Speech perception by infants: Categorization based on nasal consonant place of articulation

James Hillenbrand
Northwestern University, Department of Communicative Disorders, 2299 Sheridan Road, Evanston, Illinois 60201

(Received 27 September 1983; accepted for publication 6 December 1983)

This study examined the ability of six-month-old infants to recognize the perceptual similarity of syllables sharing a phonetic segment when variations were introduced in phonetic environment and talker. Infants in a "phonetic" group were visually reinforced for head turns when a change occurred from a background category of labial nasals to a comparison category of alveolar nasals. The infants were initially trained on a [ma]–[na] contrast produced by a male talker. Novel tokens differing in vowel environment and talker were introduced over several stages of increasing complexity. In the most complex stage infants were required to make a head turn when a change occurred from [ma,mi,mu] to [na,ni,nu], with the tokens in each category produced by both male and female talkers. A "nonphonetic" control group was tested using the same pool of stimuli as the phonetic condition. The only difference was that the stimuli in the background and comparison categories were chosen in such a way that the sounds could not be organized by acoustic or phonetic characteristics. Infants in the phonetic group transferred training to novel tokens produced by different talkers and in different vowel contexts. However, infants in the nonphonetic control group had difficulty learning the phonetically unrelated tokens that were introduced as the experiment progressed. These findings suggest that infants recognize the similarity of nasal consonants sharing place of articulation independent of variation in talker and vowel context.

PACS numbers: 43.70.Dn, 43.70.Ve

INTRODUCTION

Central to the study of any perceptual system is the search for physical properties that define or "cue" perceptual categories. A familiar puzzle confronting scientists working in this area has been called the "invariance" or "perceptual constancy" problem. In general terms, perceptual constancy refers to circumstances in which perceptual responses remain stable across variations in important stimulus dimensions. For example, an important problem in vision research is to explain how the perceived size of an object remains constant as object–observer distance, and therefore retinal size, is varied. Since perception experiments with infants can often help to define the role of learning in perceptual constancies, developmental evidence has played a prominent role in the vision literature.

The speech perception literature provides numerous examples of perceptual constancy problems. The physical properties associated with phonetic categories often show sizeable variation with changes in phonetic environment, talker, syllable position, and nonsegmental dimensions such as speaking rate and stress. In spite of this, listeners appear to have little difficulty recognizing speech sounds produced by a variety of talkers in a variety of different environments. (For general reviews, see Liberman *et al.*, 1967; Liberman, 1970; Shankweiler *et al.*, 1977.)

I. PERCEPTUAL CONSTANCY FOR SPEECH-SOUND CATEGORIES BY INFANTS

A recent series of experiments by Kuhl and her colleagues constitute most of the developmental evidence on perceptual constancy for speech-sound categories (Kuhl and Miller, 1982; Kuhl, 1977, 1979; Holmberg *et al.*, 1977; Kuhl and Hillenbrand, 1979; Hillenbrand, 1983). An experiment by Kuhl (1979) used an operant head-turn procedure to test whether six-month-old infants could detect a change from one category of vowels to another when random variations were introduced in the talker and pitch contour of the vowels. Infants were initially trained to make a head turn for visual reinforcement when a change occurred from repetitions of a single token of [a] to repetitions of a single token of [i]. Both tokens were synthesized to simulate a male voice and both had rise–fall pitch contours. Novel tokens differing in talker and pitch contour were gradually introduced over several stages of increasing complexity. In the final stage of the experiment, the infants successfully responded to changes from the [a] category to the [i] category when the tokens in both classes randomly varied in talker between male, female, and child voices and in pitch contour between rising and rise–fall contours. This experiment suggests that infants can ignore acoustically prominent variation in nonrelevant dimensions and categorize speech sounds on the basis of vowel color. Other experiments using this kind of procedure have found similar evidence for an [a]–[ɔ] contrast across variations in talker and pitch contour (Kuhl, 1977), several fricative contrasts across variations in vowel environment and talker (Holmberg *et al.*, 1977; Kuhl *et al.*, in preparation) and a stop-nasal feature contrast involving variations in place of articulation and talker (Hillenbrand, 1983).

In addition to these experiments, a study by Fodor *et al.* (1975) used a different set of techniques to test 3 1/2- to 4 1/2-month-old infants on their recognition of the similarity of

0001-4966/84/051613-10$00.80

TABLE I. Experimental stages for the phonetic condition. An "M" indicates a male voice and an "F" indicates a female voice.

	Category 1	Category 2
Stage 1	[ma] (M)	[na] (M)
Stage 2	[ma] (M) [mi] (M) [mu] (M)	[na] (M) [ni] (M) [nu] (M)
Stage 3	[ma] (M) [ma] (F)	[na] (M) [na] (F)
Stage 4	[ma] (M) [mi] (M) [mu] (M) [ma] (F) [mi] (F) [mu] (F)	[na] (M) [ni] (M) [nu] (M) [na] (F) [ni] (F) [nu] (F)

voiceless stop consonants across variation in vowel context. On any trial the infants were presented with one syllable from a three-syllable set such as [pi,pu,ka]. In all of the triads, two of the syllables shared a phone. Infants in a "same phones" group were visually reinforced for head turns following either of the two syllables that shared a phone, [pi] and [pu] in this example. Head turns following [ka] were not reinforced. Infants in a "different phones" group were reinforced for head turns to two sounds that did not share a phone; for example, [pu] and [ka]. Although the overall performance of infants in both groups was relatively poor, the results showed that infants in the "same phones" group performed better than infants in the "different phones" group.

The Fodor *et al.* (1975) results suggest that, to some degree, infants recognize the similarity of syllables such as [pi] and [pu] regardless of the differences in vowel environment. In the experiments reported here an operant head-turn paradigm was used to test whether six-month-old infants recognized the similarity of nasal consonants sharing a place-of-articulation value in spite of random variation in vowel environment and talker. The nasal-consonant place contrast was of interest partly because of the bearing of this evidence on a related study examining a stop/nasal contrast (Hillenbrand, 1983) and partly because of the difficulty that has been encountered in specifying a set of context-invariant cues to this distinction.

II. PHYSICAL CORRELATES OF NASAL CONSONANT PLACE OF ARTICULATION

Acoustic correlates of nasal-consonant place of articulation can be found in the nasal murmur, produced while the mouth is occluded, and in the spectrum change that occurs as the articulators move toward positions appropriate for the next segment. The spectrum of the nasal murmur is characterized by a relatively low-frequency first formant (200–300 Hz), other diffuse resonances in the region of 1000–2000 Hz, and an antiformant that varies in frequency with place of articulation (Fujimura, 1962; Fant, 1960). Of particular interest to this study, the spectrum of the nasal murmur shows considerable variability with changes in talker and phonetic environment. For example, Fujimura (1962) reported that,

> "... the spectra of nasal murmurs may vary considerably from one sample to the next depending on the individual nasal consonant and its context; the spectra also depend on the individual speaker who utters the sound, or even his temporary physiological state. The spectrum envelope can be altered significantly by a slight modification of the pole-zero pattern. The variability of relative levels, even within one nasal murmur presumably causes an inherent difficulty for ... a recognition scheme that is based on a straightforward analysis." (p. 1874; see also Nord, 1976, for similar comments).

The problem of contextual variability is also quite evident when the spectrum change following articulatory release is examined. The formant transitions for nasal-vowel syllables are quite similar to those of homorganic stop-vowel syllables. For that reason, the well-documented effects of vowel context on the formant transitions of stop consonants are seen in nasal consonants as well (e.g., Potter *et al.*, 1947; Liberman *et al.*, 1954). The basic problem is that the spectrum changes associated with articulatory movement from consonant to vowel are influenced in part by the location of the consonantal stricture and in part by the nature of the adjacent vowel. A recent approach to this problem, proposed by Stevens and Blumstein (1978; Blumstein and Stevens, 1979, 1980) suggests that the gross shape of the short-term spectrum at consonant release provides information about place of articulation that is independent of vowel environment and talker. While the spectrum-shape templates proposed by Stevens and Blumstein were reasonably successful in classifying stop consonants into the appropriate place category, a preliminary study of [m] and [n] showed that the templates were much less successful in classifying nasal consonants (Blumstein and Stevens, 1979).

III. DESIGN

The purpose of the present study was to determine whether prelinguistic infants could learn to produce a head-turn response when a change occurred from a category of labial nasals to a category of alveolar nasals (or from alveolars to labials) in the presence of random variations in vowel context and talker. Infants were tested in several stages of increasing complexity using an operant head-turn paradigm similar to that used in the series of experiments by Kuhl and her associates. The procedure, which will be described in detail below, involved the use of a visual reward to train an infant to make a head-turn response when a change occurred from a category of repeating background stimuli (e.g., [m]) to repetitions from a comparison category (e.g., [n]). The experimental stages for what was called the "phonetic" condition are shown in Table I. The first or "initial training" stage contrasted a single token of [m] with a single token of [n]. Both tokens were naturally produced by the same male talker in the environment of the vowel [a]. In the next stage, there was still a single talker, but the vowel environment was varied randomly between [a], [i], and [u]. This meant that the infant was required to make a head turn when a change occurred from [m] to [n] independent of random variation in vowel context. In the third stage, the vowel context was held constant at [a], but tokens produced by a female talker were added to each category. The final stage simply combined vowel and talker variation: male and female [ma,mi,mu] were contrasted with male and female [na,ni,nu]. Half of the infants were trained with [m] as the comparison category and half were trained with [n] as the comparison category.

The basic idea behind the multiple-token procedure is that good performance on the more complex stages suggests that infants can ignore variation in the nontarget dimensions—vowel environment and talker in this case—and organize the speech sounds on the basis of the target dimension, place of articulation. This, in turn, would suggest that infants recognize the similarity of labial or alveolar nasals independent of differences in vowel environment and talker. However, a more theoretically neutral explanation is possible. The basic problem is that there is nothing to prevent the infant from simply memorizing which individual tokens are reinforced and which ones are not. Memorization of tokens, of course, would not necessarily require any recognition of perceptual similarity by the infant. To test for the possibility that infants might simply memorize individual stimuli, a separate group of infants was run in a "nonphonetic" condition. Infants in the nonphonetic group were tested using the same procedures and the same pool of 12 stimuli that were used in the phonetic condition. The only difference was that the categories could not be organized by phonetic or acoustic characteristics. The experimental stages for the nonphonetic condition are shown in Table II. The initial-training stage contrasted a [ma] produced by a male voice with a [nu] produced by a female voice. The subsequent stages of this condition were analogous to those of the phonetic condition in terms of the number of tokens that were added in each stage. However, the stimuli were added in such a way that, by the final stage, the two categories could not be organized in any obvious way. In the final stage, each category contained an equal number of labials and alveolars, an equal number of male and female talkers, and an equal number of stimuli in each of the three vowel environments. It was reasoned that infants could succeed in the nonphonetic task only by memorizing individual tokens. If infants in the phonetic group performed better than infants in the nonphonetic group, the difference could be attributed to the recognition of phonetic similarity by infants in the phonetic group.

TABLE II. Experimental stages for the nonphonetic condition. An "M" indicates a male voice and an "F" indicates a female voice.

	Category 1	Category 2
Stage 1	[ma] (M)	[nu] (F)
Stage 2	[ma] (M)	[nu] (F)
	[ni] (F)	[mi] (M)
	[na] (F)	[na] (M)
Stage 3	[ma] (M)	[nu] (F)
	[mi] (F)	[ma] (F)
Stage 4	[ma] (M)	[nu] (F)
	[ni] (F)	[mi] (M)
	[na] (F)	[na] (M)
	[mi] (F)	[ma] (F)
	[mu] (M)	[mu] (F)
	[nu] (M)	[ni] (M)

IV. METHOD

A. Stimuli

The stimuli were natural tokens of [ma,mi,mu,na,ni,nu] produced by one adult male and one adult female. Several tokens of each syllable were recorded on audio tape and then digitized at a 20-kHz sample frequency. Computer-derived measurements were used to select a set of stimuli in which the [m] and [n] categories did not differ systematically in fundamental frequency contour, intensity contour, overall rms intensity, or duration. Formal listening tests showed that all stimuli were reliably identified (90% or better) by a panel of five adult listeners.

Audio tapes for discrimination testing were prepared by recording stimuli from the two categories on separate channels of a four-channel tape deck (TEAC 3340-S) with a constant 1.7-s onset-to-onset interstimulus interval. The onsets of the stimuli on the two channels were synchronized using a cueing procedure described by Hillenbrand *et al.* (1979). Gain settings at the input to the tape deck were adjusted so that the two stimuli that contrasted in the initial-training stage were balanced for loudness. Signal calibration procedures and additional technical details about stimulus preparation are available in Hillenbrand (1983).

B. Procedures

1. General

The infant was held on the parent's lap facing an assistant. An experimenter, housed 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 infant. In front of the speaker was an electrically operated stuffed toy bear housed in a smoked Plexiglas box. When activated, the box was illuminated and the bear tapped on a drum. 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 s. 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" (quiet and attending to the toys) he pressed a button that signaled the experimenter to initiate a 5-s observation interval. Two kinds of trials could occur during the interval: change trials and control trials. 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-s observation interval to the assistant; a small light mounted on the video monitor signaled the start of the interval to the experimenter. If both the experimenter and assistant judged that a head turn occurred during the observation interval, they independently pressed buttons that activated the visual reinforcer for 3 s. And-gate circuitry ensured that the reinforcer would be activated only on change trials in which both judges voted during the 5-s observation interval. During a control interval, the infant continued to hear stimuli from the background category. On control trials, both the experimenter and assistant made a judgment on the presence of a head turn, but reinforcement was not provided, regardless of the infant's response. For the final stage of testing in both conditions (stage 4) stimuli were presented using a special three-repetition trial structure described by Kuhl (1979). In stage 4 each stimulus in the order was repeated three times before cycling to the next token. Since each trial consisted of the repetition of a single token, 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 the judgment of 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 audiomonitor in the control room and therefore could have been biased in his judgment of head turns. Experimenter bias in this task would have been revealed by failure of the experimenter 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 that the infant would make errors, the experimenter was instructed to override the probability generator for a single trial after three consecutive change or control trials.

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 which 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 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

Subjects were given a maximum of 20 trials at each of the first three stages. If a subject met an accuracy criterion of nine correct responses in ten consecutive trials he/she was immediately advanced to the next stage of the experiment. With the exception of stage 1 of the nonphonetic condition, very few infants met this nine-out-of-ten criterion.[1] 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 68–75 trials, with an average of 72.3 trials.

4. Retraining

At various stages of testing infants would often 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 presenting conditioning trials—change trials in which the visual reinforcer was manually activated if the infant did not respond within about 4 s 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 at the end of 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 on, 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 min, with an average of 20 trials per session. Seven or eight sessions were generally required to complete the experiment.

C. Subjects

The subjects were normal 5 1/2- to 6 1/2-month-old infants obtained by mail solicitation to parents in the Seattle area. A parent questionnaire was used to screen out infants who: (1) had been treated for middle-ear problems, (2) had a family history of congenital hearing loss, or (3) were born more than two weeks premature or two weeks post-term. Subjects were randomly assigned to either the phonetic or 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. Five subjects failed to pass the conditioning criterion on the [ma]-[na] contrast for the phonetic condition. The nonphonetic task offered subjects a much grosser, multidimensional contrast and, consequently, only two subjects in this group failed to pass conditioning in the allotted 25 trials.

V. RESULTS

A. General

The main finding of this experiment was that infants in the phonetic group performed much better than infants in the nonphonetic group on the final stage. Results from the final stage of the phonetic and nonphonetic conditions will be discussed first, followed by a description of results from stages 1–3. Table III shows the percentage of head turns on change and control trials for subjects in the phonetic and nonphonetic groups. These data are from the final stage of each condition. Infants in both groups tended to respond more often on change trials than control trials, but this difference was more pronounced for the phonetic group. It can be seen that the two groups differed primarily in the proportion of responses on change trials.[2] A two-way analysis of variance yielded significant main effects for trial type [$F(1,14) = 24.3$, $p < 0.005$] and group [$F(1,14) = 9.6$, $p < 0.01$]. There was also a significant group X trial-type interaction [$F(1,14) = 7.1$, $p < 0.05$] indicating that the trial-type effect was significantly larger for the phonetic group. *Post hoc* analysis showed that the trial-type effect was highly significant for the phonetic group [$F(1,7) = 20.2$, $p < 0.005$] but fell just short of significance for the nonphonetic group [$F(1,7) = 4.5$, $p < 0.07$].

B. Responses to individual stimuli

Overall, these findings indicate a very strong effect for the grouping of the stimuli: it appears that infants perform well only when the stimuli can be organized along some salient dimension. It was also of interest to determine how infants in the two groups distributed their responses among the individual stimuli in the comparison and background categories. Figure 1 shows these data for the four infants in the phonetic group who were reinforced for responses to the initial consonant [m]. Again, the results are from the final stage only. The six shaded columns to the left show the percentage of head turns to each of the six labial consonants presented on change trials; the six unshaded columns to the right show the same data for the six alveolar consonants presented during control intervals. The stimulus is given on the horizontal axis. The figure shows a fair amount of variability in response rates to the stimuli in the reinforced category. However, the response rates to all of the reinforced stimuli except [mi] (male) were substantially greater than the rate of responding on control trials. The data for the subgroup of four infants who were reinforced for responses to the [n] category are shown in Fig. 2. This graph shows a more even distribution of responding to the stimuli in the two categories. The only features of note are a somewhat lower than average response rate to [ni] (female) in the reinforced category and a somewhat higher than average response rate to [ma] (male) in the unreinforced category.

A three-way analysis of variance was run on the data shown in Figs. 1 and 2 to test for the effects of trial type, talker, and vowel context. The only significant main effect was for trial type [$F(1,7) = 37.2$, $p < 0.005$]. The trial-type X vowel interaction approached but did not quite reach significance [$F(2,14) = 2.2$, p NS]. Examination of the means showed that this was due to the tendency of subjects to respond less often on change trials to stimuli in the context of [i]. Subjects responded on 42% of the change trials when the vowel context was [i], compared to 59% and 58% for [a] and [u], respectively. Although the trial-type X vowel interaction did not reach significance, examination of data from individual subjects suggests that some of the infants had relatively prominent vowel biases while others did not. Figure 3

TABLE III. Percent head-turn responses on change and control trials for infants in the phonetic (N = 8) and nonphonetic (N = 8) groups. The data are from the final stage of testing (stage 4).

	Trial type	
	Change	Control
Phonetic group	53.4	22.1
Nonphonetic group	31.9	22.6

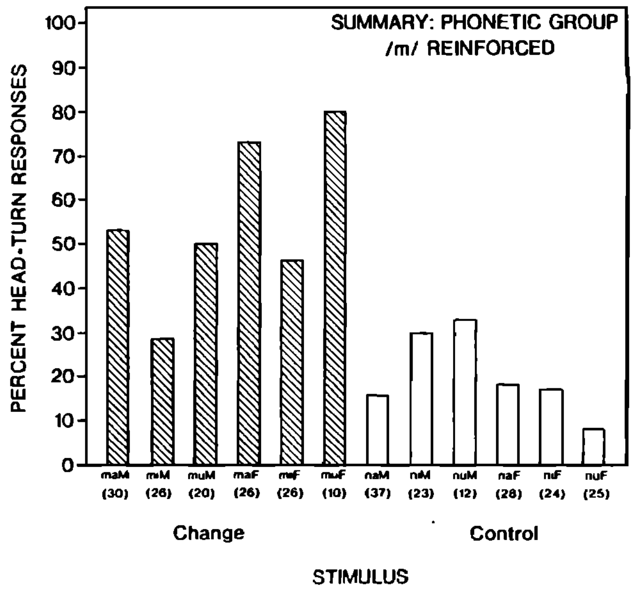

FIG. 1. Percent head-turn responses to each of the stimuli presented during change trials (shaded columns) and control trials (unshaded columns) for the phonetic subgroup in which the [m] category was reinforced ($N = 4$). The figures in parentheses indicate the number of times each stimulus was presented. (Since the stimuli were arranged in random order on the audio tape, the experimenter had no control over which stimulus would be presented on any given trial. As a consequence, the number of presentations of the stimuli varied somewhat.)

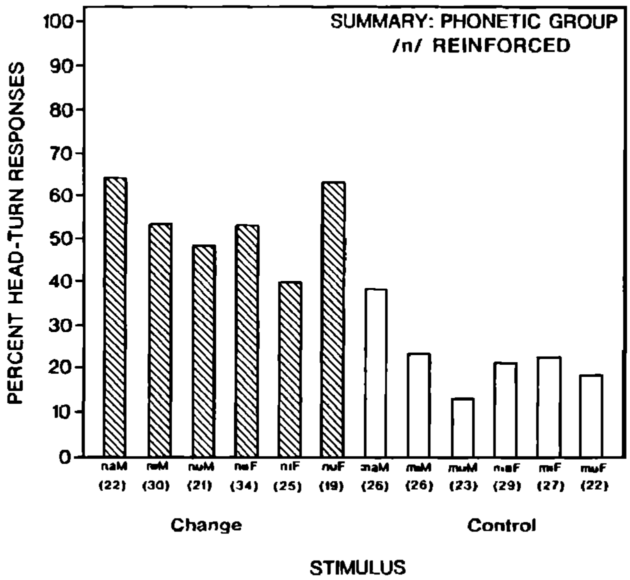

FIG. 2. Percent head-turn responses to each of the stimuli presented during change trials and control trials for the phonetic subgroup in which the [n] category was reinforced ($N = 4$).

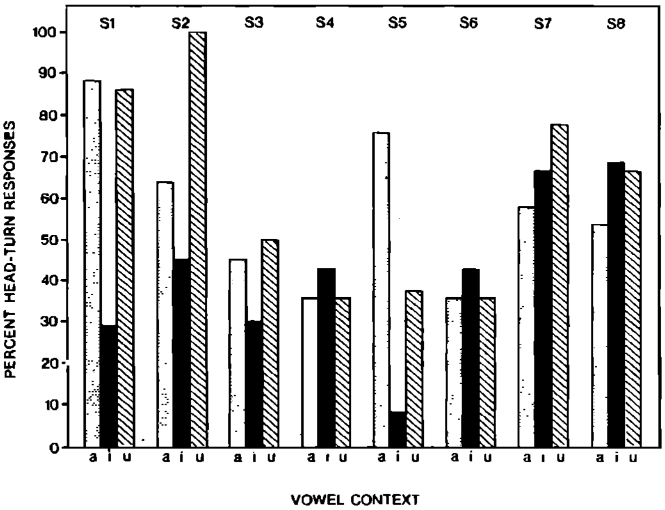

FIG. 3. Percent head-turn responses in each of the three vowel contexts for each subject in the phonetic group.

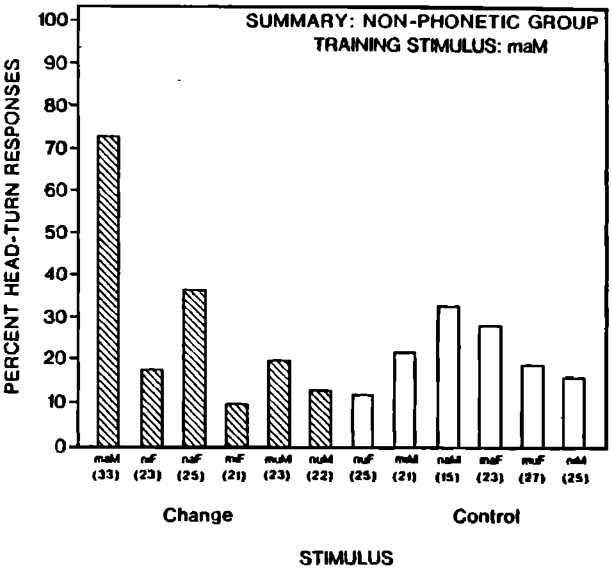

FIG. 4. Percent head-turn responses to each of the stimuli presented during change trials and control trials for the nonphonetic subgroup in which [ma] (male) served as the training stimulus ($N = 4$).

shows the percentage of head turns separately to stimuli in the [a], [i], and [u] contexts for each subject in the phonetic condition. As this figure shows, some of the infants, particularly subjects 1, 2, and 5, tended to respond less often to stimuli in the [i] context. The data from subject 1 are especially interesting. This infant responded to only five of 17 [i] presentations (29%) compared to 21 or 24 presentations (88%) in the [a] and [u] contexts. Strong evidence also comes from subject 5 who responded to only one of 13 stimuli in the [i] context (8%) compared to 16 of 25 presentations in the [a] and [u] contexts (64%). Some of the infants, however, did not appear to show any constraint vowel biases. Subjects 7 and 8, for example, responded to 50% or more of the presentations in each of the three contexts. These results suggest that, while vowel bias did not reach significance summed across all of the subjects, some of the infants showed this effect, while others showed no vowel biases.

The pattern of responding for subjects in the nonphonetic group was quite different from that of the phonetic group. Figure 4 summarizes the data from the four infants in the nonphonetic group who were reinforced for the category of stimuli in which [ma] (male) served as the training token. It is quite apparent that many more responses were cued by the training stimulus (shown at the extreme left) than any other stimulus in the reinforced category. In fact, this is the only stimulus in the change category that seems to emerge from the average response rate on control trials. A similar but less pronounced effect is seen in Fig. 5, which shows the response distribution for the subgroup reinforced for the category in which [nu] (female) was the training stimulus. Again, more responses were cued by the stimulus that was reinforced in the initial-training contrast. The rates of responding to the other stimuli in the change category appear to be roughly equivalent to the response rates to stimuli in the control category.

As was noted previously, the eight subjects in the nonphonetic group responded to 32% of the change trials, compared to 22% of the control trials. However, when data are removed from trials in which training stimuli were presented, the rate of responding on change trials falls to 23%, virtually identical to the 22% response rate on control trials.

C. Summary of results from the final stage

Results from the final stage of testing showed that: (1) the overall performance of infants in the phonetic group was significantly better than that of the nonphonetic group, (2) as a group, infants in the phonetic condition responded consistently to all of the stimuli in the reinforced category, with no statistically significant effects for talker or vowel context, (3) some of the individual subjects in the phonetic group seemed to show fairly strong biases toward turning less often to reinforced stimuli in the [i] context, and (4) subjects in the nonphonetic condition showed a strong preference for the training stimulus, and tended to respond at relatively low levels to the other stimuli. The next section describes the results from the conditioning phase of the experiment and from the first three preliminary stages of testing.

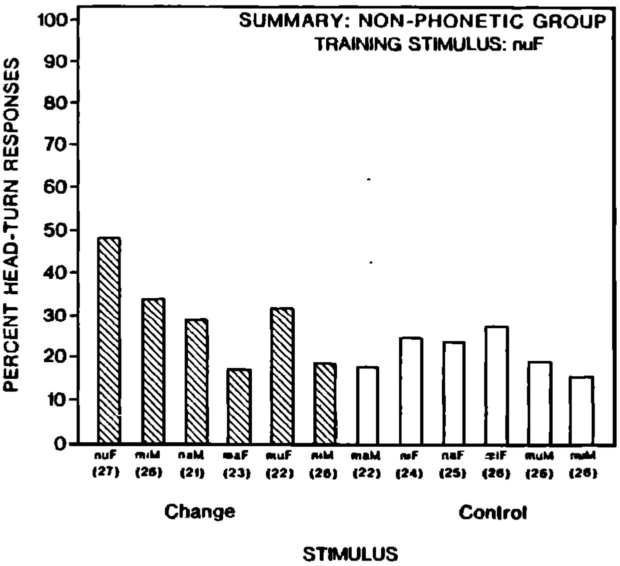

FIG. 5. Percent head-turn responses to each of the stimuli presented during change trials and control trials for the nonphonetic subgroup in which [nu] (female) served as the training stimulus ($N = 4$).

D. Results from preliminary stages

Table IV shows the number of trials required to meet criterion on conditioning and on the first three experimental stages for each infant in the phonetic and nonphonetic groups. An asterisk indicates that the subject failed to reach criterion on that stage. For the conditioning phase the criterion was three consecutive anticipatory head turns; for the three experimental stages the criterion was nine correct responses in ten consecutive trials. The conditioning data show that infants in the phonetic group required more trials to reach criterion than infants in the nonphonetic group (15.4 vs 10.3). This was not surprising since the initial-training contrast for the phonetic condition involved a difference in a single feature dimension and the initial-training contrast for the nonphonetic condition involved differences in three dimensions. The data for the next three stages show that, with the exception of stage 1 in the nonphonetic condition, infants usually did not meet the nine-out-of-ten accuracy criterion in the allotted 20 trials. This is consistent with previous research showing that infants often require more than 20 trials to reach criterion on consonant contrasts (Holmberg *et al.*, 1977; Hillenbrand, 1983).

Given the number of infants who failed to reach criterion, the trials-to-criterion data did not constitute a particularly meaningful measure of performance. A clearer picture of the infants' performance throughout the course of the experiment can be seen by examining overall percent correct for each group at each stage of the experiment. Mean percent correct for each experimental stage is displayed in Fig. 6. (Results from the nonphonetic condition are shown by the curve labeled "Nonphonetic Group 1." The curve labeled "Nonphonetic Group 2" is from an additional nonphonetic control condition that is described in the next section.) It is obvious that the pattern shown by infants in the phonetic group is quite different from that of the nonphonetic group. As was found in a related study (Hillenbrand, 1983) there was no tendency for the performance of subjects in the phonetic condition to decrease as the experiment became more complex. In fact, with the exception of a small drop in performance from stage 3 to stage 4, there was a slight tendency for infants to become more accurate as the experiment progressed. The performance of infants in the nonphonetic group is closely related to the number of stimuli that are being contrasted in each stage. Performance is relatively high in stage 1, which involves one token in each category, and in stage 3, which involves two tokens per category. Performance is much poorer in stage 2 and stage 4 which involve three and six tokens per category. These results suggest that subjects in the nonphonetic group had learned the procedure but were unable to memorize the unrelated tokens that were added as the experiment progressed.

TABLE IV. Trials to criterion for subjects in the phonetic and nonphonetic groups. An asterisk indicates that the subject failed to reach criterion.

Subject	Experimental stage Conditioning	1	2	3
Phonetic group				
1	9	*	*	*
2	17	19	*	*
3	14	*	*	*
4	23	*	*	*
5	6	*	*	*
6	23	*	*	*
7	13	*	*	20
8	18	*	*	*
Nonphonetic group				
1	10	10	*	10
2	9	10	*	*
3	11	19	*	10
4	12	18	*	*
5	13	13	*	*
6	7	10	*	*
7	11	18	*	*
8	9	15	*	*

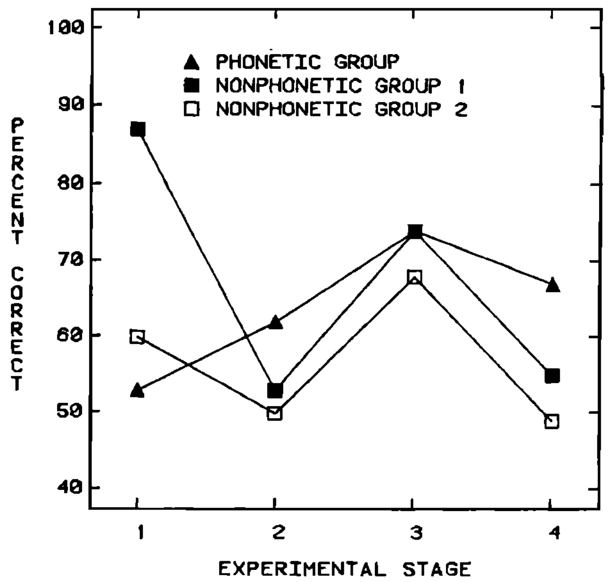

FIG. 6. Overall percent correct for each experimental stage for the phonetic group and for two nonphonetic control groups.

E. Analysis of first trial data

Although this study used a transfer-of-learning paradigm, it is not a simple matter to determine the extent to which training transferred from less complex to more complex stages. For example, the near-chance performance of the phonetic group on initial training suggests that these infants may actually have begun acquiring the [m]-[n] distinction in stage 2. In general, it is difficult to determine whether training actually transferred from one stage to the next, or whether infants may simply have relearned the category assignments each time a new stage was encountered.

One way to study this question is to examine the infants' responses on trials on which novel stimuli were presented for the first time. Assuming that training did not transfer from one stage to the next, infants should perform at chance to the first presentations of novel stimuli. Novel tokens were introduced in stage 2 (miM,niM,muM,nuM), stage 3 (maF,naF), and stage 4 (miF,niF,muF,nuF). Including performance on both change and control trials, infants responded correctly on 51 of the 80 trials in which novel tokens were presented for the first time. This outcome is significantly above chance ($z = 2.6, p < 0.01$), suggesting that training did, in fact, transfer from less complex to more complex stages. However, these results clearly do not show the very high level of generalization that was seen in the first-trial data reported by Kuhl (1979) in a study examining vowel categorization by infants.

F. An additional nonphonetic control condition

The purpose of the phonetic–nonphonetic comparison was to test for the possibility that infants might simply memorize individual tokens rather than categorize them. The relatively poor performance of subjects in the nonphonetic group suggests that infants could not memorize large numbers of unrelated tokens. It appears as though infants in the phonetic group recognized similarities among the various tokens of [m], for example, and used this information in deciding whether or not to turn to particular tokens. However, there is a possibility that the phonetic–nonphonetic difference was due to bias in the selection of subjects. Recall that infants in the phonetic group were initially trained on an [m]-[n] contrast in which both talker and vowel context were held constant. In the nonphonetic condition, the stimuli used in the initial training pair differed in place of articulation, talker, and vowel context ([ma], male versus [nu], female). Since infants in the phonetic group were initially trained on a more difficult contrast, more infants failed to meet the conditioning criterion in the phonetic group as compared to the nonphonetic group. It is possible that the differential subject attrition resulted in the selection of better subjects in the phonetic group. A subject-selection bias of this type could account for all or part of the phonetic–nonphonetic difference.

An additional nonphonetic control condition was designed to test for this possibility. A separate group of infants was run in a nonphonetic condition very similar to the one described previously except that, as in the phonetic condition, the initial training pair contrasted [ma] (male) versus [na] (male). As in the previous nonphonetic condition, tokens were added in subsequent stages in such a way as to produce two arbitrary six-token categories in stage 4. The experimental stages are shown in Table V. Testing procedures were identical to those described previously except that the tape deck and electronic logic were replaced by a laboratory computer equipped with a high-speed disk drive and digital-to-analog converter. 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. Seven 5 1/2- to 6 1/2-month-old infants began testing; three of these subjects failed to pass the conditioning criterion.

TABLE V. Experimental stages for an additional nonphonetic control condition. An "M" indicates a male voice and an "F" indicates a female voice.

	Category 1	Category 2
Stage 1	[ma] (M)	[na] (M)
Stage 2	[ma] (M) [ni] (F) [na] (F)	[na] (M) [mi] (M) [nu] (F)
Stage 3	[ma] (M) [mi] (F)	[na] (M) [ma] (F)
Stage 4	[ma] (M) [ni] (F) [na] (F) [mi] (F) [mu] (M) [nu] (M)	[na] (M) [mi] (M) [nu] (F) [ma] (F) [mu] (F) [ni] (M)

Results from the four subjects who completed the experiment strongly suggest that the phonetic–nonphonetic difference was not the result of differences in the initial-training contrast. Overall percent correct data on the four stages are shown in Fig. 6 by the curve labeled "Nonphonetic Group 2." On the initial training stage, infants in the new nonphonetic condition averaged 60% correct, a little better than the 55% correct for infants in the phonetic condition. For stages 2–4, however, the pattern shown by infants in the new nonphonetic condition looks very similar to that shown by infants in the original nonphonetic group.

The results of the additional nonphonetic control condition strongly suggest that the phonetic-nonphonetic difference was not the result of subject-selection bias, or any other effects that may have resulted from differences in the initial-training contrast. The relatively good performance of subjects in the phonetic group appears to be related to the ability of these infants to categorize phonetically similar stimuli.

VI. DISCUSSION

The very sizeable differences between the phonetic and nonphonetic groups suggest that infants, at some level, are aware of similarities between certain phonetically related speech sounds. As Fodor *et al.* (1975) have pointed out, this kind of interpretation is based on the assumption that infants, like adults, find it easier to learn category assignments when the stimuli within each category have a property in common. Consequently, the performance differences between the phonetic and nonphonetic groups imply that infants are sensitive to this common property. In other words, it appears as though infants, to some degree at least, recognize that the segments [m] and [n] retain their identity in different vowel environments and when the tokens are produced by different talkers.

These findings are interesting from the point of view of the infants' cognitive development since they indicate a rather sophisticated ability to focus on a critical dimension while ignoring acoustically prominent variation in other dimensions. The results are also interesting in light of the variations in acoustic correlates of nasal-consonant place of production that occur as a result of changes in talker and vowel context. Recall that place cues for nasal consonants are associated with the spectrum of the nasal murmur, which is conditioned by both talker and vowel context (Fujimura, 1962; Nord, 1976), and the spectrum immediately following release, which is strongly conditioned by vowel environment (Liberman *et al.*, 1954). Although the infants were initially trained on a male voice, the introduction of a female talker in stage 3 did not cause a drop in performance. Further, analysis of results from the final stage showed no indication of a preference for the male talker who was used in the conditioning task and in the initial-training stage. It appears, then, that infants had either extracted a talker-invariant place cue from the nasal murmur, or that their sorting rules were based on information in the spectrum following release. In this case, results from the different vowel environments become interesting. While the group data from the phonetic condition did not show a significant vowel-context effect, there was evidence that some individual infants performed poorly on nasals in the [i] context. This finding might indicate that some cue that the infants had isolated in the [a] environment did not allow appropriate classification of nasals in the context of [i]. For example, suppose that these infants were sorting the speech sounds on the basis of spectral templates similar to those proposed by Stevens and Blumstein (1978; Blumstein and Stevens, 1979). Blumstein and Stevens found that, for initial stop consonants, both the labial and alveolar templates classified the stimuli with relatively less accuracy in the context of [i]. Although Blumstein and Stevens tested their templates against both stop and nasal consonants, they report the results of statistical analyses as a function of vowel context for the stop consonants only. On the basis of theoretical considerations, however, these authors argue that "... the spectrum sampled at the release of a nasal consonant should show the same properties as that for a stop consonant with the same place of articulation..." (Blumstein and Stevens, 1977, p. 1011). Assuming that the vowel-context effects with stops apply also to nasals, the responses of infants in the present study are consistent with sorting rules based on templates such as those proposed by Stevens and Blumstein.

The present findings are also consistent with an approach to the acquisition of phonetic categories proposed by Stevens (1975). Stevens speculated that infants initially sort speech sounds on the basis of relatively simple acoustic properties that allow correct classification into place-of-articulation categories in most, but not all contexts. Correct classification of speech sounds in other contexts is thought to involve the acquisition of "... secondary, context-dependent cues for features...that are available for use when the primary cues are weak or absent" (Stevens, 1975, p. 325). Stevens further speculated that the acquisition of these secondary cues might be related to the acquisition of articulatory knowledge.[3] A productive line of research might involve testing for developmental changes in these types of categorization abilities and determining whether any effects that may be found are associated with the development of speech-production abilities.

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 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.

---

[1] In previous studies using this technique infants were required to meet criterion before progressing to more complex stages (e.g., Kuhl, 1979; Holmberg *et al.*, 1977). In the present study subjects were advanced automatically after 20 trials simply to reduce problems resulting from subject fatigue. Experience has shown that infants seldom maintain interest in the task beyond seven or eight sessions. It was felt that if subjects were required to meet criterion in stages 1–3, many infants might have lost interest in the visual reinforcer by the time the most critical testing began in stage 4.

[2] The fact that the two groups differed in their performance on change trials but not on control trials may at first seem puzzling. If the infants' false alarms are viewed in the same way as false alarms produced by adult subjects in a signal-detection task, the results might seem to imply that infants in the nonphonetic group had learned the stimuli in the negative category without learning the stimuli in the positive category. However, there are important differences between the operant head-turn paradigm and the typical signal-detection task. Experience with the head-turn task suggests that, once an infant is aware of the presence of the animated toy, he/she will periodically turn in the direction of the reinforcer. Some of these head turns will undoubtedly occur during control intervals. It is important to note that, unlike adults in a signal-detection task, our infants receive neither feedback nor penalty when head turns occur during control intervals. In fact, some of these random head turns will occur during change intervals and will therefore be rewarded. Consequently, we have found it difficult to get average false alarm rates much below 20%–25%. If head turns on control trials are viewed as a source of noise in the data, rather than responses to specific stimuli, there is no *a priori* reason to predict that false alarm rates should be higher for infants in the nonphonetic group.

[3] It should be recalled that some of the infants did not show vowel biases and, further, that the vowel biases which were found may have a simpler explanation than the one offered above. Variation in a nontarget dimension can cause difficulties for the subject in two different ways. As suggested by the previous discussion, the nontarget dimension can result in the transformation of a critical cue that the subject is using to sort the stimuli. However, the nontarget dimension can also serve simply to distract the infant; that is, the introduction of consonants in the [i] context might have been sufficiently distracting that the infants were unable to focus on the critical property that they had isolated. The data from the present study do not make it clear which of these explanations is appropriate.

Blumstein, S. E., and Stevens, K. N. (1979). "Acoustic invariance in speech perception: Evidence from measurements of the spectral characteristics of stop consonants," J. Acoust. Soc. Am. 66, 1001–1017.

Blumstein, S. E., and Stevens, K. N. (1980). "Perceptual invariance and onset spectra for stop consonants," J. Acoust. Soc. Am. 68, 648–662.

Fant, G. (1960). *Acoustic Theory of Speech Production* (Mouton, The Hague).

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

Fujimura, O. (1962). "Analysis of nasal consonants," J. Acoust. Soc. Am. 34, 1865–1875.

Hillenbrand, J. (1983). "Perceptual organization of speech sounds by infants," J. Speech Hear. Res. 26, 268–282.

Hillenbrand, J., Minifie, F. D., and Edwards, T. J. (1979). "Tempo of spectrum change as a cue in speech-sound discrimination by infants," J. Speech Hear. Res. 22, 147–165.

Holmberg, T. L., Morgan, K. A., and Kuhl, P. K. (1977). "Speech perception in early infancy: Discrimination of fricative consonants," J. Acoust. Soc. Am. Suppl. 1 62, S99.

Kuhl, P. K. (1977). "Speech perception in early infancy: Perceptual constancy for the vowel categories /a/ and /ɔ/," J. Acoust. Soc. Am. Suppl. 1 62, S39.

Kuhl, P. K. (1979). "Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories," J. Acoust. Soc. Am. 66, 1668–1679.

Kuhl, P. K., and Hillenbrand, J. (1979). "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.

Kuhl, P. K., Holmberg, T. L., Morgan, K. A., Hillenbrand, J., and Cameron, P. (in preparation). "Perception of equivalence for fricatives in CV syllables."

Kuhl, P. K., and Miller, J. D. (1982). "Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants," Percept. Psychophys. 31, 279–292.

Liberman, A. M. (1970). "The grammars of speech and language," Cogn. Psychol. 1, 301–323.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). "Perception of the speech code," Psychol. Rev. 74, 431–461.

Liberman, A. M., Delattre, P. C., Cooper, F. S., and Gerstman, L. J. (1954). "The role of consonant–vowel transitions in the perception of stop and nasal consonants," Psychol. Mono. 68, 1–13.

Nord, L. (1976). "Experiments with nasal synthesis," Speech Transmission Laboratory Progress and Status Report No. 2-3, 14–19.

Potter, R. K., Kopp, G. S., and Green, H. C. (1947). *Visible Speech* (Van Nostrand, New York).

Shankweiler, D., Strange, W., and Verbrugge, R. (1977). "Speech and the problem of perceptual constancy," in *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, edited by R. Shaw and J. Bransford (Erlbaum, Hillsdale, NJ), pp. 315–346.

Stevens, K. N. (1975). "The potential role of property detectors in the perception of speech," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tathum (Academic, New York), pp. 303–330.

Stevens, K. N., and Blumstein, S. E. (1978). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. 64, 1358–1368.