Published in final edited form as:

Child Dev. 1981 ; 52(4): 1135–1145.

Discrimination of Voice Onset Time by Human Infants: New Findings and Implications for the Effects of Early Experience

Richard N. Aslin, David B. Pisoni, Beth L. Hennessy, and Alan J. Perey
Indiana University

Abstract

Discrimination of voice onset time (VOT) by 6–12-month-old infants was examined in 2 experiments. An operant head-turning technique assessed discrimination along a synthetic VOT continuum ranging from –70 msec to +70 msec. Infants from an English-speaking environment provided reliable within-subject evidence for discrimination of VOT contrasts located at both the plus and minus regions of the VOT continuum. These results provide strong evidence that infants from an English-speaking environment are capable of discriminating VOT contrasts that are not phonemic in English. Threshold Δ VOT values indicated that the infants were more sensitive to VOT differences in the plus region of the VOT continuum than in the minus region. Threshold Δ VOT values from English-speaking adults indicated greater sensitivity at every location along the VOT continuum. In addition, the adults showed heightened sensitivity to VOT differences near the voiced-voiceless boundary in the plus region of the VOT continuum, a finding that was not evident in the infants' data.

Eimas, Siqueland, Jusczyk, and Vigorito (1971) demonstrated that young infants can discriminate small differences in voice onset time (VOT), a parameter that is sufficient to differentiate the phonetic categories of voiced (*b, d, g*) from voiceless (*p, t, k*) stop consonants.¹ They could discriminate VOT differences that straddled the voiced-voiceless boundary for English-speaking adults (approximately +25 msec), whereas comparable differences in VOT selected from within the phonemic categories of English were not discriminated. The finding that infants from an English-speaking environment do not discriminate VOT contrasts drawn from within English-adult phonemic categories is surprising because some of these same contrasts are discriminated by adults from other language communities (e.g., Thai and Spanish; see Abramson & Lisker [1970] and Williams [1977]). This failure to discriminate nonphonemic contrasts implies that either (1) phonemic categories are determined primarily by genetic factors (a view favored initially by Eimas) or (2) some (or all) phonemic categories are determined largely by early linguistic experience in a particular language environment.

© 1981 by the Society for Research in Child Development, Inc.

Address reprint requests to R. N. Aslin or D. B. Pisoni, Department of Psychology, Indiana University, Bloomington, Indiana 47405. Beth Hennessy is now at the Institute of Child Development, University of Minnesota.

Reference Notes

¹Voice onset time (VOT) refers to the delay between release from stop closure and the onset of laryngeal pulsing (voicing). In synthetic consonant-vowel (CV) syllables, VOT is realized acoustically by varying the delay between the onset of energy in the first formant and higher formants and the presence of aspiration noise during this interval. In phonetic terms there are three categories of voicing in syllable-initial stop consonants: voiced, voiceless unaspirated, and voiceless aspirated. In English the phonetic distinction between voiced and voiceless unaspirated stops is not used phonemically. Thus, these two phonetic categories are collapsed into a single phonemic category that has been labeled voiced. For English, therefore, the terminology voiced-voiceless is a phonemic one and should not be confused with the more general phonetic categorization applicable to all languages.

Since the report by Eimas et al., four studies have been carried out to examine the abilities of infants to discriminate VOT contrasts that do not occur in the infant's native language (Eilers, Gavin, & Wilson 1979; Eimas 1975; Lasky, Syrdal-Lasky, & Klein 1975; Streeter 1976). The results of these four discrimination studies could be interpreted as evidence that infants from all language environments are able to discriminate VOT differences in the plus region of the VOT continuum. However, only infants from a language environment that uses the voiced-voiceless unaspirated distinction phonemically are apparently able to discriminate VOT differences in the minus region of the VOT continuum. Despite the absence of precise correspondence between the adult category boundary and the presumed peak in VOT discriminability for infants in the Lasky et al. and Streeter studies, the lack of evidence for VOT discrimination in the minus VOT region by infants from English-speaking environments has suggested that the voiced-voiceless unaspirated distinction must be "learned" through experience in a specific linguistic environment (see Eilers et al. 1979). However, several important methodological and interpretive issues have been raised recently concerning the Eilers et al. study or any study that claims to demonstrate that infants *fail* to discriminate specific VOT contrasts (see critique by Aslin & Pisoni [1980a] and reply by Eilers, Gavin, & Wilson [1980]). Thus, it is of particular theoretical importance at this time to determine whether infants from an English-speaking environment are, in fact, capable of discriminating VOT contrasts in the minus region of the VOT continuum.

In the present study, the capacities of infants to discriminate VOT differences were examined over a broad range of a synthetic VOT continuum encompassing voiced, voiceless unaspirated, and voiceless aspirated bilabial (*b*, *p*, *pʰ*) stop consonants. The limits of the infants' ability to detect small differences in VOT were assessed in an operant head-turning paradigm using an adaptive staircase technique. If infants show reliable evidence for discrimination of VOT contrasts in the minus region of the VOT continuum, then the claim that the voiced-voiceless unaspirated distinction must be learned from early experiences in the language-learning environment would be substantially weakened. In addition, such results would imply that previous failures to demonstrate that infants can discriminate phonemically irrelevant VOT contrasts may well have been subject to subtle methodological artifacts which led to underestimation of the infants' sensory capabilities.

Experiment 1

Method

Subjects—A total of 92 infants ranging in age from 5.5 to 11.5 months participated in Experiment 1. Infants were solicited on a volunteer basis by letter and phone from the birth announcements in the local newspaper. All infants were presumed to have normal hearing. Forty-two of the infants failed to show any reliable evidence of acquiring the head-turn response after two sessions and were dropped from the study. This attrition rate of 45% compares favorably to other methods used in testing speech discrimination with infants, such as high-amplitude sucking and heart rate. Although other investigators (Eilers et al. 1979; Eilers, Wilson, & Moore 1979; Kuhl 1980) have reported virtually no dropout in their studies using the head-turning technique, they have employed either a very small number of trials (six or eight) or more easily discriminable vowel contrasts.

Stimuli—The stimuli consisted of a set of 15 synthetic labial stop consonant-vowel syllables that were generated on a cascade-parallel software synthesizer (Klatt 1980; Kewley-Port, Note 1). The 15 stimuli differed in 10-msec steps of VOT from –70 msec to

¹Kewley-Port, D. KLTEXC: executive program to implement the KLATT software speech synthesizer (Research on Speech Perception Progress Report No. 4). Bloomington: Indiana University, 1978.

+70 msec. The values used for synthesis of these stimuli were chosen from measurements of natural speech originally made by Klatt (Note 2) as well as our own measurements from spectrograms of the speech of a native English speaker who was a trained phonetician, producing various voicing contrasts. The stimuli consisted of a 255-msec steady-state pattern with formant values appropriate for the vowel /a/ (F1 = 700 Hz, BW1 = 90 Hz; F2 = 1,200 Hz, BW2 = 90 Hz; F3 = 2,600 Hz, BW3 = 130 Hz; F4 = 3,300 Hz, BW4 = 400 Hz; F5 = 3,700 Hz, BW5 = 500 Hz). The formant transitions into the vowel were 40 msec in duration and had starting frequencies appropriate for the bilabial stop in stressed syllable-initial position (F1 = 438 Hz, F2 = 1,025 Hz, F3 = 2,425 Hz). Voicing lead was simulated by passing a combination of the sinusoidal voicing source and normal voicing through F1 which was set at 180 Hz with a bandwidth of 150 Hz. The amplitude values of the voicing source were chosen to match natural productions measured from broad-band spectrograms and average amplitude contours. A 10-msec release burst was generated by passing a turbulent noise source through the bypass channel of the parallel branch of the synthesizer, which has a broadband (5 kHz) flat spectrum. The amplitude of the release burst was chosen on both theoretical and empirical grounds after a long series of listening tests. Finally, the aspiration associated with voiceless stops was generated by passing a noise source through the cascade branch of the synthesizer to simulate the turbulence produced at the glottis. The amplitude of aspiration was again chosen to match measurements of spectrograms of natural speech. During the period of aspiration, the bandwidth of F1 was also widened to 300 Hz. To simulate breathiness at the end of the syllable, the bandwidth of F1 was widened linearly from 90 Hz to 180 Hz and some aspiration noise was added to the final 35 msec of the stimulus. The pitch contour had a slight rise at the onset of the release of the consonant from 120 Hz to 125 Hz and then fell linearly to 100 Hz over the remaining steady-state portion of the vowel.

Apparatus—The infants were tested in a single-wall sound-attenuated booth (IAC Model 402, internal dimensions 2×2 m, ambient noise level = 40 dB, C scale) while seated on a parent's lap. An assistant was seated facing the infant and parent and used small silent toys to attract the infant's gaze during the testing procedure. The booth was equipped with a single loudspeaker (Radio Shack MC-1000), a cassette tape deck that presented masking music over headphones to the assistant and parent, and a visual reinforcer. The visual reinforce consisted of a smoked Plexiglas enclosure containing a motorized toy animal that could be activated by the experimenter from outside the testing booth. When the reinforce was activated, the inside of the Plexiglas enclosure was illuminated by two fluorescent bulbs and the toy animal became animated (drummed, clapped, or jumped).

The entire experimental procedure was controlled online by a PDP-11/34 computer. The experimenter viewed the infant on a closed-circuit TV monitor during the entire testing procedure, initiated trials, and coded the infant's head-turn responses with a button box that was interfaced to the computer. The computer program presented the speech sounds at a 10-kHz sampling rate through a 12-bit D/A converter which was low-pass filtered at 4.8 kHz before going to the loudspeaker in the booth. The program coded the experimenter's responses, presented all stimuli, and controlled the presentation of the visual reinforcer.

Procedure—We employed a variant of the operant head-turning technique originally developed by Moore, Wilson, and Thompson (1977), with the addition of procedures to eliminate experimenter bias and an adaptive staircase testing algorithm to estimate the limits of the infant's discriminative abilities. The basic operant head-turning procedure involves

²Klatt, D. H. Analysis and synthesis of CV syllables in English. Unpublished manuscript, Massachusetts Institute of Technology, 1978.

the repetitive presentation of a background signal that is interrupted by several repetitions of a novel target signal. The discriminative response is a unidirectional head turn toward the visual reinforcer and loudspeaker, which in our laboratory were located 90 degrees to the infant's left. A criterial head-turn response was a 45-degree turn. Evidence of positive discrimination in this procedure consists of a head turn on stimulus-change (experimental) trials and the absence of a head turn on no-change (control) trials.

Our procedure involved three distinct phases: (1) shaping, (2) criterion testing, and (3) staircase testing. Throughout all three phases the speech stimuli were presented with a constant interstimulus interval of 1,200 msec, thus eliminating any potential temporal cue that might occur during the shift from background to target. During the shaping phase the experimenter viewed the infant on a TV monitor and, when the infant's gaze was steadily directed toward the toys being manipulated by the assistant, she initiated an experimental trial which consisted of three repetitions of the target stimulus. In the first testing session for each infant, shaping trials began with the background at 70 dB (A scale) and the target at 80 dB. This redundant intensity cue was used to facilitate the infant's acquisition of the basic orienting response toward the sound change. On subsequent shaping trials, the intensity of the target was reduced in 5-dB steps until the target and background were equated at 70 dB. The experimenter typically reduced the target intensity after two or three trials on which a correct head-turn response was observed. Throughout the shaping phase the experimenter could hear the stimuli being presented to the infant, thus enabling her to judge how well the infant was performing. In addition, the experimenter sometimes presented the visual reinforcer simultaneously with the onset of the sound change to establish the contingency between sound change and visual reinforce for the infant. When the experimenter judged that the infant's head-turn response was consistently present on trials in which the target and background were equal in intensity, the computer program was signaled to proceed to the criterion testing phase.

During the criterion testing phase the experimenter wore headphones through which a tone was presented in synchrony with the background and target stimuli. The experimenter continued to initiate trials whenever the infant's gaze was directed toward the toys being manipulated by the assistant in the booth. However, trials now consisted of two types: experimental (change) or control (no change). These two trial types were presented according to one of eight pseudorandom orders (Fellows 1967) with the constraint that no more than two control trials would be presented successively. The observation interval for scoring the infant's head-turn response began with the onset of the first target repetition and ended 2 sec after the third (and final) target repetition (total scoring interval = 5.285 sec). If the experimenter judged that a head turn had occurred during the scoring interval on an experimental trial, a button was pressed and the computer immediately presented the visual reinforce for 3 sec. If the head-turn button was pressed on a control trial, the computer did not present the visual reinforcer. The computer program also kept a running tally of the infant's percentage of correct responses on the previous five experimental trials and the previous five control trials. When this percent correct met an 80% correct criterion for *both* experimental and control trials, the criterion testing phase was terminated and the staircase testing phase was begun. This 80% criterion, applied to the five experimental and five control trials, results in a probability of .055, based on the binomial expansion, $p = .5$, $q = .5$, of falsely accepting the hypothesis that the infant was responding above chance. Note that the experimenter, parent, and assistant were unaware of the type of trial (experimental or control) presented to the infant during the criterion testing phase.

In the staircase testing phase, the VOT value of the target stimulus was varied as a function of the accuracy of the infant's performance on preceding trials. A staircase algorithm was used to measure the smallest VOT difference between the background and target that the

infant could reliably discriminate. The algorithm followed a two up/one down rule in which two correct responses at a given target VOT value resulted in a change in the VOT value of the target that decreased the background-target VOT difference. A single incorrect response at a given target VOT value resulted in an increase in the background-target VOT difference. This staircase algorithm generates a probability of correct response equal to 70.7% (see Levitt 1970).

Figure 1 illustrates the response sequence of a particular infant tested in the staircase phase on a VOT series with 0 msec as the background stimulus and -70 msec as the initial target stimulus. The VOT step size in this series was 10 msec, so that two correct responses (hits) or one incorrect response (miss) resulted in a 10-msec change in the target VOT value, either up or down, relative to the constant background signal. This infant consistently responded to the shift from background to target at target VOT values of -70, -60, -50, -40, and -30 msec before failing to respond on the first trial at -20 msec. The “miss” at -20 msec resulted in a return of the target VOT value to -30 msec on the subsequent trial. As shown in figure 1, the two up/one down algorithm generates a series of reversal points which can be used to estimate the target VOT value that is discriminated from the background at the 70.7% level. At least five reversals were required in a given staircase testing session for the infant’s data to be considered reliable. The staircase phase was terminated automatically when six reversal points had been collected. The estimate of the VOT threshold value was computed as the mean of the midpoints of the five or six reversal points.

Finally, we also included *probe* trials which occurred whenever the infant failed to respond correctly on two consecutive trials. Probe trials consisted of presenting the VOT value of the initial target (the largest VOT difference). These probe trials not only helped to maintain the infant’s interest in the task during brief periods of inattention but also provided an objective behavioral criterion that could be used to terminate a session prior to reaching the five-reversal-point requirement. If an infant responded correctly on a probe trial, the subsequent trial returned to the previous location in the staircase sequence (see fig. 1). If, however, the infant failed to respond on a probe trial, another probe trial was presented. Errors on three consecutive probe trials resulted in termination of the session. When termination occurred in this manner the data were judged to be unacceptable on the grounds that the infant had not maintained selective attention to the criterial properties of the stimuli.

Design—All infants were initially tested on the full VOT series, with assignment of target (-70 or +70 msec) and background (+70 or -70 msec) randomized across subjects. Of the 50 infants who progressed beyond the shaping phase, 44 completed the criterion testing phase on one of the two full VOT series. Twenty-seven of these infants who met criterion in the testing phase completed the staircase testing phase on the full VOT series. This additional subject attrition resulted from the high demands of completing both the criterion and staircase testing phases in a single session (average number of sessions to completion = 3.5). The step size used in the staircase testing phase with the full VOT series was 20 msec. Those infants who failed to meet the 80% correct criterion (6 of 50) and those who failed to complete the staircase phase (17 of 44) either were unable to return for additional testing sessions or showed little further interest in the visual reinforcer.

After completion of the staircase testing phase on the full VOT series, each infant returned for further testing sessions on several truncated versions of the full VOT series. Each of these truncated VOT series spanned a range of 70 msec from background to target and employed a step size of 10 msec. If the infant’s background VOT value on the full series was +70 msec, then the background value on subsequent truncated series was +70, +50, +20, and 0 msec. If the initial background VOT value was -70 msec, then on the truncated series

it was 0, -20, -50, and -70 msec. Thus, the initial truncated series was located in the plus region of the VOT continuum, and subsequent truncated series were gradually shifted to the minus region of the VOT continuum.

Finally, it became clear that many infants became satiated with both the visual reinforce and the entire testing procedure after several sessions. Thus, within a session on the truncated VOT series, the criterion testing phase was eliminated for approximately half of the infants. These infants, however, had met criterion during the full VOT series. For these infants, testing on the truncated series proceeded directly from the shaping phase to the staircase testing phase.

Adult controls—Two groups of 10 adults were tested with the same VOT stimuli in a procedure identical to that used with the infants. Each adult was tested individually in the same sound-attenuated booth and listened to the stimuli through the same speaker. The adults were instructed to respond to any difference in the background signal by raising their hand or turning their head to view the reinforcer. The reinforcer was modified by placing a “correct” sign inside the Plexiglas enclosure in front of the animated toy. One group of adults was assigned to the -70-msec background and the other group to the +70-msec background for testing on the full VOT series. Subsequent tests on the four truncated series proceeded in the same manner as with the infants. Since the adults had little difficulty in discriminating the five VOT contrasts, they proceeded directly from the shaping phase to the staircase testing phase. Data for all five VOT series were collected in a single 1-hour testing session.

Results and Discussion

Mean thresholds for the 27 infants who completed the staircase phase on one of the two full VOT series are shown in table 1. Note that the target VOT values at which the infants ceased to respond (i.e., the estimates of their thresholds) were located at or near the VOT value typically associated with the voiced-voiceless boundary in English-speaking adults. The mean threshold value for these 27 infants was +21.6 msec. On the basis of these data alone, it would appear not only that this location in the plus region of the VOT continuum is very distinctive perceptually, but that infants from an English-speaking environment are incapable of reliably discriminating VOT differences in the minus region of the VOT continuum. However, a stronger test of this hypothesis requires a closer examination of infants’ discriminative abilities within more restricted regions of the VOT continuum. For example, if infants process VOT differences categorically, then the threshold value obtained in our staircase procedure could be considered equivalent to a category boundary as determined from adult labeling data. This boundary value should not vary significantly as the background and initial target VOT values are manipulated, provided of course that the initial contrast straddles the boundary. The data obtained on the truncated VOT series address this issue of possible equivalence between threshold values and category boundaries.

Eleven of the 27 infants who completed the full VOT series went on to successfully complete the staircase testing phase on one of the two truncated VOT series in the 0 to +70 msec region of the VOT continuum. Seven infants subsequently completed another truncated series in the -20 to +50 msec region of the VOT continuum. Mean threshold values on these four VOT contrasts that straddled the English-adult voicing boundary are shown in table 1. First, note that on the two +70 series the threshold values were similar to those obtained on the full (-70 vs. +70) series. The mean threshold value for the 11 infants who completed the +70 series was +29.6 msec compared to the +21.6-msec mean on the full VOT series. These data from the truncated +70 series are presumably more accurate

estimates of the threshold value because the staircase step size was now set at 10 msec rather than 20 msec, the value used in the full VOT series. Second, note that the threshold value was shifted slightly in the +50 series compared to the +70 series. The mean value in the +50 series was +12.4 msec, significantly less than the +29.6 msec in the +70 series, $t = 2.59$, $df = 13$, $p < .05$. These results from two truncated VOT series that spanned 0 to +70-msec and -20 to +50-msec regions indicate that threshold values vary as a function of the initial location of the VOT contrast. This shift in VOT threshold values indicates that infants do not simply parse the VOT continuum into discrete perceptual categories. Further support for this claim is provided by the data from the other truncated VOT series.

Although not all infants provided threshold values on all four truncated VOT series because of scheduling conflicts and satiation to the reinforcer, several infants proceeded from the +70 and +50 series to the two series in the minus region of the VOT continuum. Eight infants successfully completed the staircase phase in one of the two -50 series, and three infants completed the staircase phase in one of the two -70 series (see table 1). Note that 10 different infants provided threshold values for stimulus contrasts that ranged from -70 to +20 msec. Neither of these stimulus contrasts crossed the category boundary for voicing differences employed by adult speakers of English. Although these threshold values were more variable than those obtained with contrasts in the plus region of the VOT continuum, these infants nevertheless demonstrated reliable evidence of discriminating these phonemically irrelevant VOT contrasts. Finally, the mean threshold values in the -50 and -70 series were -6.6 msec and -32.0 msec, respectively, a difference that failed to reach significance due to the very small sample size.

A comparison between the infants and the adults tested in the staircase procedure is also shown in table 1. These comparisons must be made somewhat cautiously because of the low sample size for several of the VOT series from the infants. Nevertheless, there are consistencies in the data that suggest a tentative conclusion regarding the ability of infants and adults to discriminate differences in VOT. For adults, mean threshold values for contrasts that straddled the English adult voicing boundary of +25 msec ranged from +7.4 to +42.9 msec. For infants, the comparable mean threshold values ranged from +4.9 to +29.6 msec. And, for adults and infants, the mean threshold values for contrasts in the minus region of the VOT continuum ranged from +2.8 to -44.8 msec and from +11.1 to -45.5 msec, respectively. To clarify these results with respect to each VOT contrast, we also computed Δ VOT values from the mean infant and adult thresholds. The Δ VOT values were calculated by subtracting each threshold value from its corresponding background VOT value. Adults showed Δ VOT values that were, across all contrasts, 20.1 msec *less* than the infants' Δ VOT values. Thus, the magnitude of VOT difference between the target and background that was required to discriminate the target at the 70.7% level was consistently greater for infants than for adults.

The Δ VOT values shown in table 1 illustrate several important points regarding the infants' capacities to discriminate differences in VOT. Note that infants tested on the four truncated series that straddled the English adult voicing boundary showed Δ VOT values that were smaller than on the four series that did not straddle the English boundary (weighted means = 38 and 52 msec). Adults showed the same pattern, but overall they were more sensitive to VOT differences than the infants (weighted means = 20 and 35 msec). Finally, as shown in figure 2, the adults showed a trough in their Δ VOT values between backgrounds of 0 and +50 msec, indicating greater sensitivity to VOT differences in this region. This enhanced sensitivity to VOT differences corresponds to a region of the VOT continuum utilized by English-speaking adults in making the phonemic distinction between voiced and voiceless stop consonants. Note that in this region the adults' performance was very close to the limit of the staircase procedure (minimum = one step), since their threshold VOT values were less

than two steps from the background. If a smaller step size had been used with the adults, the dip in Δ VOT values may have been more pronounced, presumably near the +25-msec voiced-voiceless boundary.

In contrast to the adults, the infants showed only a slight dip in the Δ VOT function at 0 msec. Thus, in addition to the infants' poorer overall sensitivity to VOT differences compared to adults, they also displayed little or no evidence of heightened sensitivity to VOT differences at the region of the VOT continuum associated with the boundary between phonemic categories in English-speaking adults. The overall level of sensitivity to VOT differences could reflect either actual developmental differences or attentional and task variables. However, the age difference in the trough in the two Δ VOT functions shown in figure 2 may reflect an actual developmental difference in VOT sensitivity. However, the absence of an adult-like trough in the infants' Δ VOT function was based largely on the -50-msec versus +20-msec VOT series, a contrast that does not cross the English adult voicing boundary. A complementary contrast located entirely within the plus VOT region (e.g., +90 msec vs. +20 msec) may have facilitated discrimination of the endpoints and may have resulted in a phonemically relevant dip in the infants' Δ VOT function. Several other studies of infants have, in fact, demonstrated reliable VOT discrimination near the English adult boundary using fixed contrasts differing by as little as 20 or 30 msec. The smallest mean Δ VOT value obtained in the present study was 29.6 msec. It should be noted, however, that individual infants showed Δ VOT values as small as 9 msec, and our procedure generated a 70% correct threshold rather than the 50% correct threshold used in past studies. Nevertheless, it remains for more detailed investigations to reconcile the noncategorical nature of the infants' flat Δ VOT function (see fig. 2) with previous reports demonstrating more categorical-like VOT discrimination in infants.

Experiment 2

The results from Experiment 1 provide strong evidence that infants from an English-speaking environment are capable of discriminating VOT differences that cross both the English adult voicing boundary *and* the voiced-voiceless unaspirated boundary employed phonemically in other languages such as Spanish and Thai. In addition, the infants tested in our procedure appeared to show less sensitivity to VOT differences that were located in the minus region of the VOT continuum compared to equivalent VOT differences located in the plus region. This differential sensitivity mirrors the performance of adults tested under identical experimental conditions. Despite the similarities between the adult and infant data, it could be argued that the infant data were potentially confounded by the effects of testing VOT differences in the minus region after having tested VOT differences in the plus region. The order of presentation of VOT contrasts across sessions may have biased the infants to attend primarily to voicing lags, and the repeated testing sessions may have satiated the infants to the testing situation and the visual reinforcer.

In an effort to provide a stronger test of the hypothesis that infants from an English-speaking environment are capable of discriminating VOT contrasts in the minus region of the VOT continuum, we conducted a second experiment with two additional groups of infants. Each infant was tested on only two VOT series: (1) the full -70-msec versus +70-msec series and (2) the -70-msec versus 0-msec series (both counterbalanced for background VOT value). With the experimental conditions designed in this way, infants should be less influenced by the order of presentation of the VOT contrasts. Moreover, with substantially fewer VOT contrasts, requiring fewer testing sessions, the infants should become less satiated to the visual reinforcer.

Method

Subjects—A total of 33 infants ranging in age from 6 to 11.5 months participated in Experiment 2. Infants were obtained from birth announcements in the local newspaper as in Experiment 1. Seventeen infants were assigned to the +70-msec versus –70-msec group (background = –70 msec) and 16 infants to the –70-msec versus +70-msec group (background +70 msec).

Stimuli—The stimuli used in Experiment 2 were the same set of 15 synthetic bilabial stop consonant-vowel syllables used in Experiment 1.

Apparatus—The apparatus used in Experiment 2 was identical to that described in Experiment 1.

Procedure—The procedure used in Experiment 2 was identical to that used in Experiment 1 except that only two VOT series were tested with each infant. The infant was initially assigned to one of the two full VOT series (–70 vs. +70 or +70 vs. –70). The infant proceeded through the shaping phase to the criterion testing phase, which also employed an 80% correct criterion as in Experiment 1. The staircase phase immediately followed the criterion testing phase for all infants who had met the 80% criterion. However, infants who failed to provide five or six reversals in the staircase phase within the same session as the criterion testing phase were not required to repeat the criterion testing phase again before proceeding to the staircase phase in the next session. Those infants who successfully completed five or six reversals in the staircase phase returned for additional testing on a single truncated VOT series. Infants assigned to the full series with –70 msec as the background were subsequently tested on the 0-msec versus –70-msec series, which also had a –70-msec background. Infants assigned to the full series with +70 msec as the background were subsequently tested on the –70-msec versus 0-msec series, which had a background of 0 msec. All infants were required to meet the 80% correct response level during the criterion testing phase on the truncated series (–70 vs. 0 or 0 vs. –70) before proceeding to the staircase testing phase.

Results and Discussion

Twenty-two of the 33 infants met the 80% criterion on one of the two full VOT series. The average number of sessions required to meet this criterion was 2.2 (range = 1–4). Although each of the 22 infants who met criterion was performing at the 80% correct level at the end of the criterion testing phase (i.e., the last 10 trials), it could be argued that their overall performance in reaching this criterion may have been below chance. To test this possibility, each infant's percentage correct was computed for all trials during the session in which the criterion was met. This mean percentage correct across all 22 infants was 79.2%, a value that is significantly above the chance response level of 50%, $t = 16.1$, $df = 21$, $p < .001$. Finally, the mean number of trials within the criterion testing phase required to meet the 80% criterion was 15.9 (SD = 6.3).

Fifteen of the 22 infants who met the 80% criterion completed the staircase testing phase on one of the two full VOT series. The mean threshold value for these 15 infants was +20.4 msec, a value that is nearly identical to the +21.6-msec value obtained earlier in Experiment 1. The average number of sessions required to complete the staircase phase was 2.9 (range = 1–4).

Twelve of the 15 infants who completed the staircase phase on one of the full VOT series met the 80% criterion on one of the two truncated series located in the minus region of the VOT continuum (–70 vs. 0 or 0 vs. –70). The three infants who did not meet the 80%

criterion were unable to return for additional testing sessions and thus could not be tested on either of the truncated series. All 12 infants who were tested on the truncated series met the 80% criterion, and they did so within an average of 1.8 additional sessions (range = 1–4). Again, for the session during which the 80% criterion was met, the overall percentage correct was 72.7%, a value that is significantly above the chance level of 50%, $t = 8.2$, $df = 11$, $p < .001$. Finally, the mean number of trials within the criterion testing phase required to meet the 80% criterion was 14.7 (SD = 5.8).

Six of the 12 infants who met the 80% criterion on one of the truncated VOT series completed the staircase testing phase. Threshold values for the three infants who completed the series with a 0-msec background VOT were –51, –54, and –63 msec, values similar to the –45.5-msec mean from Experiment 1. Threshold values for the other three infants with a –70-msec background VOT were –5, –6, and –12 msec, values also similar to the –5-msec mean obtained in Experiment 1. Collapsing across Experiments 1 and 2, there were a total of 9 infants who provided Δ VOT values in one of the two –70-msec to 0-msec series and 11 infants in one of the two 0-msec to +70-msec series. The mean Δ VOT for the –70 series was 56.8 msec, a value significantly greater than the mean Δ VOT for the +70 series of 33.6 msec, $t = 4.21$, $df = 16$, $p < .001$.

In summary, the data from Experiments 1 and 2 provide strong evidence that infants from an English-speaking environment are capable of discriminating VOT contrasts that do *not* cross the English adult voicing boundary. Of particular importance is the finding from Experiment 2 that all infants tested met the 80% criterion on the VOT contrasts in the minus region of the continuum and that their overall performance across all trials was significantly above chance. However, the findings show that the infants' sensitivity to VOT differences is poorer in the minus region of the VOT continuum than in the plus region. Moreover, infants do not show strong evidence of heightened sensitivity to VOT differences near the location of the VOT boundary for voicing in English-speaking adults.

General Discussion

The results of the present set of experiments offer the first demonstration that infants from an English-speaking environment can reliably and consistently discriminate a phonologically irrelevant VOT contrast. What factors are responsible for the failure of other investigators to show that infants from an English-speaking environment can discriminate VOT contrasts in the minus region of the VOT continuum? It appears that the acceptance of categorical perception as a basic mechanism in both adults' and infants' processing of VOT differences has typically led to an examination of very small VOT contrasts. The idealized version of categorical perception implies that any VOT difference that crosses the category boundary will be discriminated. However, there are very few studies in the adult literature that have shown discrimination performance which matches this idealized conception of categorical perception. Discrimination is never perfect near the category boundary, and, in fact, it rarely exceeds 90% correct for contrasts as large as 20 msec. Given the rigorous task demands confronting infants in speech perception experiments, as well as the absence of any control over their attentive and motivational states, one would not expect their performance to approach that of adults. Thus, the robust evidence of VOT discrimination in the plus region of the VOT continuum indicates a remarkable sensitivity to this acoustic information even in infants. And the absence of robust evidence of VOT discrimination in the minus region of the VOT continuum, at least in past studies, indicates that the acoustic information contained in such contrasts is less salient to both the infant and the adult. The present study corroborates these earlier findings, but the application of a new assessment procedure provides strong evidence that contrasts in the minus VOT region *can* be discriminated reliably by prelinguistic infants.

The present results imply that the idealized view of discrete perceptual categories in infant speech perception should be moderated. The discrimination of VOT differences may be more fruitfully conceptualized according to a Δ VOT function which attempts to specify VOT sensitivity at many locations along the VOT continuum. Both adults and infants from an English-speaking environment are less sensitive to VOT differences in the minus region of the VOT continuum. However, both infants and adults from an English-speaking environment are capable of discriminating VOT differences in the minus VOT region at above-chance levels. Until detailed psychophysical studies are conducted on samples of infants from language environments that employ minus VOT contrasts phonemically, it will be impossible to determine whether the lower sensitivity in the minus VOT region is the result of selective early linguistic experience or maturation of underlying auditory mechanisms (see Aslin & Pisoni 1980b). Nevertheless, at the present time the weight of the evidence from infant and adult speech-perception studies suggests that the lower sensitivity in the minus region of the VOT continuum is acoustically rather than phonetically based. For example, Pisoni (1977) has shown that adults exhibit categorical perception of pure tone nonspeech signals and that those signals which are analogous to minus VOT stimuli are less discriminable than are signals analogous to plus VOT stimuli. In addition, Lisker (1978) and Stevens and Klatt (1974) have noted that syllable-initial stop consonants in the plus region of the VOT continuum contain additional cues (aspiration and absence of the F1 transition) that provide salient acoustic information *not* present in stops from the minus region of the VOT continuum (see also Summerfield & Haggard 1977).

Finally, the Δ VOT results from the present study provide evidence that linguistic experience may operate to enhance basic sensitivity to VOT differences only at regions along the VOT continuum that are used phonemically (see Aslin & Pisoni 1980b). Whether the complementary process of *attenuation* of sensitivity to VOT differences that are not used phonemically also operates during the period of early linguistic experience remains unclear at the present time. However, recent experiments with adults (Pisoni, Aslin, Perey, & Hennessy, in press) strongly suggest that the basic auditory capacities to perceive phonemically irrelevant VOT distinctions are not eliminated by lengthy periods of naturally occurring phonological deprivation (i.e., the absence of nonphonemic contrasts in the adult productions heard by the developing child). Rather, such distinctions can be “reacquired,” as evidenced by the accurate and reliable discrimination and identification functions obtained from adults after minimal training experience. Moreover, infants can discriminate pure tone nonspeech signals (analogous to VOT) in both the minus and plus regions of a tone-onset-time continuum (Jusczyk, Pisoni, Walley, & Murray 1980). This finding from infants supports the interpretation that acoustic rather than phonetic mechanisms underlie VOT discrimination in both the minus and plus regions of the VOT continuum. It is the task of future studies to delineate more accurately the processes by which sensitivity to phonetic contrasts changes during the period of early linguistic experience. Such findings should help us to better understand the interactions between early experience and genetic predispositions in speech perception, as well as the level of analysis mediating the infant’s discrimination and categorization of speech sounds.

Acknowledgments

This research was supported by research grants from NIH (HD-11915 and NS-12179) and NIMH (MH-24027 and MH-30424). The first author was also supported by an NICHD Research Career Development Award (HD-00309). An earlier version of Experiment 1 was presented at the biennial meeting of the Society for Research in Child Development, San Francisco, March 1979. We thank Jerry C. Forshee and David Link for their technical assistance, Diane Kewley-Port for assisting in the synthesis of the speech stimuli, Wendy Crawford and Kathy Mitchell for organizing and implementing the data collection process, and Nancy Layman for manuscript preparation. Address reprint requests to R. N. Aslin or D. B. Pisoni, Department of Psychology, Indiana University, Bloomington, Indiana 47405. Beth Hennessy is now at the Institute of Child Development, University of Minnesota.

References

- Abramson, AS.; Lisker, L. Proceedings of the 6th International Congress of Phonetic Sciences. Academia: Prague; 1970. Discriminability along the voicing continuum: cross language tests.
- Aslin RN, Pisoni DB. Effects of early linguistic experience on speech discrimination by infants: a critique of Eilers, Gavin, & Wilson, (1979). *Child Development*. 1980; 51:107–112. (a). [PubMed: 7363730]
- Aslin, RN.; Pisoni, DB. Some developmental processes in speech perception. In: Yeni-Komshian, GH.; Kavanagh, JF.; Ferguson, CA., editors. *Child phonology. Perception*. Vol. Vol. 2. New York: Academic Press; 1980. (b)
- Eilers RE, Gavin WJ, Wilson WR. Linguistic experience and phonemic perception in infancy: a cross-linguistic study. *Child Development*. 1979; 50:14–18. [PubMed: 446199]
- Eilers RE, Gavin WJ, Wilson WR. Effects of early linguistic experience on speech discrimination by infants: a reply. *Child Development*. 1980; 51:113–117. [PubMed: 7363731]
- Eilers RE, Wilson WR, Moore JM. Speech discrimination in the language-innocent and the language-wise: a study in the perception of voice-onset-time. *Journal of Child Language*. 1979; 6:1–18. [PubMed: 429462]
- Eimas, PD. Speech perception in early infancy. In: Cohen, LB.; Salapatek, P., editors. *Infant perception: from sensation to cognition*. Vol. Vol. 2. New York: Academic Press; 1975.
- Eimas PD, Siqueland ER, Jusczyk PW, Vigorito J. Speech perception in infants. *Science*. 1971; 171:303–306. [PubMed: 5538846]
- Fellows BJ. Chance stimulus sequences for discrimination tasks. *Psychological Bulletin*. 1967; 67:87–92. [PubMed: 6045339]
- Jusczyk PW, Pisoni DB, Walley A, Murray J. Discrimination of relative onset time of two-component tones by infants. *Journal of the Acoustical Society of America*. 1980; 67:262–270. [PubMed: 7354194]
- Klatt DH. Software for a Cascade/Parallel formant synthesizer. *Journal of the Acoustical Society of America*. 1980; 67:971–995.
- Kuhl, PK. Perceptual constancy for speech-sound categories. In: Yeni-Komshian, GH.; Kavanagh, JF.; Ferguson, CA., editors. *Child phonology. Perception*. Vol. Vol. 2. New York: Academic Press; 1980.
- Lasky RE, Syrdal-Lasky A, Klein RE. VOT discrimination by four and six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*. 1975; 20:215–225. [PubMed: 1194843]
- Levitt H. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*. 1970; 49:467–477. [PubMed: 5541744]
- Lisker L. In qualified defense of VOT. *Language and Speech*. 1978; 21:375–383. [PubMed: 750798]
- Moore JM, Wilson WR, Thompson G. Visual reinforcement of head-turn responses in infants under twelve months of age. *Journal of Speech and Hearing Disorders*. 1977; 42:328–334. [PubMed: 881813]
- Pisoni DB. Identification and discrimination of the relative onset time of two-component tones: implications for voicing perception in stops. *Journal of the Acoustical Society of America*. 1977; 61:1352–1361. [PubMed: 881488]
- Pisoni DB, Aslin RN, Percy AJ, Hennessy BL. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*. in press.
- Stevens KN, Klatt DH. Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*. 1974; 35:653–659. [PubMed: 4819867]
- Streeter LA. Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature*. 1976; 259:39–41. [PubMed: 1256541]
- Summerfield Q, Haggard M. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*. 1977; 62:435–448. [PubMed: 886081]

Williams L. The perception of stop consonant voicing by Spanish-English bilinguals. *Perception and Psychophysics*. 1977; 21:289–297.

\$watermark-text

\$watermark-text

\$watermark-text

Fig. 1.

An example of an individual infant's responses during the adaptive staircase testing phase on a specified VOT contrast. Circles indicate test trials and squares indicate probe trials. A plus (+) indicates a correct head-turn response and a minus (-) indicates the absence of a head turn.

Fig. 2. Mean Δ VOT values for infants and adults tested on the staircase procedure with five truncated VOT series. Error bar equals ± 1 SE about the mean.

\$watermark-text

\$watermark-text

\$watermark-text

TABLE 1

Mean VOT Thresholds and ΔVOT Values for Infants and Adults

TARGET	BACKGROUND	INFANTS			ADULTS			ΔVOT	
		Mean	SD	N	Mean	SD	N	Infant	Adult
Full VOT series:									
-70	+70	12.0	23.0	12	33.6	10.4	10	58.0	36.4
+70	-70	29.3	17.4	15	7.4	23.2	10	99.3	77.4
Truncated series straddling English adult boundary:									
0	+70	29.5	4.2	4	42.9	11.4	10	40.5	27.1
+70	0	29.7	13.0	7	17.2	7.4	10	29.7	17.2
-20	+50	4.9	14.0	4	33.2	7.2	10	45.1	16.8
+50	-20	22.3	12.7	3	9.6	12.9	10	42.3	29.6
Truncated series not straddling English boundary:									
-50	+20	-24.3	20.9	4	2.8	16.1	10	44.3	17.2
+20	-50	11.0	4.1	4	-8.9	24.7	10	61.0	41.1
-70	0	-45.5	27.6	2	-42.3	18.2	9	45.5	42.3
0	-70	-5.0	...	1	-44.8	23.0	9	65.0	25.2