Published in final edited form as:
J Acoust Soc Am. 1980 January ; 67(1): 262–270.

Discrimination of relative onset time of two component tones by infants

Peter W. Jusczyk,
Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

David B. Pisoni,
Indiana University, Bloomington, Indiana 47401

Amanda Walley, and
Dalhousie University, Halifax, Nova Scotia B3H 3J1, Canada

Janice Murray
Dalhousie University, Halifax, Nova Scotia B3H 3J1, Canada

Abstract

A great deal of research has focused on the perception of voice onset time (VOT) differences in stop consonants. Yet, the nature of the mechanisms responsible for the perception of these differences is still the subject of much debate. Recently Pisoni [J. Acoust. Soc. Am. **61**, 1352–1361 (1977)] has presented evidence which suggested that the perception of VOT differences by adult listeners may reflect a basic limitation on processing temporal order information by the auditory system. For adults, stimuli with onset differences approximately greater than 20 ms are perceived as successive events (either leading or lagging), while stimuli with onset differences less than about 20 ms are perceived as simultaneous events. Thus, differences in voicing may have an underlying perceptual basis in terms of three well-defined temporal attributes corresponding to leading, lagging, or simultaneous events at onset. The present experiment was carried out to determine whether young infants can discriminate differences in temporal order information in nonspeech signals and whether their discrimination performance parallels the earlier data obtained with adults. Discrimination was measured with the high-amplitude sucking (HAS) procedure. The results indicated that infants can discriminate differences in the relative onset of two events; the pattern of discrimination also suggested the presence of three perceptual categories along this temporal continuum although the precise alignment of these categories differed somewhat from the values found in the earlier study with adults.

INTRODUCTION

A considerable body of psychophysical evidence has accumulated in the last few years on the perception of nonspeech signals that have properties similar to those found in speech (Cutting and Rosner, 1974; Cutting, Rosner, and Foard, 1976; Miller *et al.*, 1976; Pisoni, 1977). The results of these experiments have shown that the mechanisms used in speech perception appear to be constrained in numerous principled ways by the basic capabilities of the auditory system to process incoming sensory information (Searle, Jacobson, and Rayment, 1979; Miller *et al.*, 1977). These findings mesh nicely with the theoretical work of

The final paper was written while the second author was a Guggenheim Fellow at the Research Laboratory of Electronics, M. I. T. An earlier version of this paper was presented at the Biennial Meeting of the Society for Research in Child Development in San Francisco, California on 17 March 1979.

Stevens (1972) who suggests that the constraints imposed on acoustic signals by the auditory system may initially delineate some of the basic kinds of acoustic events and properties that languages have exploited in realizing phonetic distinctions.

During this period of time there has also been a great deal of research dealing with the perception of voice onset time (VOT) in stop consonants by human adults and infants as well as animals such as chinchillas (Kuhl and Miller, 1975; 1978), and monkeys (Morse and Snowdon, 1975; Sinnott *et al.*, 1976; Waters and Wilson, 1976). However, despite the prevalent interest in the perception of VOT, the precise nature of the sensory and perceptual mechanisms responsible for these findings from seemingly diverse organisms is still a matter of some controversy among numerous investigators (Stevens and Klatt, 1974; Lisker, 1975; Miller *et al.*, 1976; Summerfield and Haggard, 1977; Kuhl and Miller, 1978).

Recently, one of us suggested that the perception of the phonetic feature of voicing—a complex set of temporal and spectral events used to distinguish between voiced and voiceless stops—may have its origin in a basic property of the auditory system to respond to differences in the temporal order of events (Pisoni, 1977). Earlier research by Hirsh (1959), and Hirsh and Sherrick (1961), has shown that the auditory system responds differently when two events occur within 20–25 ms of each other than when the relative onsets of the two events are greater than 20–25 ms. Subjects cannot identify the temporal order of two distinct acoustic events when their onsets are separated by less than 20–25 ms—the stimuli are perceived as having simultaneous onsets. However, subjects can identify the temporal order of two events when their onsets differ by more than 20–25 ms, in which ease the events are perceived as occurring successively and ordered in time.

The significance of these earlier findings for speech perception is that voicing distinctions among stop consonants in a number of languages might reflect a basic limitation on the ability of the auditory system to process temporal order information. For the perception of voicing in stop consonants, the time of an occurrence of an event, (i.e., the onset of periodicity) must be judged in relation to the temporal attributes of other articulatory events, (i.e., the release from stop closure). Since these articulatory events, as well as a number of others involved in producing the voicing distinction in stops, are ordered precisely in time, highly distinctive and discriminable changes may only be produced at certain regions along a temporal continuum including the acoustic continuum represented by variations in (VOT). Indeed, Stevens and Klatt (1974) have even remarked that the inventory of phonetic features found in natural languages seems to consist of the presence or absence of sets of acoustic attributes or cues rather than simply continuous changes in a small set of parameters or dimensions, a view that is similar in spirit to the founders of distinctive feature theory (Jakobson, Fant, and Halle, 1952). One such set of distinctive attributes that may be coded by the auditory system could be temporal information about the timing of laryngeal and supralaryngeal events in the production of stop consonants.

To study the underlying perceptual basis of the voicing feature, Pisoni (1977) used a set of nonspeech stimuli differing in the relative onsets of two component tones of different frequencies, a temporal dimension known to be an important acoustic cue to the perception of voicing in stops. Earlier experiments with synthetic speech stimuli had established the importance of the so-called *F*1 "cutback" cue as a perceptual dimension to voicing in stops so there was sufficient justification for focusing on the same temporal variable in these nonspeech stimuli (Liberman, DeLattre, and Cooper, 1958). The results obtained in identification and discrimination experiments with these tone-onset-time (TOT) stimuli were quite similar to the results observed earlier with synthetic speech stimuli differing in VOT (Lisker and Abramson, 1970; Abramson and Lisker, 1970). Subjects were able to consistently identify these nonspeech stimuli into well-defined perceptual categories.

Moreover, discrimination of pairs of these stimuli was very nearly categorical, with performance close to chance for pairs of stimuli selected from within a perceptual category and excellent for pairs of stimuli selected from different perceptual categories. Furthermore, in other experiments involving categorization and temporal order judgments, it was possible to identify a basis for the underlying perceptual categories found with these nonspeech stimuli in terms of whether the acoustic events at stimulus onset were perceived as simultaneous or successive, and if the latter, whether the temporal order of the component events could be identified as leading or lagging. These three properties of stimulus onsets—lead, lag, and simultaneity—have been found to characterize the major differences in voicing among stops in a large number of languages as represented by the VOT dimension (Lisker and Abramson, 1964, 1970). Thus, it seemed likely that these perceptual results with nonspeech stimuli could be used as an account of the perceptual findings obtained with speech stimuli differing in VOT.

The temporal order hypothesis of voicing perception proposed by Pisoni (1977) at that time was also able to account for a seemingly diverse set of findings on the perception of VOT that had been reported in the literature over the last few years. For example, it had been known for some time that cross-language differences exist in the perception of VOT by adults (Lisker and Abramson, 1967). Moreover, a number of perceptual experiments were also carried out on the discrimination of VOT by infants, chinchillas, and monkeys indicating a strong possibility of a psychophysical or sensory basis for the observed discrimination data. These somewhat diverse results could be accommodated by simply postulating a common underlying basis for the discrimination involving a basic constraint on the auditory system's ability to respond to differences in temporal order between two events at onset.

Recent interest in the basic sensory capabilities of the auditory system has also provided additional information about the underlying psychophysical basis of categorical perception, a finding once thought to be unique only to the perception of speech sounds. Several studies have demonstrated that categorical perception is not confined exclusively to the perception of speech signals *per se*, but instead may be a very general characteristic of the way sensory systems respond to changes in one component of a complex stimulus when other properties of the stimulus remain constant (Miller *et al.*, 1976; Pastore, 1976; Pastore *et al.*, 1977). Moreover, the prevalent view that categorical perception of speech was primarily a consequence of identification or labeling brought about through phonetic categorication has now been seriously questioned by the demonstration of marked changes in sensitivity (d′) and bias in the region corresponding to the boundary separating perceptual categories (Wood, 1976). Thus, these results imply that the perceptual categories employed in the phonological systems of languages may have a natural and well-defined basis in terms of what is known about the sensory capacities of the auditory system itself, above and beyond considerations related to the interpretation of these acoustic signals of speech.

Taken together, these recent studies promote the general view that many of the basic functions and mechanisms of the auditory system are used in processing both speech and nonspeech signals alike. While there are, no doubt, important differences in perception between speech and nonspeech signals, there may also be many similarities based on common psychophysical processes that could help to specify the exact sensory and perceptual basis of the acoustic correlates of distinctive features that occur in speech. Such perceptual considerations may also be relevant to explanations of numerous phonetic and phonological processes that seem to occur universally in language (Lieberman, 1976).

In addition to the theoretical interest in the possible sensory and perceptual correlates of distinctive features in speech, the recent findings on the perception of nonspeech signals

differing in relative onset time are also relevant to several well-known findings in perceptual development, particularly the demonstration by Eimas *et al.* (1971) that one-month old infants perceive differences in VOT categorically. These results, as well as a number of other findings with infants, have been interpreted as evidence for the existence of a "speech mode" of perception and the operation of "specialized" perceptual mechanisms for processing speech signals in humans (for a review see Eimas, 1978; Liberman *et al.*, 1967).

In the initial study involving stimuli differing in VOT, Eimas *et al.* (1971) demonstrated that infants could discriminate between two speech sounds selected from across an adult phoneme boundary but could not discriminate two stimuli selected from within the same adult perceptual category even though the acoustic differences between the stimuli were equal, at least in terms of the physical dimension of VOT. These results were quite provocative at the time, suggesting that infants might have access to mechanisms of phonetic categorization at an extremely early age. Moreover, these results were interpreted by Eimas and others as support for the idea that the mechanisms responsible for categorical perception of speech sounds might be specified innately in humans.

One of the most important claims of these early infant experiments on VOT was the assertion that the infants were responding to these speech signals in a "linguistically relevant manner" that involved the phonetic coding of these stimuli into abstract perceptual categories comparable to those observed in adult subjects. An alternative view—that these infants were simply responding to the psychophysical differences between these signals in the absence of explicit phonetic categorization—was proposed by Stevens and Klatt (1974) in light of the results they obtained in several perceptual experiments with adults. These investigators argued that the infants in the Eimas *et al.* experiments were simply responding to the presence or absence of a voiced *F*1 formant transition at onset rather than to VOT *per se*. In a reply to this paper, Lisker (1975) has shown that it is primarily *F*1 onset frequency that adult listeners respond to as a positive cue voicing rather than the *F*1 frequency shift observed by Stevens and Klatt. Summerfield and Haggard (1977) have confirmed and extended Lisker's earlier findings in a series of experiments that systematically varied both spectral and temporal cues to voicing. Although these perceptual experiments have provided useful information about the numerous cues to voicing in stops and their potential interactions, the data were all collected with adult subjects who no doubt had a very long history in mastering English phonology. Thus, the claim that infants are responding to VOT differences on a phonetic basis still remains largely unresolved.

The results of two cross-language experiments using the same stimuli differing in VOT have also provided additional evidence that young infants can discriminate differences in this acoustic dimension. Moreover, the results have been interpreted as support for the claim that infants are sensitive to three primary modes of voicing in stop consonants. In one study, Lasky, Syrdal-Lasky, and Klein (1975) studied 4 to $6\frac{1}{2}$ month-old infants born to Spanish-speaking parents and found evidence suggesting the presence of three voicing categories in their discrimination data. One area of high sensitivity occurred in the region of +20 to +60 ms, which corresponds to the English voiced–voiceless distinction, whereas the other area of high sensitivity occurred in the region between roughly −20 and −60 ms. These discrimination results are interesting because Spanish has only one phoneme boundary separating its voiced and voiceless stops and that boundary does not coincide with either of the two boundaries that Lasky *et al.* inferred from their discrimination data. The apparent discrepancy between the adult and infant data suggests that the infants in this study were probably responding to some set of acoustic attributes or cues in these VOT stimuli independently of their phonetic status or exposure to them in the language learning environment.

In another study Streeter (1976) found that Kikuyu infants also showed evidence of discriminating three categories of voicing for labial stops. Her results are also of some importance, because there are no voicing contrasts for labial stops in Kikuyu, although there are voicing contrasts for stops at other places of articulation in this language. Since this particular contrast was not phonologically distinctive in the adult language, and therefore probably occurred quite infrequently in the language learning environment of these infants, the infants' discrimination of VOT must have been entirely based on the acoustic and psychophysical attributes of the stimuli themselves. This conclusion is strengthened by the fact that the regions of high discriminability found in this study were similar to those obtained in the earlier study by Lasky *et al.* despite the differences between the two languages.

The results of both cross-language investigations of the perception of voicing in young infants, as well as the initial findings of Eimas *et al.*, indicate that young infants can discriminate differences in VOT. However, the underlying basis of the infants' discrimination performance may simply be a consequence of the presence of psychophysically defined regions of high discriminability that exist in the VOT continuum itself rather than processes that involve phonetic categorization or interpretation of these signals as speech. A clear precedent for this notion already exists in an earlier study of infants' perception of nonspeech stimuli conducted by Jusczyk *et al.* (1977). These investigators found evidence that infants' discrimination of sine wave stimuli differing in rise times was categorical. This demonstration that infants display categorical discrimination of nonspeech as well as speech sounds, supports the view that the infants' perceptual behavior in these situations may be the consequence of mechanisms attuned to psychophysical properties in the acoustic signal. Thus, the infants in the previous VOT studies may not have perceived these signals linguistically as Eimas *et al.* have claimed, but instead may have been responding to some complex set of psychophysical properties that separates each of the three primary modes of voicing. One such property of these VOT stimuli may be the relative timing of the component events at stimulus onset.

If the auditory system responds to temporal order information in both speech and nonspeech signals in terms of coding simultaneous and successive events as salient perceptual attributes, we would expect to find that such mechanisms are also present and operative in young infants, given the earlier results on the discrimination of VOT summarized above. Moreover, such an outcome in young infants would be consistent with the nonspeech results of Jusczyk *et al.* (1977) and with predictions based on the nonspeech results obtained with adults by Pisoni (1977). The present experiment was therefore carried out to determine whether infants can discriminate differences in temporal order in nonspeech signals having properties similar to those found in speech. In addition, we were also interested in determining whether the pattern of discrimination along this nonspeech continuum would be comparable to that found earlier in adult subjects.

I. METHOD

A. Procedure

Each infant was tested individually in a small laboratory room. The infant was placed in a reclining chair which faced a rear projection screen approximately 0.5-m away. An image of a man was displayed on the screen for the entire test session. The projection screen was situated just above a loudspeaker through which the test stimuli were played. Each infant sucked on a bind nipple held in place by an experimenter who wore headphones and listened to recorded music throughout the test session. A second experimenter in an adjacent room monitored the apparatus.

The experimental procedure was a modification of the high-amplitude sucking technique devised by Siqueland and DeLucia (1969). For each infant, the high-amplitude sucking criterion and the baseline rate of high-amplitude sucking were established prior to the presentation of any test stimuli. The criterion for high-amplitude sucking was adjusted so as to produce rates of 15–35 sucks/min. After a baseline rate was established, the presentation of stimuli was made contingent upon the rate of high-amplitude sucking. Since the stimuli had a maximum duration of 300 ms and a 750-ms interstimulus interval was used, the maximum stimulus presentation rate was approximately one stimulus per second. If the infant produced a burst of sucking responses with interresponse times of less than 1 s, then each response did not produce one presentation of the stimulus. Rather, the timing apparatus was reset so as to provide continuous auditory feedback for 1 s after the last response of the sucking burst. Use of a progammable logic board ensured that all stimulus presentations were uninterrupted.

The criterion for satiation to the first stimulus was a decrement in sucking rate of 25% or more over 2 consecutive minutes compared to the rate in the immediately preceding minute. At this point the auditory stimulation was changed without interruption by switching channels on the tape recorder. For infants in the experimental conditions, the change resulted in the presentation of a second acoustically different stimulus. For infants in the control condition, the channels on the tape recorder were switched, but no acoustic change occurred since the same signal had been recorded on both channels of the tape. The postshift period lasted for 4 min. The infant's sensitivity to the change in auditory stimulation was inferred from comparisons of response rates of subjects in the experimental and control conditions during the postshift period.

B. Stimuli

The stimuli were two-tone sequences that were generated digitally on a PDP 11/10 computer with a program that permits the specification of the amplitude and frequency of two sinusoids at successive moments in time (Kewley-Port, 1976). These stimuli were similar to ones used in the earlier experiment by Pisoni (1977). A schematic display of the stimuli is shown in Fig. 1. Each stimulus consisted of two tones, a lower one set at 500 Hz and a higher one set at 1500 Hz. The amplitude of the latter was 12 dB lower than the former so that the amplitude relations between the two might parallel those found in a neutral vowel. Both tones were terminated together at the same time. In addition, the duration of the 1500-Hz tone was always held constant at 230 ms. To form the test signals, the duration of the 500-Hz tone was varied systematically in 10-ms steps from 300 to 160 ms across the series of stimuli. Thus, the stimuli could be arranged along a temporal continuum according to the degree to which the onset time of the 500-Hz tone either led or lagged behind that of the 1500-Hz tone. The endpoint values of this (TOT) continuum were −70 ms (in which case the 500-Hz tone leads the 1500-Hz tone by 70 ms) and +70 ms (in which case the 500-Hz tone lags behind the 1500-Hz tone by 70 ms). Digitized waveforms of the stimuli were converted into analog form via a D-A converter, low-pass filtered and then output to a Crown (model 822) tape recorder in order to prepare the two-channel audiotapes employed in this experiment.[1]

C. Design

Each infant was seen for one experimental session. Sixteen infants were assigned randomly to each of six test groups. One of these groups (group I) served as a control condition in

[1]Since we had resynthesized our stimuli, and thus they were not the identical waveforms used in the Pisoni (1977) study, we, of course, tested a group of adults on these new tokens. The data from the adult subjects were consistent with the original Pisoni study. For this reason, we saw no need to include the data of these additional adult subjects in our present paper.

which subjects were randomly assigned to one of the 11 two-tone stimuli for the entire session (e.g., +70 vs +70). Subjects in the remaining five test groups were presented with pairs of stimuli differing in TOT values by 30 ms. The stimulus pairs were chosen so as to permit comparisons of the discriminability of both between-category and within-category contrasts in TOT. The stimulus values selected for each experimental group are displayed in Table I Based on the results of Pisoni's (1977) earlier experiment with adults, groups II, IV, and VI were presented "within category" contrasts of TOT stimuli. For group II, both stimuli were chosen from the "lead category." For group IV, all stimuli were selected from the "simultaneous category." For group VI, stimuli from the "lag category" were employed. In contrast, subjects in groups III and V received stimulus pairings selected from different TOT categories. In the case of group III, one member of each stimulus pair was selected from the "lead category" (i.e., −40 or −30 ms), and the other from the "simultaneous category" (i.e., −10 or 0 ms). For group V, the pairings were between the "simultaneous category" (i.e., 0 or +10 ms) and the "lag category" (i.e., +30 or +40 ms). The presentation order of stimuli was always counterbalanced across subjects for each of the groups.

On the basis of these stimulus pairings selected from the earlier adult data, we expected that infants would discriminate only "between category" constrasts that were selected from opposite sides of either the −20-ms boundary (i.e., lead versus stimultaneous) or the +20-ms boundary (i.e., simultaneous versus lag). In contrast, we also expected that infants would not discriminate any of the "within category" contrasts that were selected from the same adult perceptual category.

D. Apparatus

A blind nipple was connected to a Grass PT5 volumetric pressure transducer which, in turn, was coupled to a type DMP-4A physiograph. A Schmitt trigger provided a digital output of criterial high-amplitude sucking responses. Additional equipment included a Teac 3340 tape recorder, a Kenwood (KA-3500) power amplifier, an Ads 200 loudspeaker, a Grason–Stadler (model #1200) programmable logic board, a power supply, two relays, a counter, and a physiograph dc preamplifier. Each criterial response activated a timer on the logic board for a 1-s period or restarted the period. Auditory stimulation at a level of 75 ± 2 dB (A) SPL (approximately 15-dB above the background noise level caused by the ventilation system) was available whenever the timer was in an active state. By using the logic board to monitor the auditory signals on the tape recorder, it was possible to ensure that the timer was never activated in the middle of a TOT stimulus.

E. Subjects

The subjects were 96 infants, 49 males and 47 females. Mean age was 10.0 weeks (range: 7–13 weeks). In order to obtain 96 infants for the study, it was necessary to test 231. Subjects were excluded from this study for the following reasons: crying (33%) or falling asleep (33%) prior to shift, ceasing to suck during the course of the experiment (i.e., 2 consecutive minutes with less than 2 sucks/min) (10%), failure to maintain a minimal criterial sucking rate of 15 responses/min during the satiation period (7%), equipment failure (6%), experimenter error (6%), and miscellaneous (3%).

II. RESULTS

Figure 2 displays the mean number of high-amplitude sucking responses as a function of minutes and experimental groups. For purposes of statistical comparison, we examined each subject's rate of sucking during five intervals: baseline minute, third minute before shift, average of minutes 1 and 2 before shift, average of minutes 1 and 2 after shift, and average of all 4 min after shift. Difference scores were then calculated for each subject for each of

the following rate comparisons: (1) acquisition of the sucking response—third minute before shift less baseline; (2) satiation—third minute before shift less the average of the last 2 min before shift; (3) release from satiation—average of first 2 min after shift less the average of the last 2 min before shift; (4) release from satiation for the full 4 min—average of 4 min after shift less the average of the last 2 min before shift.

In each of groups III, IV, and V, half of the subjects were tested on one stimulus pair and half of the subjects on another. For each of these groups, Randomization tests for independent samples (Siegel, 1956) were used to determine whether the data from the two kinds of stimulus pairs could be pooled for further analysis. Since no significant differences between stimulus pairs emerged for any of these groups, the data were combined for further statistical treatment.

As is usually the case in studies employing the HAS procedure, subjects in all sessions acquired the conditioned high-amplitude sucking response and satiated to the first stimulus prior to shift. An indication of the mean change in response rate during the postshift period for each of the six groups is provided in Table II. Randomization tests for independent samples (Siegel, 1956) were employed to assess performance during the postshift periods. Postshift performance of each of the experimental groups (II, III, IV, V, and VI) was compared to that of the control group (I) for both the first 2-min and the full 4-min periods. These tests indicated that the only reliable ($p < 0.05$, one tailed) differences occurred between the control group and two of the "within category" groups (II and VI) for both the first 2-min and the full 4-min periods. Neither of the two "between category" groups (III and V) nor the other "within category" group (IV) performed reliably differently than the control group. Although the mean change in response rate after shift was somewhat smaller for group II than for group VI, subsequent comparisons of these two groups by means of randomization tests for independent samples, indicated that no reliable differences existed between them for either the first 2-min or full 4-min periods. Thus, as was the case for adult subjects, infants were capable of discriminating differences in the relative onset of two events. Moreover, the pattern of the discrimination data suggests the presence of three perceptual categories along this temporal continuum. However, the regions of highest discriminability observed in these infants apparently differ somewhat from our initial expectations based on the adult discrimination data.

III. DISCUSSION

The overall results of the present study are generally consistent with our predictions based on the temporal order hypothesis of voicing perception. We have shown that infants are capable of discriminating differences in temporal order information in nonspeech signals having speech-like properties. The pattern of results indicates the presence of three well-defined perceptual categories along this temporal continuum, corresponding to leading, simultaneous, and lagging events. Thus, in general, these findings provide additional support for the claim that the underlying basis of the perception of VOT in stop consonants reflects a basic limitation of the auditory system to respond to differences in temporal order at stimulus onset. Therefore, the auditory system of young infants may be predisposed, in some sense, to respond to salient and well-defined properties of acoustic signals that represent the acoustic correlates of the distinctive features of speech. One such salient acoustic property appears to be the relative timing of events at stimulus onset, corresponding, in the case of voicing perception, to the temporal ordering of laryngeal and supralaryngeal events, a nearly universal property of all languages.

Although the major findings of the present study demonstrate that young infants can discriminate relatively small differences (i.e., 30 ms) in temporal order information, the

specific details of the results differ somewhat from those anticipated at the outset. Specifically, we predicted that the infants would be able to discriminate only the stimulus contrasts that were selected from opposite sides of either the −20 or +20 ms boundary, the value assumed to represent the threshold for temporal order in adults. However, the present results indicated that infants discriminate only the −70/−40 ms lead and the +40/+70 ms lag contrasts, stimulus pairings that were initially assumed to represent "within category" comparisons. These findings indicate that infants' sensitivity to temporal order information is shifted slightly toward larger stimulus values on this test continuum. It is unlikely that these results are due to some artifact in the specific stimulus contrasts employed or details of the HAS measurement procedure since the shifts in discrimination occurred for both lead and lag contrasts. Moreover, these shifts were displaced in opposite directions in each case toward larger stimulus differences in temporal order. Nevertheless, it should be noted that the small discrepancy between the adult and infant discrimination data could be simply a consequence of the degree of imprecision that is present in the HAS procedure itself.[2] Discrimination data collected in this paradigm does not permit an exact specification of the infants' sensitivity or threshold. Rather, these discrimination measures provide only a rough indication of the range over which large differences in sensitivity might be observed. Thus, the exact values obtained in any HAS discrimination study must be interpreted with care and direct comparisons between adults and infants made with some caution.[3]

The present investigation was undertaken not only to determine infants' responsiveness to temporal order information in nonspeech signals, but also to examine whether temporal order information might serve as the underlying basis for the perception of VOT in speech stimuli. The previous findings of Pisoni (1977) with adults indicated a close correspondence between identification and discrimination of temporal order information in nonspeech signals and suggested a possible account of the perception of speech signals differing in VOT. However, the present results revealed a slight divergence, at least for infants, in the precise location of the region of highest discriminability for the nonspeech stimuli. While we would want to interpret this discrepancy cautiously, the results raise the possibility that temporal order *per se* may not be the only property that young infants respond to in discriminating VOT. As mentioned earlier, Stevens and Klatt (1974) have suggested that infants could be responding to the presence or absence of an *F*1 transition and not VOT. Although Lisker (1975) has questioned the importance claimed for this particular acoustic cue in controlling adults' perception of voicing differences, it may be the case that the presence or absence of a rapid spectrum change at onset serves as one of several salient properties that infants initially respond to in discriminating stop consonants. We might speculate further that in the course of perceptual development, the *F*1 transition information is combined in some way with other acoustic cues such as those related to processing temporal order information and that these complex or integrated cues gradually assume a larger and larger role in controlling the perception of voicing as the child's perceptual system develops. It should be noted, however, that any account of VOT perception based on

[2]A reviewer has suggested two possible reasons for the discrepancy which we found between the infant and adult discrimination data. The first suggestion is that differences in listening conditions—the adults were tested using headphones and the infants under free field conditions—may account for the discrepancy. While this may have indeed been a factor in the present results, it would also have to be true for almost every other infant–adult comparison with respect to speech perception since these same listening conditions have held for almost all previous studies. Thus, if there were a systematic effect due to differences in listening conditions, then one might expect there to be a similar discrepancy in infant-adult comparisons involving VOT. In fact, there is no evidence of such a discrepancy in the VOT studies, so that it is unlikely that differences in listening conditions is the source of the present results.
The reviewer's second suggestion was that low-level background noise from the ventilating system may have masked the lower tone, thereby requiring a longer TOT before the two onsets were decidedly nonsimultaneous. Again it seems unlikely that this could account for the results of the present study, since it was conducted under the same conditions as the original Eimas *et al.* study (which also had a low level of background noise present due to a ventilating system).

[3]The use of more refined measures such as those employed by Aslin *et al.* (1979) may permit a more exact specification of the infant's threshold.

the *F*1 transition cue is incomplete since it can only be invoked to deal with the discrimination of differences in the lag region of the stimulus continuum where the duration of the *F*1 transition varies inversely with VOT.[4]

With regard to accounting for the cross-language data on infants' perception of VOT, it may be necessary to assume that infants first respond to speech signals on the basis of the sensory or psychophysical properties of the stimuli without any subsequent phonetic coding or interpretation. Experience in the language-learning environment would enable infants to utilize other acoustic attributes that might be prominent in phonetic environments defined by the phonological constraints of the specific language. Differential weighting might then be assigned to these acoustic cues according to their salience in marking a distinctive contrast in the language or particular dialect. Thus, a change in the relative weightings of the acoustic cues for a particular phonetic contrast could shift the region of sensitivity along some selected stimulus continuum. According to this view, infants' reliance on a common set of psychophysical properties could be responsible for the apparent universality of VOT discrimination by infants from different language-learning environments. Differences in the relative weights assigned to the various acoustic attributes to voicing could account for the cross-language differences observed in adult speakers. Questions surrounding perceptual tuning by environmental input and a more detailed discussion of the developmental course of speech perception in infants are taken up in a recent chapter by Aslin and Pisoni (1978).

An alternative to the psychophysical account summarized above is one that assumes that the infant's discrimination of VOT is, in fact, based on some form of phonetic coding or interpretation of the stimuli as speech, a view first proposed by Eimas *et al.* (1971). Given the current procedures available for studying speech perception in infants, it is extremely difficult to determine whether an infant's perceptual behavior is controlled entirely by the psychophysical or phonetic properties of the stimuli. Moreover, there is little known at this time about how these two levels of perceptual analysis interact during the course of perceptual development. One promising avenue currently open to investigators is to search for correspondences in discrimination between speech and comparable nonspeech signals. When such correspondences can be found, they would strongly imply some sensory or psychophysical basis to the perception of a particular set of acoustic correlates to a distinctive feature. For example, in the earlier study of Jusczyk *et al.* (1977) on the discrimination of rise time by infants, evidence was found with nonspeech stimuli indicating that infants can discriminate differences in tempo of frequency change. This result was subsequently verified for speech stimuli by Hillenbrand, Minifie, and Edwards (1977) who reported that infants can discriminate the differences between [ba] and [wa]. Thus, one could account for these results by a common underlying factor involving the detection of rate of frequency change.

While one may still be forced into accepting at least some type of phonetic coding account for certain aspects of the infant's perceptual behavior, it may be difficult to reconcile this position with the results of recent comparative studies that have examined the perception of speech signals by animals (Kuhl and Miller, 1975; 1978). In the absence of alternative proposals, the most parsimonious explanation for the parallels observed in the perception of speech signals by animals and humans is one that also assumes a common underlying basis for the two sets of results in terms of some general psychophysical process. At the present time, there is strong evidence that both humans and chinchillas respond in somewhat similar ways to VOT, a temporal contrast. However, it remains to be seen in future work if similar

[4]It is worth noting here that discrimination performance in the lag region of the VOT continuum tends to be better than that for the lead region (e.g., Abramson and Lisker, 1970; Aslin and Pisoni, 1978). We might speculate that the addition of the *F*1 transition cue to the lag region, but not the lead region, may help to account for the better discriminability of contrasts in the lag region.

evidence can be adduced for other phonetic contrasts that have well-defined acoustic properties. A psychophysically based explanation of the infant's ability to discriminate the acoustic correlates of place or manner may be somewhat more difficult to develop as the comparable nonspeech experiments examining the possible underlying perceptual properties that define these categories have not yet been conducted (see Walley and Aslin, 1979). Nevertheless, an important goal of future research will be to specify more precisely the sensory and perceptual correlates of the distinctive features in speech and how the abilities to perceive these salient properties develop in the young infant.

In summary, the present study has demonstrated that young infants can discriminate differences in temporal order information in nonspeech signals. The pattern of results suggests the presence of three well-defined perceptual categories corresponding to leading, simultaneous, and lagging temporal events. Although the overall results of this study were similar to earlier work obtained with adults, several differences were observed in the precise location of the perceptual categories that could be inferred from the infant discrimination data. Despite these differences, the main findings provide some additional support for the hypothesis that the perception of VOT, one of the major cues to voicing in stop consonants, involves the perception of the relative temporal order of the component events at stimulus onset.

Acknowledgments

This research was supported, in part by an NSERC grant (A0282) to the first author and by NIH grants NS-12179-04 and HD-11915-01 and NIMH grant MH-24027-04 to Indiana University. We also wish to acknowledge our appreciation to the staff and administrators of the Grace Maternity Hospital for their support and assistance.

References

Aslin, RN.; Hennessey, B.; Pisoni, DB.; Perey, AJ. Individual infant's discrimination of voice onset time: Evidence for three modes of voicing. Paper presented at the Biennial Meeting of the Society for Research in Child Development; San Francisco, California. 1979.

Aslin, RN.; Pisoni, DB. Some developmental processes in speech perception. Paper presented at N.I.C.H.D. Conference on Child Phonology: Perception, Production and Deviation; Bethesda, Maryland. 1978.

Abramson, A.; Lisker, L. Discriminability along the voicing continuum: Cross-language tests. Proceedings of the 6th International Congress of Phonetic Sciences; Academia, Prague. 1970.

Cutting JE, Rosner BS, Foard CR. Perceptual categories for musiclike sounds: Implications for theories of speech perception. Q. J. Exp. Psychol. 1976; 28:361–378. [PubMed: 1005649]

Cutting JE, Rosner BS. Categories and boundaries in speech and music. Percept. Psychophys. 1974; 16:564–571.

Eimas, PD. Developmental aspects of speech perception. In: Held, R.; Leibowitz, H.; Teuber, HL., editors. Handbook of Sensory Physiology: Perception. New York: Springer-Verlag; 1978.

Eimas PD, Siqueland ER, Jusczyk P, Vigorito J. Speech perception in infants. Science. 1971; 171:303–306. [PubMed: 5538846]

Hillenbrand, J.; Minifie, FD.; Edwards, TJ. Tempo of frequency change as a cue in speech sound discrimination by infants. Paper presented at the Biennial Meeting of the Society for Research in Child Development; New Orleans. 1977.

Hirsh IJ. Auditory perception of temporal order. J. Acoust. Soc. Am. 1959; 31:759–767.

Hirsh IJ, Sherrick CE. Perceived order in different sense modalities. J. Exp. Psychol. 1961; 62:423–432. [PubMed: 13907740]

Jakobson, R.; Fant, G.; Halle, M. Preliminaries to Speech Analysis. Cambridge: M. I. T.; 1952.

Jusczyk PW, Rosner BS, Cutting JE, Foard CF, Smith LB. Categorical perception of nonspeech sounds by 2-month old infants. Percept. Psychophys. 1977; 21:50–54.

Kewley-Port, D. Research on Speech Perception: Progress Report #3. Bloomington, IN: Department of Psychology, Indiana University; 1976. A complex-tone generating program.

Kuhl P, Miller JD. Speech perception by the chinchilla: Voiced–voiceless distinction in alveolar-plosive consonants. Science. 1975; 190:69–72. [PubMed: 1166301]

Kuhl PK, Miller JD. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. J. Acoust. Soc. Am. 1978; 63:905–917. [PubMed: 670558]

Lasky RE, Syrdal-Lasky A, Klein RE. VOT discrimination by four to six and a half month old infants from Spanish environments. J. Exp. Child. Psychol. 1975; 20:213–225.

Liberman AM, Cooper FS, Shankweiler DP, Studdeft-Kennedy M. Perception of the Speech Code. Psychol. Rev. 1967; 74:431–461. [PubMed: 4170865]

Liberman AM, DeLattre PC, Cooper FS. Some cues for the distinction betweeen voiced and voiceless stops in initial position. Lang. Speech. 1958; 1:153–167.

Lieberman P. Phonetic features and physiology: a reappraisal. J. Phonetics. 1976; 4:91–112.

Lisker L. Is it VOT or a first-formant transition detector? J. Acoust. Soc. Am. 1975; 57:1547–1551. [PubMed: 1141504]

Lisker L, Abramson A. A cross language study of voicing in initial stops: Acoustical measurements. Word. 1964; 20:384–422.

Lisker, L.; Abramson, A. The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences; Academia, Prague. 1970.

Miller JD, Engebretson AM, Spenner BF, Cox JR. Preliminary analyses of speech sounds with a digital model of the ear. J. Acoust. Soc. Am. Suppl. 1. 1977; 62(S1):13.

Miller JD, Wier L, Pastore R, Kelly W, Dooling K. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. J. Acoust. Soc. Am. 1976; 60:410–417. [PubMed: 993463]

Morse P, Snowdon C. An investigation of categorical speech discrimination by rhesus monkeys. Percept. Psychophys. 1975; 17:9–16.

Pastore, RE. Categorical perception: A critical reevaluation. In: Hirsch, SK.; Eldredge, DH.; Hirsh, IJ.; Silverman, SR., editors. Hearing and Davis: Essays Honoring Hallowell Davis. St. Louis: Washington University; 1976. p. 253-264.

Pastore RE, Ahroon WA, Buffuto KJ, Friedman CJ, Puleo JS, Fink EA. Common factor model of categorical perception. J. Exp. Psychol. 1977; 4:686–696.

Pisoni DB. Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in steps. J. Acoust. Soc. Am. 1977; 61:1352–1361. [PubMed: 881488]

Searle CL, Jacobson JZ, Rayment SG. Phoneme recognition based on human audition. J. Acoust. Soc. Am. 1979; 65:799–809. [PubMed: 447910]

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw–Hill; 1956.

Sinnott J, Beecher M, Moody D, Stebbins W. Speech sound discrimination by humans and monkeys. J. Acoust. Soc. Am. 1976; 55:653–659.

Siqueland ER, DeLucia CA. Visual reinforcement of nonnutritive sucking in human infants. Science. 1969; 165:1144–1146. [PubMed: 5801599]

Stevens, KN. The quantal nature of speech. In: David, EE., Jr; Denes, PB., editors. Human Communication: A unified view. New York: McGraw–Hill; 1972.

Stevens KN, Klatt DH. Role of formant transitions in the voiced–voiceless distinction for steps. J. Acoust. Soc. Am. 1974; 55:653–659. [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 QS, Haggard M. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. J. Acoust. Soc. Am. 1977; 62:435–448. [PubMed: 886081]

Walley A, Aslin RN. Infants' discrimination of full and partial cues to place of articulation in stop consonants. 1979 (unpublished).

Waters RS, Wilson WA. Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. Percept. Psychophys. 1976; 19:285–289.

Wood CC. Discriminability, response bias, and phoneme categories in discriminations of voice onset time. J. Acoust. Soc. Am. 1976; 60:1381–1389. [PubMed: 1010890]

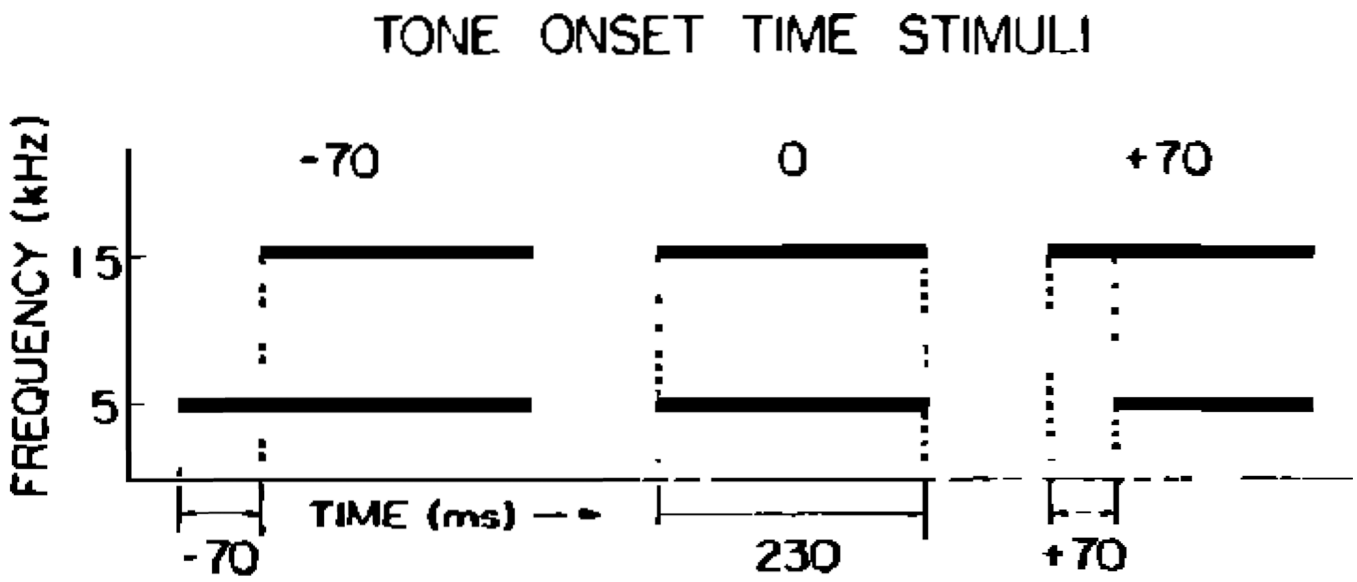

FIG. 1.
Schematic representations of three stimuli differing in relative onset time: leading (−70 ms), simultaneous (0 ms), and lagging (+ 70 ms).

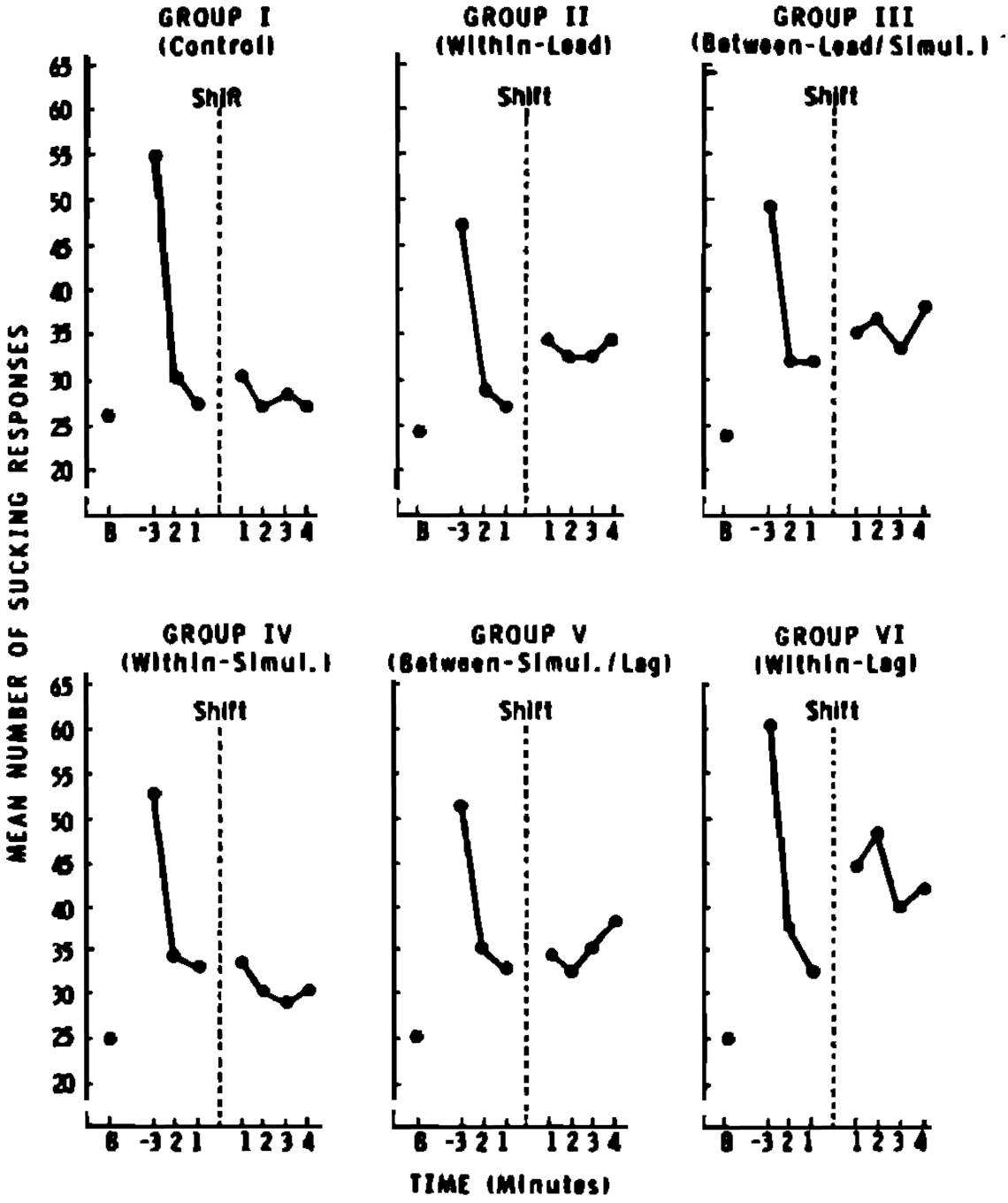

FIG. 2.

Mean number of high-amplitude sucking responses as a function of time and experimental group. Time is measured with reference to the moment of the stimulus shift, marked by the vertical dashed line. The baseline rate of sucking is indicated by the letter B.

TABLE I

Design and breakdown of experimental groups.

Group	Type of contrast	Stimulus pairings
I	Control (e.g., $lead_1$ versus $lead_1$)	−70 vs −70,
		−40 vs −40, etc.
II	Within ($lead_1$ versus $lead_2$)	−70 vs −40
III	Between (lead versus simul.)	−40 vs −10 (8 Ss)
		−30 vs 0 (8 Ss)
IV	Within ($simul._1$ versus $simul._2$)	−20 vs +10 (8 Ss)
		−10 vs +20 (8 Ss)
V	Between (simul. versus lag)	0 vs +30 (8 Ss)
		+10 vs +40 (8 Ss)
VI	Within (lag_1 versus lag_2)	+40 vs +70

TABLE II

Mean change in response rate after shift.

	Release from satiation (minutes after shift)	
Group	**First 2**	**Full 4**
I (Control)	−0.03	−0.34
II (Within-lead)	6.00 [a]	6.28 [a]
III (Between-lead/simul.)	3.63	3.73
IV (Within-simul.)	−0.88	−2.02
V (Between-lag/simul.)	2.28	2.55
VI (Within-lag)	11.50 [a]	8.97 [a]

[a]Indicates a reliable difference ($p < 0.05$ or better) when compared to the performance of control subjects for the same period.