dog (Baru, 1975) and the chinchilla (Burdick & Miller, 1975; Kuhl & Miller, 1975; 1978) suggests that nonhuman listeners are able to sort speech sounds on the basis of phonetic similarity across variations in noncritical dimensions. Taken together, the infant and animal findings suggest that acoustic invariants are available in the speech signal and, further, that the mammalian auditory system seems capable of extracting these properties in a variety of contexts.

Implications for Phonological Development

The phonetic condition contrasted a category of voiced stop consonants with a category of nasal consonants. The performance of subjects in this task indicates that infants are capable of organizing speech sounds on the basis of categories at least this broad or "abstract." The feature categories tested, however, are phonologically organized within even broader feature classes, such as [±continuant] or [±sonorant]. It would be interesting to determine whether infants are capable of organizing speech sounds based on very broad feature categories such as these. For example, would infants reinforced for head turns to nasal consonants also respond to presentations of other sonorants, such as liquids and semivowels, but not to presentations of obstruents, such as fricatives and affricates? The importance of determining the infant's proclivities for classifying speech sounds is that these kinds of perceptual abilities may form the basis for acquiring phonological rules that appeal to feature categories.

There are a number of phonological rules that appeal to the nasal/oral distinction. For example, in most dialects of American English, voiced stops that precede homorganic syllabic nasals are released nasally rather than orally (e.g., "sadden"). Most descriptions of phonological rule systems suggest that rules such as these are specified in terms of values on feature dimensions rather than individual phonetic segments. While the present results do not argue that infants are born with anything that could be described as "phonological knowledge," it is possible that the acquisition of phonological rules may be aided by the infant's recognition of the inherent perceptual similarity of speech sounds sharing particular feature values. On a related issue, some investigators have argued that children do not learn the sound system of their language in a straightforward "segment-by-segment" fashion, but rather by learning the hierarchical organization of features and feature contrasts (Blache, 1978; Jakobson, 1968; Smith, 1973). More detailed studies of the type presented here might reveal a relationship between the acquisition of phonological rules and phonetic segments and the relative difficulty of organizing speech sounds along various feature dimensions.

ACKNOWLEDGMENTS

This work is a portion of a dissertation conducted at the University of Washington's Child Development and Mental Retardation Center under the direction of Patricia Kuhl. Her careful guidance is gratefully acknowledged, as is the advice of Fred Minifie, Wesley Wilson, and Philip Dale. I would also like to thank Jean Tully, Tristan Holmberg, Chris Prall, and Kyum-Ha Lee for their valuable contributions to this project. This work was supported by a research contract from the National Institute of Child Health and Human Development to Dr. Fred Minifie (NICHD HD-3-2793), a grant from the National Science Foundation to Dr. Patricia Kuhl (BNS 79-13767), and by an Annual Fund Doctoral Fellowship to the author from the Graduate School of the University of Washington.

REFERENCE NOTES

1. KUHL, P. K., & HILLENBRAND, J. *Speech perception by young infants: Perceptual constancy for categories based on pitch contour.* Paper presented at the biennial meeting of the Society for Research in Child Development, San Francisco, 1979.
2. HILLENBRAND, J. *Speech perception by infants: Categorization along a nasal consonant place dimension.* Manuscript submitted for publication.
3. KUHL, P. K., HOLMBERG, T. L., MORGAN, K. A., HILLENBRAND, J., & CAMERON, P. *Perception of equivalence for fricatives in CV syllables.* Manuscript in preparation.
4. HILLENBRAND, J. *Perceptual organization of speech sounds by young infants.* Unpublished doctoral dissertation, University of Washington, 1980.
5. PRALL, C. W., & HILLENBRAND, J. *AUDED: A time-domain analysis and editing program for audio signals.* Technical report, Northwestern University, Evanston, IL, 1980.

REFERENCES

BARTLETT, F. C. *Remembering.* Cambridge, England: Cambridge University Press, 1932.

BARU, A. V. Discrimination of synthesized vowels [a] and [i] with varying parameters in dog. In G. Fant & M. A. A. Tathum (Eds.), *Auditory analysis and perception of speech.* New York: Academic Press, 1975.

BLACHE, S. E. *The acquisition of distinctive features.* Baltimore: University Park Press, 1978.

BORNSTEIN, M. H. Two kinds of perceptual organization near the beginning of life. In W. A. Collins (Ed.), *Aspects of the development of competence.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1981.

BOWER, T. G. R. Discrimination of depth in premotor infants. *Psychonomic Science,* 1964, *1*, 368.

BRANSFORD, J. D., & FRANKS, J. J. Memory for syntactic form as a function of semantic context. *Journal of Experimental Psychology,* 1974, *103*, 1037-1039.

BURDICK, C. K., & MILLER, J. D. Speech perception by the chinchilla: Discrimination of sustained /a/ and /i/. *Journal of the Acoustical Society of America,* 1975, *58*, 415-427.

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

EIMAS, P. D., & MILLER, J. L. Discrimination of information for manner of articulation. *Infant Behavior and Development,* 1980, *3*, 367-375.

FANT, G. *Acoustic theory of speech production.* The Hague: Mouton, 1960.

FANT, G. Auditory patterns of speech. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form.* Cambridge: MIT Press, 1967.

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

Fujimura, O. Analysis of nasal consonants. *Journal of the Acoustical Society of America,* 1962, *34,* 1865-1875.

Hailman, J. P. How an instinct is learned. *Scientific American,* 1969, *221,* 98-106.

Hillenbrand, J. Categorization of stop and nasal consonants by young infants. *Journal of the Acoustical Society of America,* 1980, 68(Suppl. 1), S31(A).

Hillenbrand, J., Minifie, F. D., & Edwards, T. J. Tempo of spectrum change as a cue in speech-sound discrimination by infants. *Journal of Speech and Hearing Research,* 1979, *22,* 147-165.

Holmberg, T. L., Morgan, K. A., & Kuhl, P. K. Speech perception in early infancy: Discrimination of fricative consonants. *Journal of the Acoustical Society of America,* 1977, 62(Suppl. 1), S99(A).

Jakobson, R. *Child language, aphasia and phonological universals.* The Hague: Mouton, 1968.

Kuhl, P. K. Speech perception in early infancy: Perceptual constancy for the vowel categories /a/ and /ɔ/. *Journal of the Acoustical Society of America,* 1977, 62(Suppl. 1), S39(A).

Kuhl, P. K. Models and mechanisms in speech perception: Species comparisons provide further contributions. *Brain, Behavior and Evolution,* 1979, *16,* 374-408. (a)

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. (b)

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

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

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

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

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review,* 1967, *74,* 431-461.

Marler, P. A comparative approach to vocal learning: Song development in white-crowned sparrows. *Psychological Monographs,* 1970, *71,* 1-25.

Marler, P. Sensory templates in species-specific behavior. In J. Fentress (Ed.), *Simpler networks and behavior.* Sunderland: Sinauer Associates, 1976.

Miller, J. D. Perception of speech by animals: Evidence for speech processing by mammalian auditory systems. In T. H. Bullock (Ed.), *Recognition of complex auditory signals.* Berlin: Abakon Verlagsgesellschaft, 1977.

Miller, J. D., Engebretson, A. M., Spenner, B. F., & Cox, J. R. Preliminary analysis of speech sounds with a digital model of the ear. *Journal of the Acoustical Society of America,* 1977, 62(Suppl. 1), S13(A).

Morse, P. A. The discrimination of speech and non-speech in early infancy. *Journal of Child Psychology,* 1972, *14,* 477-492.

Searle, C. L., Jacobson, J. Z., & Rayment, S. G. Stop consonant discrimination based on human audition. *Journal of the Acoustical Society of America,* 1979, *65,* 799-809.

Smith, N. V. *The acquisition of phonology: A case study.* Cambridge, England: Cambridge University Press, 1973.

Stevens, K. N., & Blumstein, S. E. Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America,* 1978, *64,* 1358-1368.

Stevens, K. N., & House, A. S. Speech perception. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 2). New York: Academic Press, 1972.

Tulving, E., & Donaldson, W. (Eds.). *Organization of memory.* New York: Academic Press, 1972.

Received March 2, 1982
Accepted August 12, 1982

Requests for reprints should be sent to James Hillenbrand, Department of Communicative Disorders, Northwestern University, 2299 Sheridan Road, Evanston, IL 60201.

APPENDIX

Table A shows the results of acoustic measurements on the stop-vowel and nasal-vowel stimuli used in the infant tests. All measurements were made using the program AUDED (Prall & Hillenbrand, Note 5) written for a DEC PDP 11 computer. Fundamental frequency was measured for the vocalic portion of each utterance by displaying successive 100-msec segments of the waveform on a high-resolution graphics terminal (Tektronix 4010) and using a cross-hair cursor to mark the boundaries of each pitch period. For simplicity, the table shows only mean fundamental frequency. All utterances showed rise/fall fundamental frequency contours. Intensity was measured by a program that simply calculated an RMS value over all data points in the waveform and converted the value to a decibel scale. All values in the table are given in relation to [ba] (male), which was arbitrarily set to 65 dB. The overall duration of each utterance was measured from the same graphics displays as those used to calculate fundamental frequency.

TABLE A. Fundamental frequency, intensity, and duration measurements of the stop and nasal stimuli. Fundamental frequency means and standard deviations are given separately for the male and female talkers.

Stimuli	*Fundamental frequency (Hz)*	*RMS intensity (dB)*	*Duration (msec)*
ba (male)	80.7	65.0	479.2
da (male)	81.5	64.5	562.8
ga (male)	82.1	64.3	518.6
ba (female)	197.6	69.1	407.2
da (female)	194.8	69.5	474.4
ga (female)	197.4	68.4	560.4
mean	81.4/196.6	66.8	500.4
SD	.7/1.6	2.4	59.4
ma (male)	83.9	64.8	505.6
na (male)	84.0	64.3	561.9
ŋa (male)	81.3	64.4	535.2
ma (female)	188.0	71.2	487.7
na (female)	197.0	71.9	508.2
ŋa (female)	192.4	71.5	484.9
mean	83.1/192.4	68.0	513.9
SD	1.5/4.5	3.9	29.6