Challenging the notion of innate phonetic boundaries

Susan Nittrouer
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 27 September 2000; accepted for publication 17 April 2001)

Numerous studies of infants' speech perception abilities have demonstrated that these young listeners have access to acoustic detail in the speech signal. Because these studies have used stimuli that could be described in terms of adult-defined phonetic categories, authors have concluded that infants innately recognize stimuli as members of these categories, as adults do. In fact, the predominant, current view of speech perception holds that infants are born with sensitivities for the universal set of phonetic boundaries, and that those boundaries supported by the ambient language are maintained, while those not supported by the ambient language dissolve. In this study, discrimination abilities of 46 infants and 75 3-year-olds were measured for several phonetic contrasts occurring in their native language, using natural and synthetic speech. The proportion of children who were able to discriminate any given contrast varied across contrasts, and no one contrast was discriminated by anything close to all of the children. While these results did not differ from those reported by others, the interpretation here is that we should reconsider the notion of innate phonetic categories and/or boundaries. Moreover, success rates did not differ for natural and synthetic speech, and so a minor conclusion was that children are not adversely affected by the use of synthetic stimuli in speech experiments. [DOI: 10.1121/1.1379078]

PACS numbers: 43.71.Ft, 43.71.An, 43.71.Pc [CWT]

I. INTRODUCTION

In 1971, Eimas, Siqueland, Jusczyk, and Vigorito reported that one- and four-month-old infants could discriminate between two synthetic stop-vowel syllables that differed along an acoustic dimension associated with the voicing of initial stop consonants. The voicing of initial stops is usually described by voice onset time (VOT), which is the start time of vocal-fold vibration relative to the release of closure (Lisker and Abramson, 1964). Positive values of VOT indicate that vocal-fold vibration started after the release of closure, while negative VOT values indicate that vocal-fold vibration preceded release. The acoustic correlate of VOT manipulated by Eimas *et al.* was $F1$-cutback, which is the time of first formant ($F1$) onset relative to the release of closure. As with English-speaking adults, these infants were found to discriminate between stimuli with VOTs of +20 ms and +40 ms. These VOTs placed the syllables on opposite sides of the phoneme boundary for English /p/ and /b/. When a 20-ms difference between syllables was used that placed both stimuli on the same side of the phoneme boundary, infants failed to discriminate between them, as adults fail to do. From these results, Eimas *et al.* concluded that infants are sensitive to the acoustic dimension that defines adult voicing categories, even before they have experience with language.

That report sparked a great deal of research over the next two decades investigating infants' capacities for speech perception. The collective conclusion of these many studies was that infants approximately nine months of age or younger were able to discriminate virtually all phonetic contrasts presented to them, regardless of whether or not the contrasts were in the infant's native language (e.g., Werker, 1991). This result was demonstrated with natural and synthetic stimuli, across a range of contrasts (e.g., Eilers *et al.*, 1982, 1977; Kuhl, 1979b; Moffitt, 1971; Morse, 1972; Streeter, 1976). Although not investigated as frequently, evidence was also found to support the second of Eimas *et al.*'s results, that infants fail to discriminate within-category acoustic differences. Unlike the between-category experiments, this kind of test can be conducted only with synthetic stimuli. When stimuli differ by the same acoustic distance as a between-category pair, but both fall within the same category, infants fail to discriminate them (e.g., Aslin *et al.*, 1981; Eimas, 1974, 1975).

Those early studies of infant speech perception led to the widely accepted view that infants are born with sensitivities to phonetic boundaries for all languages (i.e., the universal set). Experience listening to a native language during the first year of life, the theory holds, maintains those boundaries supported by the ambient language, and causes those boundaries not supported by the ambient language to dissolve. This view of perceptual development is what Aslin and Pisoni (1980) call a ''universal'' theory. Reviews of the work supporting this theoretical position are numerous (e.g., Eimas *et al.*, 1987; Jusczyk, 1995; Kuhl, 1979a, 1987; Morse, 1985; Werker, 1989). Even if only by default it has become the predominant theory of infant speech perception.

As early as the 1970s, however, there were a few discrepant findings that presented some challenge to the intransigent nature of the speech processing mechanism suggested by this model. First, it was found that some phonetic contrasts were discriminated more readily than others by infants. For example, Holmberg *et al.* (1977) counted the number of trials required for 6-month-olds to learn to discriminate pairs of stimuli. Using a criterion of eight correct responses out of ten consecutive trials (half change and half no-change),

Holmberg *et al.* reported that it required 64 trials, on average, for 6-month-olds to meet the criterion for /f/ versus /θ/, but only 33 trials, on average, for /s/ versus /ʃ/. Similarly, Eilers *et al.* (1977) found that infants were able to discriminate some contrasts (out of the ten presented to them), but not others.

Even when infants were found to discriminate stimuli, responses were not always strictly categorical. For example, Eimas and Miller (1980) found that 2- to 4-month-olds were able to discriminate synthetic tokens located on the same side of a /b/-to-/m/ boundary better than would be expected if perception was strictly categorical. Furthermore, even when responses were of a categorical nature, category boundaries did not always appear where they would be expected. For example, Lasky *et al.* (1975) investigated the abilities of infants in a Spanish language environment to discriminate three voicing contrasts for syllable-initial stops. One contrast placed stimuli on opposite sides of the Spanish VOT boundary: −20 ms versus +20 ms. Two contrasts placed stimuli within a Spanish voicing category: −60 ms versus −20 ms VOT and +20 ms versus +60 ms VOT. A particularly important manipulation in this study was that the +20 ms versus +60 ms contrast placed stimuli on opposite sides of the English VOT boundary. The infants in Lasky *et al.*'s study (ages 4 to $6\frac{1}{2}$ months) discriminated this contrast, even though it was not in their native language. They also discriminated the −60 ms versus −20 ms contrast, even though both of these stimuli fall within the Spanish voiced category. In fact, the one contrast they failed to discriminate was the −20 ms versus +20 ms, which crosses the Spanish voicing boundary. In spite of these seemingly contradictory findings, however, the notion of innate phonetic boundaries has persisted.

Of course, descriptions offered by various authors differ somewhat, particularly with respect to whether the focus is on the boundary or on the contrast. For example, Jusczyk (1995) writes ''Findings of this sort [as those described at the outset] have led to the view that infants are born with the capacity to discriminate contrasts that could potentially appear in any of the world's languages. Experience with language appears to have its impact by getting the infant to focus on those contrasts that play a critical role in distinguishing words in the native language.'' (p. 269) Similarly Best (1994) states ''Current findings suggest that infants begin life with language-universal abilities for discriminating segmental phonetic contrasts but that, by the second half-year of life, listening experience with the native language has begun to influence the perception of contrasts that are non-distinctive in the native phonological system.'' (p. 168) Kuhl (e.g., 1979b, 1980; Kuhl and Miller, 1982) reminds us repeatedly that an important component of any theory of innate capacities for speech-sound categorization must be a demonstration of similarity judgments for acoustically disparate members of a category. Also, there have been various modifications of the basic view. For example, Werker (1994) suggests that the loss of non-native boundaries may not be a permanent loss, as originally thought, and that boundaries may differ in how long it takes them to dissolve. Jusczyk (1998) specifically invokes the Lasky *et al.* (1975) finding to suggest that rather than there being innate phonetic categories that line up with the categories of specific languages, perhaps there is a ''language-general categorization of speech information'' (p. 56). According to this view, the ability of the infants in the Lasky *et al.* study to discriminate the English voicing contrast, but not the Spanish, can be explained as evidence that the English voicing contrast comes closer to infants' innate perceptual boundaries than does the Spanish contrast. Kuhl proposes a model in which ''acoustic space'' is linear at birth, but the regions around phonetic boundaries become warped as a result of language experience during the first six months of life (Grieser and Kuhl, 1989; Kuhl, 1991, 1993). These variations, however, fail to contradict the basic tenets of the universal theory. The predominant view continues to rely on notions of innate, universal boundaries as the starting point for human speech perception, with some form of loss as the main mechanism for developmental change.

The purpose of this brief report is to encourage reconsideration of the ''universal'' theory as it applies to infant speech perception. The experiment reported here evolved from efforts in this laboratory to extend findings and hypotheses concerning the speech perception of children roughly $3\frac{1}{2}$ to 7 years to even younger listeners. This work has shown that, at least for some phonetic distinctions, children in this age range weight the various acoustic properties upon which phonetic decisions are made differently than adults do (Nittrouer, 1992, 1996; Nittrouer *et al.*, 1998, 2000; Nittrouer and Miller, 1997a, b; Nittrouer and Studdert-Kennedy, 1987). This finding has been corroborated by others (e.g., Greenlee, 1980; Krause, 1982; Morrongiello *et al.*, 1984; Parnell and Amerman, 1978; Wardrip-Fruin and Peach, 1984), who also report that when making the same phonetic decision, children, compared to adults, pay more attention to some acoustic properties and less attention to others. Combining these two general findings (that infants are born with innate phonetic boundaries and that children weight acoustic properties differently from adults in making phonetic decisions) led to an apparent contradiction: If indeed infants are born with capacities to recognize all the phonetic contrasts in their native language (i.e., the mechanism of maintenance alone accounts for their presence into childhood), how is it that differences in phonetic decision-making are observed for children and adults? The hypothesis that emerged was that perhaps infants' discrimination abilities are based on different weighting strategies than those of adults. That is, even though infants make the same discriminations as adults, the way that they come to make these discriminations could be different. Thus efforts were undertaken to examine the relative weighting of acoustic properties in infants' phonetic decisions.

The initial assumption was that infants would surely be able to discriminate the contrast of interest. Much of the work examining developmental shifts in perceptual weighting strategies has been done using /s/-vowel versus /ʃ/-vowel contrasts, and Holmberg *et al.* (1977) showed that this contrast is well within the capabilities of 6-month-olds to discriminate. So that contrast was selected for use with infants. Quickly, however, it became clear that infants could not

readily make this discrimination, even when all properties covaried appropriately (e.g., when natural tokens were used). As a result, other contrasts were introduced that were presumed to be even more discriminable: specifically, two vowels from the corners of the vowel triangle and a VOT contrast. The focus of the study then shifted to examining the ''universal theory.''

A secondary goal of this work was to determine if infants perform differently in their perception of speech when synthetic signals are used instead of natural tokens. This question arose largely from concern expressed informally (e.g., in manuscript reviews) that children may perform differently from adults in tests using synthetic speech because children have less experience hearing these signals. The implication of such statements has clearly been that children perform worse with synthetic stimuli than they would with natural stimuli. Eilers *et al.* (1977), on the other hand, suggested that perhaps infants perform better with synthetic stimuli than they would with natural stimuli because there is no variation in acoustic properties across stimuli, save the one on which the discrimination must be made.

Finally, in addition to infants, children just 3 years of age participated in this study. In earlier work we have tested children no younger than $3\frac{1}{2}$ years largely because younger children do not perform well on the labeling tasks used. However, in addition to difficulty with the task, it has appeared that some of the poor performance of children in this age range may actually be due to phonetic categories that are more poorly specified for children than for adults. We wanted to explore that possibility with a discrimination task, using just slightly younger children than in our labeling experiments.

II. METHOD

A. Participants

Two groups of children participated: 46 infants and 75 preschoolers. Infants were between the ages of 6 and 14 months. Several investigators have successfully used headturning procedures with infants up to 14 months of age (e.g., Eilers *et al.*, 1977; Moore *et al.*, 1975), and Kuhl (1985) states that it is appropriate for infants between 5.5 and 18 months. Preschoolers were between 2 years, 6 months and 3 years, 4 months. All children (infants and preschoolers) were full-term births, with no prenatal or perinatal histories that would put them at risk for language problems. All children lived in homes with English as the only language. Children were excluded if they had a sibling or parent with a speech or language problem, or if they were not developing speech as expected. Specifically parents were asked about two well-recognized milestones, if the infant or preschooler was old enough to have reached the milestone. Children must have shown evidence of canonical babbling by nine months of age and must have started using two-word utterances by two years. All children were free from significant histories of otitis media, defined as having no more than one episode during the first year of life and no more than three episodes total. All children passed a hearing screening of the frequencies 0.5, 1.0, 2.0, 4.0, and 6.0 kHz presented free-field at 25 dB HL using either a visually reinforced headturning procedure (infants) or play audiometry (preschoolers).

B. Stimuli

All stimuli were digitized at a 20-kHz sampling rate, and low-pass filtered at 10 kHz.

1. Natural stimuli

Five sets of natural stimuli were made. Two sets consisted of stimuli that differed only in the vowel, either /sɑ/ versus /su/ or /ʃɑ/ versus /ʃu/. Because relatively stable regions of spectral information signal these contrasts they should be readily discriminated. One set of stimuli consisted of syllables with initial alveolar stops that differed in VOT, /tɑ/ versus /dɑ/. This voicing contrast was included because it has a long and thorough history of investigation, dating back to the seminal work of Lisker and Abramson (1964), and it is widely accepted that infants (e.g., Eimas *et al.*, 1971) as well as nonhuman animals (e.g., Kuhl and Miller, 1978) can make VOT discriminations. Two sets of stimuli differed only in fricative place of constriction, /sɑ/ versus /ʃɑ/ and /su/ versus /ʃu/. Much of the earlier work in this laboratory with $3\frac{1}{2}$- to 7-year-olds has focused on the /s/-/ʃ/ contrast, and so we were particularly anxious to investigate this contrast with younger listeners. For each set, three tokens of each syllable were obtained from the same speaker. Using multiple tokens made it possible to present stimuli with a roving standard and a roving comparison to ensure that discrimination was based on phonetic change, rather than on changes in other, irrelevant acoustic properties. At the same time, restricting samples to those from one speaker minimized factors that could interfere with the encoding of speech by listeners (Jusczyk *et al.*, 1992).

2. Synthetic stimuli

Synthetic versions of /su/ versus /ʃu/ and /sɑ/ versus /ʃɑ/ were prepared using a Klatt (1980) software synthesizer. The fricative noises have been used in other labeling experiments (Nittrouer, 1992, 1996; Nittrouer and Miller, 1997a, b), had a single pole, and were 230 ms long. The center frequency of the /s/ noise was 3.8 kHz, and the center frequency of the /ʃ/ noise was 2.2 kHz. Vocalic portions were 270 ms long, and two portions were synthesized for each vowel: one with a second-formant ($F2$) transition appropriate for /ʃ/ and one with an $F2$ transition appropriate for /s/. For the two /u/ portions, fundamental frequency ($f0$) started at 120 Hz and fell throughout to an ending frequency of 100 Hz. The $F1$ was constant at 250 Hz, and the third formant ($F3$) was constant at 2100 Hz. Because $F3$ was similar in frequency to the pole of the /ʃ/ noise, energy was present in this frequency region across the entire syllable when the /ʃ/ noise was used, but not when the /s/ noise was used. Stevens (1985) has suggested that one cue to fricative identity for /s/ versus /ʃ/ is the amount of amplitude change in the $F3$ region across the noise/voicing boundary, and this cue was appropriately manipulated here. For both /u/ portions, $F2$ fell through the entire portion to an ending frequency of 850 Hz. For /(s)u/, $F2$ started at 1600 Hz; for /(ʃ)u/, it started at 1800 Hz.[1]

For the two /ɑ/ portions, $f0$ started at 100 Hz, and fell through the portion to 80 Hz. For both /ɑ/ portions, $F1$ started at 450 Hz and rose over the first 50 ms to a steady-state frequency of 650 Hz. $F3$ remained constant at 2500 Hz. Again this setting maintained the relative amplitude cue in the $F3$ region described by Stevens (1985). For both /ɑ/ portions, $F2$ fell over the first 100 ms to a steady-state frequency of 1130 Hz. For the /(s)ɑ/ portion, $F2$ started at 1300 Hz; for /(ʃ)ɑ/, $F2$ started at 1500 Hz.

3. Hybrid stimuli

The vocalic portion of each of the three tokens of each natural fricative-vowel syllable was separated from the fricative noise, and combined with the synthetic /s/ and /ʃ/ noise such that the place of constriction specified by formant transitions matched that specified by the noise. Although one purpose of this study was to compare discrimination of natural and synthetic stimuli, these hybrid stimuli were also included because such tokens are frequently used in our testing with children.

C. Equipment

All testing took place in a sound-attenuated chamber. A one-way window connected the chamber with an adjacent control room. A Madsen audiometer was used to screen hearing. Speech stimuli were presented free field using a computer, a Data Translation 2801A digital-to-analog converter, a Frequency Devices 901F filter, a Tascam PA30-B amplifier, and a JBL Control-1 speaker. A special purpose board with two boxes attached to it controlled the presentation of stimuli, recorded responses, and turned on reinforcers. One box had three foot pedals attached to it which allowed the experimenter in the chamber with the infant or preschooler to start the presentation of standard stimuli, initiate trials of comparison stimuli, and temporarily interrupt the presentation of all stimuli without the child or the second experimenter observing pedal presses.[2] The other box had four buttons, and was in the control room. By pressing any one of the buttons, the experimenter in the control room recorded that a response had occurred and presented reinforcement. Reinforcement was provided by one of three Plexiglas boxes, each containing a mechanical animal, or by a graphics monitor that displayed brightly colored shapes. A total of ten mechanical animals was kept in stock, so they could be replaced between visits for any one child. For infants, a supply of quiet toys helped maintain forward eye gaze between trials. For preschoolers, reinforcement was contingent on the press of a large button, mounted on a board. The button was not connected to anything, but when the child pressed it the experimenter in the control room recorded the response.

D. Procedures

Procedures were very similar to those of most studies using a visually reinforced headturning procedure (e.g., Kuhl, 1985; Werker and Tees, 1984). One experimenter (E1) was in the chamber with the child and the child's parent. The arrangement in the chamber (shown in Fig. 1) was modified slightly from that of other investigators, who often have the child sit on the parent's lap. We found that children were less

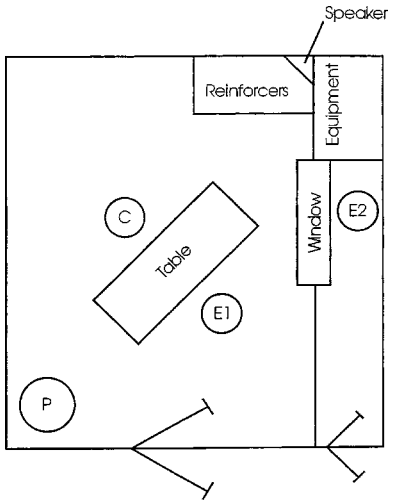

FIG. 1. Arrangement of booth during testing. C=child; P=parent; E1 and E2=experimenters.

restless if they sat in a seat by themselves (a table-mounted chair for infants; a high chair for preschoolers).[3] The child sat across a table from E1, with the parent well off to the side. The speaker and reinforcers were on the opposite side of the table from the parent. A second experimenter (E2) was in the control room.

The parent listened the entire time she was in the chamber to monologues by a male radio personality (Garrison Keillor), presented over headphones. We found that at comfortable listening levels these monologues more effectively masked the stimuli being presented than did music, probably because the $f0$'s of Garrison Keillor and of the stimuli (whether natural or synthetic) were similar. E1 listened to the stimuli during training phases, but listened to the monologues during testing. E2 listened to the stimuli during training, but then simply switched off the speaker in the control room so that stimuli were not heard during testing.

E1 used the foot pedals to initiate the presentation of the standard stimulus, and to introduce trials. Stimuli were presented at a peak intensity of 68 dB SPL, at a rate of one every 2 s. The presentation level was selected based both on Dobie and Berlin's (1979) report that a normal conversational level is between 65 and 70 dB SPL, and on Nozza's (1987) demonstration that infants' discriminations were more successful at a level close to 70 dB, rather than at the lower levels of 50 to 60 dB commonly used (e.g., Eilers *et al.*, 1977; Werker *et al.*, 1981). Three stimuli were presented during each comparison trial (i.e., 6-s trials), and stimulus presentation returned to the presentation of the standard if no response occurred. E2 pressed a button in the control room if she judged that a response had occurred (i.e., headturns on the part of infants or button-presses on the part of preschoolers). If the response was to a change trial, reinforcement lasting 3 s was presented.

Children were scheduled to participate in up to four sessions, over as many consecutive days.[4] One contrast was presented per session, although no one child received more than two phonetically different contrasts (see later in this work). During training, all trials were change trials. For preschoolers, training was straightforward: the task was ex-

plained, and they were given the training trials. For infants, training procedures required traditional conditioning. Initially, reinforcement was presented after the first presentation of the stimulus during the change trial (causing the infant to turn to look), but gradually the presentation of reinforcement became contingent on a headturn. To pass training, a child had to respond to three consecutive change trials with no prompting. Twenty trials were provided in which to meet this criterion. Again, the fricative contrast was anticipated to be the most difficult of the three contrasts used, and Holmberg *et al.* (1977) reported that it required an average of 11.2 trials for infants to train on this contrast. Consequently, it seemed reasonable to expect infants and preschoolers to train on these contrasts within the 20-trial limit, if they were going to train at all. For both infants and preschoolers, the training phase was also used to decrease the probability of false positives. For those children who initially demonstrated frequent false-positive responses, the interval between change trials was deliberately lengthened, thus diminishing those responses (Werker and Tees, 1984).

During testing, 15 trials were presented: ten change and five no-change. The criterion for passing a test phase was to get eight of ten correct responses to change trials, with no more than one response during no-change trials. This criterion is similar to that of Werker and Tees (1984).

Only E2 had a vote in deciding if a response had occurred. This experimenter was unaware of when a trial was occurring because that was controlled by E1. In many studies with infants, two experimenters must judge that a headturn occurred for it to be considered a response, although the use of just one judge for these decisions is not novel (e.g., Hirsch-Pasek *et al.*, 1987; Werker and Tees, 1984). The decision to base reinforcement on the judgment of a headturn by just one experimenter was related to the choice of ratios for change/no-change trials. We used a 2/1 ratio of change/no-change trials, instead of the more common 1/1 ratio (i.e., five change and five no-change trials) (Kuhl, 1985).[5] We chose to implement the higher ratio of change/no-change trials because the use of a 1/1 ratio gives the same weight to a lack of a headturn for a no-change trial as to a headturn for a change trial. Thus, even if 90% correct responses are required to satisfy the test criterion, a child need only respond to four changes (out of ten trials) to satisfy that criterion; that is, simply failing to turn one's head in the presence of five no-change trials would count as five correct responses. We wanted to see stronger evidence of the child responding in the presence of a change in stimulus. However, there was one drawback to using this stricter criterion. With this higher ratio of change/no-change trials, the probability of a trial being a change trial increased, so the experimenter who knew if a trial was occurring (E1) might be biased to vote that a headturn had occurred. Thus, that experimenter did not get a vote.

1. Infants

Half the infants heard a vowel contrast first and half heard the VOT contrast first. Of the infants hearing the vowel contrast first, half of them heard the contrast with syllable-initial /ʃ/ and half heard it with /s/. Also, the vowel that was the standard and the vowel that was the comparison (/ɑ/ or /u/) varied across infants. Those infants who met the test criterion for this first contrast were next presented with a fricative contrast using natural tokens, on the second day. For those infants who first heard a vowel contrast, the syllable that had served as the comparison for that contrast remained the comparison for the fricative contrast. For example, if an infant heard the vowel contrast /sɑ/ versus /su/ (with /su/ as the comparison), then the infant heard the fricative contrast /ʃu/ versus /su/ (again with /su/ as the comparison). This was done because it cannot be known whether the child is responding to a change in stimuli, or to the presence of the stimulus associated with reinforcement. If the latter, contingencies would not change for the infant from the vowel contrast to the fricative contrast. The fricative contrasts presented to infants hearing the VOT contrast first were randomly assigned. Every infant who met the test criterion for the natural fricative contrast was subsequently presented with a synthetic fricative contrast, on the third day, and the synthetic fricative-vowel syllables used with any one child remained the same as those of the natural contrast. Infants who met the test criterion for their first contrast, but failed to meet it for the natural fricative contrast, came back on the third day to repeat the first contrast. Infants who met the test criterion for these synthetic fricative stimuli were dismissed. Infants who did not meet the criterion would return for a fourth day, to be retested with the natural fricative stimuli.

2. Preschoolers

The focus of investigation with preschoolers was on fricative perception, and so the VOT contrast was not used. All preschoolers heard a vowel contrast on the first day of testing. Those who met the test criterion with vowels were presented with a fricative contrast on the second day. As with infants hearing the vowel contrast first, testing for any one child was planned so that the comparison stimulus remained the same across all contrasts. The kind of fricative stimuli first presented (natural, synthetic, or hybrid) was randomly varied across preschoolers. Children who heard natural or hybrid stimuli for the first fricative contrast, and met the test criterion, returned for a third day of testing with the synthetic fricative contrast. In this way we could ask if synthetic speech *per se* presents problems for children. As with infants, preschoolers were retested with the last contrast on which they were successful, if they failed to meet the criterion for a contrast.

For all children then, except those infants hearing the VOT contrast first, the phonetic structure of the stimulus associated with reinforcement remained constant across all conditions.

III. RESULTS

A. Infants

Of the 23 infants tested with a vowel contrast, 15 (65%) met the test criterion.[6] Of the 23 infants tested with the VOT contrast, eight (35%) met the test criterion. Of the 15 children who were able to do the vowel contrast, six (40%) were

subsequently successful with the natural fricative contrast. The nine who were not successful with the fricative contrast were all able to meet the test criterion for the vowel contrast when retested. All six infants who could discriminate the natural fricative contrast were also able to discriminate the synthetic fricatives. None of the eight infants who were successful with the VOT contrast were able to discriminate the natural fricatives, but all were able to discriminate the VOT contrast when retested.

B. Preschoolers

Forty-two preschoolers (56%) were successful with the vowel contrast. One of those children did not participate in further testing due to illness. Of the remaining 41 children tested with one of the three fricative contrasts, 24 (59%) were successful. There were no differences among the proportions of children who succeeded with the natural, synthetic, and hybrid stimuli. The 17 preschoolers who did not discriminate the fricative contrast were able to perform the vowel contrast when retested. All children tested with the natural or hybrid stimuli first were subsequently able to discriminate the fricative contrast with synthetic stimuli.

IV. DISCUSSION

The data reported here were originally collected as part of what was to be a pilot experiment, developing methods for investigating the weighting strategies of infants for the various acoustic properties that define linguistic segments. In line with the work that has been done with $3\frac{1}{2}$- to 7-year-olds, the plan was to manipulate the acoustic structure of fricative-vowel syllables to examine whether infants base discrimination judgments more on differences in the fricative noise or on differences in formant transitions. However, the principal experiment was never conducted because infants and 3-year-olds demonstrated unreliable results discriminating even clear tokens of fricative-vowel syllables, regardless of whether the fricative or vowel differed within the pair, as well as unreliable results for stimuli differing in voicing.[7] To ask the question of how much weight is given to each acoustic property in discrimination decisions would require manipulations of the stimuli that would make them somewhat perceptually ambiguous. There is every reason to believe that such manipulations would render stimuli undiscriminable, even for the infants and 3-year-olds who did discriminate these clear tokens.

What is left then are these sparse data demonstrating how difficult it is for infants, and even children as old as 3 years, to discriminate speech stimuli based on phonetic category. It is, of course, tempting to dismiss these results by suggesting that the success rates were low because of poor procedures. However, the success rates reported here are not different from those reported by others who report success rates for infants. Jusczyk and colleagues always report attrition, and they generally dismiss 40% to 45% of the infants in their studies due to ''fussiness'' (e.g., Bertoncini *et al.*, 1988; Jusczyk *et al.*, 1992; Levitt *et al.*, 1988). (Other infants may be dismissed for other reasons, as well.) The implicit assumption of that work is that infants dismissed due to fussiness would have had similar success rates as those of the nonfussy infants, if only they had not been fussy. In our work with children $3\frac{1}{2}$ to 7 years of age, however, we have not found that to be the case. We find that children may become uncooperative precisely because they cannot discriminate the stimuli presented: If these children return to the laboratory on a different day they usually become uncooperative with the same or similar stimuli, but if presented with stimuli that are not minimal pairs, these same children cooperate and perform the task appropriately. While we cannot conclusively draw a parallel from those findings with older children to work with infants, it would be inappropriate to assume that the dismissed infants would have performed as the infants who were not dismissed.

Another reason to discount the notion that procedures might have been nonoptimal, accounting for the low success rate, is that success rates differed across contrasts. If procedures accounted for a large proportion of variance in success rates, we would have expected those rates to be similar across contrasts. Of particular interest was the low proportion of infants who reached criterion on the VOT contrast. The English /dɑ/ versus /tɑ/ contrast has been used extensively to support the argument that the auditory system provides regions of enhanced sensitivity along some psychophysical continua, and those regions form natural boundaries between phonetic classes (e.g., Kuhl, 1981; Kuhl and Miller, 1978; Sinex and McDonald, 1989; Sinex *et al.*, 1991). This study was unable to address the notion of enhanced sensitivity, but clearly infants were not as successful at discriminating stimuli differing in VOT as the notion suggests they should have been.

Finally, the fact is that it is simply not that difficult to institute a headturning procedure with infants or a button-pressing procedure with 3-year-olds. These procedures are used routinely in audiology clinics to measure auditory thresholds in infants and 3-year-olds. In those settings, the procedures have proven to be fairly robust to variations in procedures, and so it was that screening children's hearing in this study presented no problems.

A minor conclusion drawn from this work was that infants and preschoolers are perfectly capable of perceiving synthetic speech. There was not one instance in which a child was able to discriminate a contrast with natural or hybrid stimuli, but unable to do so with synthetic stimuli.

Overall these results fail to provide support for claims that universal phonetic boundaries are in place at birth. It is emphasized that the findings of this study do not really differ from those of others: success rates are similar across studies. What differs is the willingness of authors to use the results to support claims of innate phonetic boundaries. In fact, some earlier studies provide evidence that could be taken to refute overtly such claims. Again, Eimas and Miller (1980) found that infants could discriminate between stimuli that both fell within an adult phonetic category, and Lasky *et al.* (1975) found that infants failed to discriminate between tokens that fell into different categories. In sum, the data across experiments do not support the proposition that infants have clearly established phonetic categories, separated by well-defined boundaries. As a field, it is important for us to bear this point

in mind because the nature of the processing that we attribute (or fail to attribute) to infants affects the questions we ask about the speech perception of older listeners. Perhaps we should not even be asking if infants have well-formed phonetic categories, separated by boundaries, but rather if any language users do. In other words, the very concept of categories, and even more so of boundaries, needs to be reconsidered.

The concept of phonetic boundaries arose from the early categorical studies using synthetic signals in which articulatorily and acoustically impossible speech sounds were constructed by manipulating a single dimension of the signal, while holding all other dimensions constant. Boundaries were, and remain, a statistical term: they are defined as the points on the distributions where half of the responses are for one phonetic category and half are for another phonetic category. In the early studies, this statistical term helped investigators to describe the acoustic correlates of phonetic segments. In more recent studies, the term helps us to understand how multiple properties influence phonetic decisions. For example, we can examine how the boundary along an acoustic continuum of one property shifts when another property is manipulated, and so understand better the nature of effect of these two properties. However, we have no evidence that boundaries exist in the natural world, or any account of how or why they might have evolved by natural selection. To extend to them any degree of psychological reality is unsupportable, and deleterious to efforts to understand how phonetic structure is indeed instantiated and retrieved from the speech signal.

ACKNOWLEDGMENTS

This work was supported by Grant No. R01 DC00633 from the National Institute on Deafness and Other Communication Disorders. Carol Manning and Gina Meyer helped with data collection. Michael Studdert-Kennedy, Carol A. Fowler, Doug Keefe, Maureen Higgins, and Arlene Carney provided helpful comments on earlier drafts of this manuscript.

[1]Throughout this manuscript, the fricative shown in parentheses indicates the one for which the *F*2 transition was appropriate.

[2]Although the capability to interrupt the presentation of stimuli during testing existed, it was rarely used. It was there merely as an option, in case it would be needed if the infant were to spit up, have a serious episode of coughing, or such.

[3]Investigators who have difficulty using a headturning task with infants older than 12 months typically do so for one of two reasons, according to Kuhl (1985): either the infant becomes restless sitting on the parent's lap or the infant wants to look for the mechanical animal in the Plexiglass box, producing frequent false-positive headturns. The first of these concerns was eliminated by our use of an infant chair, and the second concern was eliminated by the training procedure, to be discussed.

[4]As it turned out, no child had to attend more than three sessions.

[5]In fact, ratios of change/no-change trials as high as 3/1 have been used successfully (Moore *et al.*, 1975).

[6]It will not be specified here whether children failed to pass the training or the testing phase. In the end, it does not matter because in either case the child was judged not to discriminate the stimuli in the contrast.

[7]The lessons learned from this experiment did help to develop adaptive procedures that were used with 3 year olds (mean age 3 years, 7 months) to explore their perceptual weighting of some acoustic cues in speech perception (Nittrouer, 1996).

Aslin, R. N., and Pisoni, D. B. (**1980**). ''Some developmental processes in speech perception,'' in *Child Phonology*, edited by G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson (Academic, New York), Vol. 2, pp. 67–96.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., and Perey, A. J. (**1981**). ''Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience,'' Child Dev. **52**, 1135–1145.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., and Mehler, J. (**1988**). ''An investigation of young infants' perceptual representations of speech sounds,'' J. Exp. Psychol. Gen. **117**, 21–33.

Best, C. T. (**1994**). ''The emergence of native-language phonological influences in infants: A perceptual assimilation model,'' in *The Development of Speech Perception*, edited by J. C. Goodman and H. C. Nusbaum (MIT, Cambridge, MA), pp. 167–224.

Dobie, R. A., and Berlin, C. I. (**1979**). ''Influence of otitis media on hearing and development,'' in *Otitis Media and Child Development: Speech, Language and Education*,'' Ann. Otol. Rhinol. Laryngol. **88** (Suppl. 60), 48–53.

Eilers, R. E., Gavin, W., and Oller, D. K. (**1982**). ''Cross-linguistic perception in infancy: Early effects of linguistic experience,'' J. Child Lang. **9**, 289–302.

Eilers, R. E., Wilson, W. R., and Moore, J. M. (**1977**). ''Developmental changes in speech discrimination in infants,'' J. Speech Hear. Res. **20**, 766–780.

Eimas, P. D. (**1974**). ''Auditory and linguistic processing of cues for place of articulation by infants,'' Percept. Psychophys. **16**, 513–521.

Eimas, P. D. (**1975**). ''Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants,'' Percept. Psychophys. **18**, 341–347.

Eimas, P. D., and Miller, J. L. (**1980**). ''Discrimination of information for manner of articulation,'' Infant Behav. Dev. **3**, 367–375.

Eimas, P. D., Miller, J. L., and Jusczyk, P. W. (**1987**). ''On infant speech perception and the acquisition of language,'' in *Categorical Perception: The Groundwork of Cognition*, edited by S. Harnard (Cambridge U. P., New York), pp. 161–195.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., and Vigorito, J. (**1971**). ''Speech perception in early infancy,'' Science **171**, 304–306.

Greenlee, M. (**1980**). ''Learning the phonetic cues to the voiced-voiceless distinction: A comparison of child and adult speech perception,'' J. Child Lang. **7**, 459–468.

Grieser, D., and Kuhl, P. K. (**1989**). ''Categorization of speech by infants: Support for speech-sound prototypes,'' Dev. Psychol. **25**, 577–588.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright Cassidy, K., Druss, B., and Kennedy, L. (**1987**). ''Clauses are perceptual units for young infants,'' Cognition **26**, 269–286.

Holmberg, T. L., Morgan, K. A., and Kuhl, P. K. (**1977**). ''Speech perception in early infancy: Discrimination of fricative consonants,'' J. Acoust. Soc. Am. **62**, S99.

Jusczyk, P. W. (**1995**). ''Language acquisition: Speech sounds and the beginning of phonology,'' in *Speech, Language, and Communication: Handbook of Perception and Cognition, Vol. 11*, edited by J. L. Miller and P. D. Eimas (Academic, San Diego), pp. 263–301.

Jusczyk, P. W. (**1998**). *The Discovery of Spoken Language* (MIT, Cambridge, MA).

Jusczyk, P. W., Pisoni, D. B., and Mullennix, J. (**1992**). ''Some consequences of stimulus variability on speech processing by two-month-old infants,'' Cognition **43**, 253–291.

Klatt, D. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.

Krause, S. E. (**1982**). ''Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults,'' J. Acoust. Soc. Am. **71**, 990–995.

Kuhl, P. K. (**1979a**). ''The perception of speech in early infancy,'' in *Speech and Language: Advances in Basic Research and Practice*, edited by N. J. Lass (Academic, New York), pp. 1–47.

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

Kuhl, P. K. (**1980**). ''Perceptual constancy for speech-sound categories in early infancy,'' in *Child Phonology, Vol. 2: Perception*, edited by G. H. Yeni-Komshian, J. F. Kavanaugh, and C. A. Ferguson (Academic, New York), pp. 41–66.

Kuhl, P. K. (**1981**). ''Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech sound catego-

ries,'' J. Acoust. Soc. Am. **70**, 340–349.
Kuhl, P. K. (**1985**). ''Methods in the study of infant speech perception,'' in *Measurement of Audition and Vision in the First Year of Postnatal Life*, edited by G. Gottlieb and N. A. Krasnegor (Ablex, Norwood, NJ), pp. 223–251.
Kuhl, P. K. (**1987**). ''Perception of speech and sound in early infancy,'' in *Handbook of Infant Perception, Vol. 2*, edited by P. Salapatek and L. Cohen (Academic, New York), pp. 275–382.
Kuhl, P. K. (**1991**). ''Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not,'' Percept. Psychophys. **50**, 93–107.
Kuhl, P. K. (**1993**). ''Developmental speech perception: Implications for models of language impairment,'' Ann. N.Y. Acad. Sci. **682**, 248–263.
Kuhl, P. K., and Miller, J. D. (**1978**). ''Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli,'' J. Acoust. Soc. Am. **63**, 905–917.
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.
Lasky, R. E., Syrdal-Lasky, A., and Klein, R. E. (**1975**). ''VOT discrimination by four- to six-and-a-half-month-old infants from Spanish environments,'' J. Exp. Child. Psychol. **20**, 215–225.
Levitt, A. G., Jusczyk, P. W., Murray, J., and Carden, G. (**1988**). ''Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts,'' J. Exp. Psychol. [Hum Learn] **14**, 361–368.
Lisker, L., and Abramson, A. S. (**1964**). ''A cross-language study of voicing in initial stops: Acoustical measurements,'' Word **20**, 384–422.
Moffitt, A. R. (**1971**). ''Consonant cue perception by twenty- to twenty-four-week-old infants,'' Child Dev. **42**, 717–731.
Moore, J. M., Thompson, G., and Thompson, M. (**1975**). ''Auditory localization in infants as a function of reinforcement conditions,'' J. Speech Hear. Disord. **40**, 29–34.
Morrongiello, B. A., Robson, R. C., Best, C. T., and Clifton, R. K. (**1984**). ''Trading relations in the perception of speech by five-year-old children,'' J. Exp. Child Psychol. **37**, 231–250.
Morse, P. A. (**1972**). ''The discrimination of speech and nonspeech stimuli in early infancy,'' J. Exp. Child Psychol. **14**, 477–492.
Morse, P. A. (**1985**). ''Infant speech perception,'' in *Advances in the Study of Communication and Affect: Auditory Development in Children, Vol. 10*, edited by S. E. Trehub and B. Schneider (Plenum, New York), pp. 215–230.
Nittrouer, S. (**1992**). ''Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries,'' J. Phonetics **20**, 1–32.
Nittrouer, S. (**1996**). ''Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds,'' J. Speech Hear. Res. **39**, 278–297.
Nittrouer, S., and Miller, M. E. (**1997a**). ''Developmental weighting shifts for noise components of fricative-vowel syllables,'' J. Acoust. Soc. Am. **102**, 572–580.
Nittrouer, S., and Miller, M. E. (**1997b**). ''Predicting developmental shifts in perceptual weighting schemes,'' J. Acoust. Soc. Am. **101**, 2253–2266.
Nittrouer, S., and Studdert-Kennedy, M. (**1987**). ''The role of coarticulatory effects in the perception of fricatives by children and adults,'' J. Speech Hear. Res. **30**, 319–329.
Nittrouer, S., Crowther, C., and Miller, M. E. (**1998**). ''The relative weighting of acoustic properties in the perception of [s]+stop clusters by children and adults,'' Percept. Psychophys. **60**, 51–64.
Nittrouer, S., Miller, M. E., Crowther, C. S., and Manhart, M. J. (**2000**). ''The effect of segmental order on fricative labeling by children and adults,'' Percept. Psychophys. **62**, 266–284.
Nozza, R. J. (**1987**). ''Infant speech-sound discrimination testing: Effects of stimulus intensity and procedural model on measures of performance,'' J. Acoust. Soc. Am. **81**, 1928–1937.
Parnell, M. M., and Amerman, J. D. (**1978**). ''Maturational influences on perception of coarticulatory effects,'' J. Speech Hear. Res. **21**, 682–701.
Sinex, D. G., and McDonald, L. P. (**1989**). ''Average discharge rate representation of voice onset time in the chinchilla auditory nerve,'' J. Acoust. Soc. Am. **83**, 1817–1827.
Sinex, D. G., McDonald, L. P., and Mott, J. B. (**1991**). ''Neural correlates of nonmonotonic temporal acuity for voice onset time,'' J. Acoust. Soc. Am. **90**, 2441–2449.
Stevens, K. N. (**1985**). ''Evidence for the role of acoustic boundaries in the perception of speech sounds,'' in *Phonetic Linguistics*, edited by V. A. Fromkin (Academic, Orlando), pp. 243–255.
Streeter, L. (**1976**). ''Language perception shows effects of both innate mechanisms and experience,'' Nature (London) **259**, 39–41.
Wardrip-Fruin, C., and Peach, S. (**1984**). ''Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants,'' Lang. Speech. **27**, 367–379.
Werker, J. F. (**1989**). ''Becoming a native listener,'' Am. Sci. **77**, 54–59.
Werker, J. (**1991**). ''The ontogeny of speech perception,'' in *Modularity and the Motor Theory of Speech Perception*, edited by I. G. Mattingly and M. Studdert-Kennedy (Lawrence Erlbaum, Hillsdale, NJ), pp. 91–109.
Werker, J. F. (**1994**). ''Cross-language speech perception: Developmental change does not involve loss,'' in *The Development of Speech Perception*, edited by J. C. Goodman and H. C. Nusbaum (MIT, Cambridge, MA), pp. 93–120.
Werker, J. F., and Tees, R. C. (**1984**). ''Cross-language speech perception: Evidence for perceptual reorganization during the first year of life,'' Infant Behav. Dev. **7**, 49–63.
Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C. (**1981**). ''Developmental aspects of cross-language speech perception,'' Child Dev. **52**, 349–355.