RESEARCH ON SPEECH PERCEPTION

Progress Report No. 4

January 1977 - September 1978

David B. Pisoni

Principal Investigator

Department of Psychology

Indiana University

Bloomington, Indiana 47405

Supported by:

Department of Health, Education, and Welfare
U.S. Public Health Service

National Institute of Mental Health
Grant No. MH-24027-04

and

National Institutes of Health
Grant No. NS-12179-03

CONTENTS

INTRODUCTION

This is the fourth report of research activities on speech processing conducted in the Department of Psychology at Indiana University. As with our previous progress reports, our main goal has been to summarize our various research activities over the past year and make them available to interested colleagues in the field. Some of the papers contained in this report are extended manuscripts that have been prepared for formal publication as journal articles or book chapters. Other papers are simply short reports of research presented at professional meetings during the past year or brief progress reports on the status of on-going research projects in the laboratory. We also have included new information on instrumentation developments and software support when we think this information would be of interest or help to other colleagues.

We decided to issue a progress report of our research activities primarily because of the lag in journal publications and the resulting delay in the dissemination of new information and research findings. We are, or course, very interested in following the work of other colleagues who are carrying out research on speech and, therefore, would be most grateful if you would send us copies of your own recent reprints, preprints and progress reports as they become available so that we can keep up with your latest findings. Please address all correspondence to:

Professor David B. Pisoni
Department of Psychology
Indiana University
Bloomington, Indiana 47405
U.S.A.

EXTENDED MANUSCRIPTS

Adaptation of the Relative Onset Time of Two-Component Tones

David B. Pisoni

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Short Title: Adaptation of Two-Component Tones

Abstract

The results of three selective adaptation experiments employing nonspeech signals that differed in temporal onset are reported. In one experiment, adaptation effects were observed when both the adapting and test stimuli were selected from the same nonspeech test continuum. This result was interpreted as evidence for selective processing of temporal order information in nonspeech signals. Two additional experiments tested for the presence of cross-series adaptation effects from speech-to-nonspeech and then from nonspeech-to-speech. Both experiments failed to show any evidence of cross-series adaptation effects implying a possible dissociation between perceptual classes in processing temporal order information. Despite the absence of cross-series effects, it is argued that the ability of the auditory system to process temporal order information may still provide a possible basis for explaining the perception of voicing in stops that differ in VOT. The results of the present experiments taken together with earlier findings on the perception of temporal onset in nonspeech signals were viewed as an example of the way languages have exploited the basic sensory capabilities of the auditory system to signal phonetic differences.

Adaptation of the Relative Onset Time of Two-Component Tones*

David B. Pisoni
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

In the last few years there has been a great deal of interest in the mechanisms thought to underlie the perception of speech sounds, particularly as they may be used in processing distinctions involving the phonetic features of voicing, place and manner in consonants. Beginning with the initial report by Eimas and Corbit (1973), numerous studies have used the selective adaptation technique to support the hypothesis that complex feature detectors mediate speech perception (see Eimas and Miller, 1978; Ades, 1976; and Cooper, 1975 for reviews). These detector mechanisms were thought to be narrowly tuned to specific acoustic properties or attributes of the speech signal that could be related in some straightforward way to the features and categories used in spoken language. Other studies have employed this procedure to reveal the selectivity of these detector mechanisms for a specific range or subrange of stimuli having a property or attribute in common (Miller, 1975). Such efforts have also been aimed at specifying the level or levels of perceptual analysis that were susceptible to adaptation under the assumption that such perceptual effects would therefore define

the primary channels used by the perceptual system in recognizing phonetic segments and features (Miller, 1977; Sawusch, 1977a,b). Although it has been convenient to describe a number of aspects of the early stages of speech perception by appeal to the existence and operation of specialized feature detectors analogous to those found in the visual system, there is currently little consensus among investigators as to the exact nature of these detectors, their structural arrangement or levels of the perceptual system that are being tapped by the adaptation technique itself. As a consequence, the usefulness of this analogy has been called into question as the number of hypothetical detectors and levels of perceptual analysis proliferate with each newly published study (Simon and Studdert Kennedy, 1978; Diehl, Elman & Buchwald, 1978; Elman, 1979; Remez, 1979). Nevertheless, the adaptation technique has been helpful in detailing some aspects of the perceptual system's selectivity. In addition, the immunity of certain stimulus properties from contingent adaptation (see Ades, 1976 for a review) offers the interesting suggestion that the perceptual system may not employ the same analytic categories as do linguists and psychologists.

As additional psychophysical evidence accumulates on the perception of nonspeech signals having properties similar to those found in speech, it has also become increasingly apparent that mechanisms used in speech perception are constrained in numerous ways by the basic capabilities of the auditory system

(Divenyi & Sachs, 1978; Searle, Jacobson and Rayment, 1979; Miller, Engebretson, Spenner and Cox, 1977). The constraints imposed on acoustic signals by the auditory system have been assumed to delineate some of the basic kinds of acoustic events and properties that languages have exploited in realizing phonetic distinctions (Stevens, 1972). To cite one example of this approach, Pisoni (1977) has suggested recently that the phonetic feature of voicing-- a complex set of temporal and spectral events used in distinguishing voiced from voiceless stop consonants, may have its origin in a basic property of the auditory system to respond to differences in the temporal order of events. The auditory system responds differently when two events occur within 20-25 msec. of each other than when the relative onsets of two events are greater than 20-25 msec. (Hirsh, 1959). Subjects cannot identify the temporal order of two distinct acoustic events when their onsets are separated by less than 20-25 msec.-- 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 msec. (Hirsh, 1959; Hirsh and Sherrick, 1961). In this case, the two events are perceived as occurring successively and ordered in time.

The present report is a continuation and extension of earlier work on the perception of temporal order information and its potential role in signalling voicing differences in stop

consonants (see Pisoni, 1977). The results of these experiments with nonspeech signals provided an initial basis for the view that the phonological categories used to realize 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 example, in the case of the voicing feature in stop consonants, the time of an occurrence of an event, (i.e., the onset of voicing) must be judged in temporal relation to other articulatory events, (i.e., the release from stop closure). The fact that these articulatory events occur in precise temporal sequence implies that potentially distinctive and highly discriminable changes will be produced at only certain regions along a temporal continuum such as the acoustic dimension represented by voice onset time (VOT).

To study the perceptual basis of the voicing feature, Pisoni (1977) generated 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 stop consonant. Earlier experiments with synthetic speech stimuli by Liberman, Delattre & Cooper(1958) had established the importance of the F1 "cutback" as a perceptual cue to voicing in stops so there was good reason for focusing on the same temporal variable in the nonspeech stimuli.

The results obtained in identification and discrimination experiments with these nonspeech stimuli were quite similar to those observed with synthetic speech stimuli differing in VOT (Pisoni, 1977). After familiarity and some minimal training, subjects were able to consistently identify the nonspeech stimuli into well-defined perceptual categories. In addition, discrimination of pairs of these stimuli was very nearly categorical; performance was 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 it was possible to identify the basis for the underlying perceptual categories in these nonspeech stimuli in terms of whether the acoustic events at stimulus onset were perceived as simultaneous or successive and, if successive, whether the temporal order of the component events could be identified as leading or lagging. These three properties, leading, lagging and simultaneity also 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). Thus, it seemed likely that these perceptual results with nonspeech speech stimuli could offer an account of the perceptual findings obtained with synthetic speech stimuli differing in VOT. In addition, the temporal order hypothesis of voicing perception suggested at that time was also able to account for a seemingly diverse set of findings on the

perception of VOT reported in the literature. Cross-language differences observed in the perception of VOT by adults as well as a number of perceptual experiments with infants and chinchillas could now be rationalized by simply postulating a common underlying basis for the discrimination involving a basic constraint on the auditory systems' ability to resolve differences in temporal order between two events.

Although the results of these initial nonspeech experiments suggested that the perceptual categories underlying voicing distinctions in stops might depend on the extraction of temporal order information, the precise mechanism responsible for detecting differences in the relative timing between two events remained unspecified at the time. If the auditory system responds differently to simultaneous and successive events at stimulus onset as our previous results indicated, it should be possible to gain some additional information about how the auditory system preserves these salient properties by means of perceptual adaptation techniques.

The results of the earliest adaptation experiments on voicing (Eimas and Corbit; 1973) were, in fact, interpreted as support for the operation of detectors in speech perception that were tuned specifically to differences in VOT, the temporal dimension distinguishing voiced and voiceless stops (see also Miller, 1977). However, at that time, it was assumed that the

detectors used to process VOT information were specific to processing speech signals since several control conditions involving the use of nonspeech adaptors failed to produce any systematic cross-series adaptation effects on speech stimuli (Eimas, Cooper and Corbit, 1973). It seems equally possible, and perhaps even quite likely, that the auditory system contains property detecting mechanisms of the kind that encode temporal order information in both speech and nonspeech signals. Eimas et al. may simply have failed to identify a more general timing mechanism because of their reliance on speech signals as test stimuli.

The first experiment reported below was therefore specifically carried out to determine whether the relative onset time between two tones in a set of nonspeech signals would also show perceptual adaptation in a manner analogous to that reported earlier for speech stimuli differing in VOT. Two additional experiments were conducted with these nonspeech stimuli to gain additional information about the selectivity of the perceptual mechanism in processing temporal order information and its susceptibility to adaptation in cross-series tests. Such cross-series tests between speech and nonspeech signals could provide one way of determining whether the same general timing mechanism is used by the perceptual system in processing both types of signals. If such cross-series adaptation effects could be obtained with nonspeech stimuli differing in relative onset

time, the results would not only provide additional support for the temporal order hypothesis of voicing perception summarized earlier but would also establish the existence and operation of a somewhat more general timing mechanism. Such a timing mechanism would respond to temporal onsets in the auditory system regardless of perceptual class; that is, whether the signal is speech or nonspeech. Moreover, this general timing mechanism would, of course, no doubt be very closely tied to what is already known about the basic capabilities of the auditory system to respond to differences in temporal order.

Experiment I

In this experiment, subjects were first trained to identify stimuli selected from a nonspeech auditory continuum by means of a training procedure developed in our earlier studies (Pisoni, 1977). After completing the training and identification testing, subjects who met a strict predetermined criterion were asked to return for two additional sessions in which identification tests were carried out under baseline and adaptation conditions.

Method

Subjects. Twenty-four volunteers served as subjects. They were recruited by means of an advertisement in the student newspaper and were paid at a base rate of $2.00 per hour plus whatever they earned during the initial training phase of the

experiment. All subjects were right-handed native speakers of English who reported no history of a hearing or speech disorder at the time of testing.

Stimuli. The stimuli consisted of the same eleven two-tone nonspeech patterns that were used in the previous experiments reported by Pisoni (1977). These were generated with a computer program that permitted control over the amplitude and frequency of two sinusoids as a function of time. Schematic representations of the time course of these signals are displayed in Figure 1.

Insert Figure 1 about here

The frequency of the lower tone was set at 500 Hz. while the frequency of the higher tone was set at 1500 Hz. The amplitude of the 1500 Hz. tone was adjusted to be 12dB lower than the 500 Hz. tone so as to approximate the amplitude relations observed in natural speech for a neutral vowel. The stimuli differed in terms of the temporal onset of the lower tone relative to the higher tone. As shown in Figure 1, for the -50 msec. stimulus, the lower tone leads the higher one by 50 msec, for the 0 msec stimulus both tones were simultaneous at onset while for the +50 msec. stimulus the lower tone lags the higher tone by 50 msec. Both component frequencies terminated together at stimulus

offset. The duration of the 1500 Hz. tone was always fixed at 230 msec. in all of the stimuli while the duration of the 500 Hz. tone was varied to produce the test stimuli. All the remaining intermediate values differing in 10 msec. steps from -50 msec. through +50 msec. were also generated to form a complete continuum. The eleven test stimuli were generated on a PDP-9 computer at M.I.T. where they were recorded on audio tape and then later digitized via an A-D converter on a PDP-11 computer in the Speech Perception Laboratory at Indiana University.

Procedure. All experimental events involving the presentation of stimuli, collection of responses from subjects and delivery of feedback were controlled on-line by the PDP-11 computer. The digitized waveforms were output via a D-A converter, low-pass filtered and then presented to subjects through matched and calibrated Telephonics (TDH-39) headphones. The stimuli were presented at a comfortable listening level of about 80 dB (re: 0.0002 dynes/cm) throughout the experimental sessions. Testing was carried out in a quiet room equipped with six individual cubicles.

The present experiment consisted of three 1-hour testing sessions conducted on separate days. All subjects were run in small groups which received the same stimulus conditions in a particular testing session. The first day was used for training, testing and for the selection of subjects. On the second and third days subjects identified these stimuli under both baseline and adapted conditions.

In the initial shaping sessions, subjects were first presented with the endpoint stimuli, -50 and +50 msec., in a fixed sequence for 160 trials so that the differences between the two stimuli could be discriminated easily. This was followed by a training session in which the same two endpoint stimuli were presented in a random order for 160 trials. Subjects were told to learn which of two response buttons was associated with each sound. Immediate feedback was provided after the presentation of each stimulus indicating the correct response on each trial. No explicit labels or coding instructions were provided, permitting subjects to adopt their own strategies in learning to categorize these stimuli. After 320 trials, two additional intermediate stimuli (-30 and +30 msec.) were included in the training set and another 160 trials were run. Of the original twenty-four subjects, nineteen met the criterion of at least 90% or better correct performance on the four stimuli in the last block of trials. These subjects were then divided into two groups and asked to return for the remaining sessions on Days 2 and 3.

Testing on Days 2 and 3 was identical for each group and included an initial practice sequence of 80 trials using the endpoint stimuli, -50 msec. and +50 msec., with feedback in effect. This was followed immediately by a baseline identification test in which all eleven stimuli were presented fifteen times each in a random order for 165 trials. Finally, identification was measured under adaptation conditions

separately for each endpoint stimulus. One group of subjects was assigned the -50 msec. stimulus as an adaptor whereas the other group received the +50 msec. stimulus. Subjects received 100 repetitions of the adaptor followed by a single randomized presentation of nine stimuli selected from the middle of the continuum. Subjects were required to identify each of the nine test stimuli into one of the two response categories they had used earlier in the training and identification sessions although no feedback was provided. Ten sequences of adaptation and identification testing were run in a session providing ten responses for each stimulus per day. Timing and sequencing of trials in the experiment was paced to the slowest subject in a given session.

Results and Discussion

The average identification functions obtained for baseline and adaptation are shown in Figure 2 separately for each of the two adaptation conditions.

Insert Figure 2 about here

A small shift can be observed in the identification function measured after adaptation for the -50 msec. group shown in the lefthand panel of the figure. However, examination of the average identification data for the +50 msec. group does not

reveal a noticeable or consistent difference between the functions obtained for baseline or adaptation, at least when considering the group data as a whole.

In order to get a better indication of the strength of these results, the locus of the category boundary was determined by an algorithm that interpolated linearly between the two stimulus values on either side of the 50% crossover point in the identification functions.

Insert Table 1 about here

Examination of the individual boundary values given in Table 1 shows that all nine subjects in the -50 msec. or "lead" adaptor group showed a shift in their identification functions after adaptation with this endpoint stimulus. The boundary shifts in this condition were small but, nevertheless, they were in the anticipated direction displaced toward the adapting stimulus. The difference between baseline and adaptation functions for this condition was statistically significant by a one-tailed *t*-test for matched samples ($p<.005$) indicating the presence of a reliable adaptation effect.

Turning to the individual data shown in Table 1 for subjects in the +50 msec. or "lag" group, it can be seen that two of the ten subjects showed boundary shifts after adaptation in the

direction opposite to that anticipated. Moreover, one of these subjects showed a large shift in the wrong direction thus accounting for the relatively small effects shown for the group identification functions when the average data are displayed in Figure 1. However, despite the results for these two subjects, a t-test on the differences between baseline and adaptation conditions was significant ($p<.05$) indicating the presence of a reliable adaptation effect with the +50 msec. adaptor as well.

The results of this experiment indicate that small although reliable adaptation effects can be obtained with nonspeech stimuli differing in relative onset time of their component frequencies. Although adaptation effects have been observed for other nonspeech stimuli, particularly along relatively simple perceptual and sensory dimensions (Ward, 1973), the present results are of special interest because such effects were obtained with a temporal dimension closely related to one studied extensively in speech--VOT. Moreover, these findings encourage the view, already suggested from our earlier work, that the auditory system responds selectively to the temporal order of events and that this temporal dimension may reflect one of the basic patterns of auditory system response in processing of speech as well as other acoustic signals.

Experiment II

In this experiment, cross-series tests using synthetic speech stimuli were carried out to determine how selective the earlier adaptation effects were with nonspeech adaptors. On one hand, it is possible that the adaptation effects observed for temporal onset are due to a very general timing mechanism that simply responds to differences in temporal onset in the auditory system regardless of whether the signals are speech or nonspeech. On the other hand, the timing mechanism may be quite general but the specific effects revealed through the adaptation technique could well be more selective requiring, at the very least, a fair degree of spectral overlap between the adapting and test stimuli. Finally, it is also possible that two quite distinct mechanisms exist in the auditory system for processing temporal order information, one restricted exclusively to speech and the other to nonspeech signals. The results of the following two cross-series adaptation experiments should help to clarify these possibilities.

Method

Subjects. Eighteen new volunteers served as subjects. They were recruited in the same way and met all of the requirements as in the previous experiment.

Stimuli. The same eleven nonspeech stimuli differing in temporal onset were also used as test stimuli in this experiment.

However, two additional stimuli were used to test for the presence of cross-series adaptation effects from speech to nonspeech. The adaptors were two synthetically produced speech stimuli differing in VOT by -50 msec. and +50 msec. and were selected from the labial VOT series constructed by Lisker and Abramson (1967). The stimuli consisted of a 450 msec. steady-state formant pattern (F1=769 Hz, F2=1232 Hz, F3=2525 Hz) appropriate for the vowel /a/ with 45 msec. formant transitions at starting values appropriate for a bilabial stop consonant. Voicing lead was simulated by the presence of a low amplitude first formant at 154 Hz. The aspiration and voicing differences for the voicing lag stop were realized by attenuation of low frequency energy in F1 and the presence of hiss in the two higher formants. The pitch contour was fixed at 114 Hz for 320 msec. and then tapered off linearly to 70 Hz over the remainder of the steady-state portion of the vowel. The entire Lisker and Abramson series was originally recorded on audio tape at Haskins Laboratories and later digitized on the same PDP-11 used in the earlier experiment.

Procedure. The procedure was identical in all ways to that used in the previous experiment except that the -50 and +50 msec. VOT stimuli were substituted for the two nonspeech adaptors. The first day was used for training, testing and subject selection. On the second and third days, identification tests were carried out under both baseline and adaptation conditions.

Results and Discussion

The average identification functions obtained for both baseline and adaptation are shown in Figure 3 separately for each group of subjects.

Insert Figure 3 about here

Examination of the group data displayed in this figure shows no consistent shift in the identification functions for either adaptation condition. When the individual boundary values are examined, as shown in Table 2, it is quite apparent that the speech stimuli differing in VOT simply did not produce any consistent effects on identification of the nonspeech test series.

Insert Table 2 about here

A *t*-test for matched pairs was carried out on the data in each condition to confirm these initial observations, and in both cases, the test failed to reach statistical significance. Thus, the results of the present experiment clearly indicate that speech adaptors differing in VOT are not able to produce cross-series adaptation effects on a set of nonspeech stimuli differing in relative onset time.

Considering the enormous differences in spectral composition between the synthetic speech adaptors and the nonspeech test series used in this experiment, it seems reasonable to conclude that the adaptation effects found for temporal order in the first experiment with nonspeech adaptors were, in fact, quite specific in scope. Such effects were not observed when the adaptors were more complex speech stimuli differing in VOT and the test series consisted of spectrally simpler nonspeech signals. Since the temporal differences in these VOT adaptors were distributed across component formant frequencies containing substantially wider bandwidths than the two-tone components of the nonspeech stimuli, any adaptation that might have been produced may have simply been too broadly spread across the spectrum. Thus, with the sensitivity of the present adaptation technique, it may have been difficult to detect the presence any of these effects.

In addition to the temporal onset dimension under consideration, a number of other differences can also be noted between the speech and nonspeech adaptors. These factors could have influenced the overall selectivity brought about by repeated stimulation with these speech adaptors. For example, the speech adaptor with the voicing lead (i.e., -50 msec.) contained formant transitions after release while the speech adaptor with a voicing lag (i.e., +50 msec.) contained significant noise during the period of the aspiration interval after release. The extent to which variations in these attributes influence the processing of

temporal order information is currently unknown although potentially important in terms of understanding the initial sensory coding of speech signals by the auditory system.

Experiment III

The results of the previous experiment failed to reveal any cross-series adaptation effects on a nonspeech continuum when the adapting stimuli were speech signals differing in VOT. Such results could be due to a dissociation of the perceptual mechanisms used to process timing information across different perceptual classes. Alternatively, the outcome might simply be due to an asymmetry from complex to simple signals. In earlier adaptation studies, some evidence was found for adaptation effects on speech stimuli when the adapting stimuli were components of speech signals such as formant transitions or isolated formants (see Tartter and Eimas, 1975). To determine if there is a dissociation of processing temporal order information across perceptual classes such as speech and nonspeech, we carried out an additional experiment in which the adaptors were nonspeech TOT stimuli and the test series consisted of speech signals differing in VOT. If cross-series adaptation effects can be found in this experiment, the results would suggest the operation of some fairly general timing mechanism for both speech and nonspeech signals. On the other hand, the absence of cross-series effects from nonspeech to speech, taken together

with the results of the previous experiment, would imply a dissociation between perceptual classes and therefore a certain degree of selectivity for processing temporal onset information in speech and nonspeech.

Method

Subjects. Nineteen new volunteers were recruited for this experiment. They met all of the same requirements as in the earlier experiments.

Stimuli. Eleven synthetically produced labial stop CV syllables differing in VOT from -50 msec. to +50 msec. in 10 msec. steps were used as test stimuli. The stimuli were originally produced at Haskins Laboratories by Lisker and Abramson (1967) on the parallel-resonance synthesizer and then recorded on audio tape. The endpoints were identical to the stimuli used in the previous experiment although all the intermediate VOT values were included in this series as well. The two adapting stimuli were selected from the endpoints of the nonspeech TOT series used in the previous experiments. They had onset time values of -50 msec. and +50 msec., respectively.

Procedure. The general procedure was similar to that used in the previous two experiments except that the methods used for training, testing and subject selection on Day 1 were eliminated since subjects had no difficulty in identifying the test stimuli as /ba/ or /pa/. As a consequence, all baseline and adaptation

testing could be carried out in one session lasting about an hour. Baseline testing was always conducted first, followed after a short break, by adaptation testing. Subjects were required to identify all eleven VOT stimuli during baseline testing but only the middle nine test stimuli during adaptation testing. The stimuli were presented fifteen times each for identification under both testing conditions. As in the previous experiments, all timing and sequencing of trials was paced to the slowest subject in a group.

Results and Discussion

The average identification functions obtained during baseline and after adaptation testing are shown in Figure 4 separately for each of the two adaptation conditions.

Insert Figure 4 about here

Inspection of the figure reveals no consistent shift in the group identification functions after adaptation in either condition. The individual boundary values of baseline and adaptation are shown in Table 3.

Insert Table 3 about here

While there is a slight tendency for the boundary values to shift in the expected direction for some subjects after adaptation, the overall differences were not significant as revealed by t-tests for matched pairs. As in the previous experiment, cross-series adaptation effects appear to be difficult to obtain, at least under the conditions used in the present experiments. Thus, at least for processing temporal order information, there appears to be a clear dissociation between perceptual classes since the absence of adaptation effects in both cross-series tests was symmetrical from speech to nonspeech and vice-versa. The selective adaptation effects observed for VOT in speech or TOT in nonspeech signals seem to be restricted to within a particular perceptual class. Taken at face value such results would, of course, undermine any arguments in favor of revealing the operation of a very broadly tuned timing mechanism in the auditory system that was sensitive to temporal order information in both speech and nonspeech signals. Thus, when considered in this light, it is apparent that the temporal order hypothesis is not sufficient by itself to account for the complexity of the numerous temporal and spectral cues to the voicing feature in speech, an issue we will return to in the next section.

General Discussion

The overall results of the present experiments are consistent with several earlier findings on perceptual adaptation for speech and nonspeech signals. One group of studies has examined the effects of nonspeech adaptors on the perception of speech stimuli in order to gain additional information about the level of processing tapped by the adaptation technique. In general, these cross-series adaptation effects with nonspeech stimuli were smaller in magnitude than the cross-series or within-series adaptation effects obtained with speech stimuli and were closely related, in many cases, to the degree of spectral overlap of the adaptor and test series. In his review, Cooper (1975) summarized several early studies on monaural and binaural sites of adaptation with speech and suggested the possibility of at least two types of detectors, a set of low-level detectors and a set of higher-level integrative detectors. He further suggested that the low-level detectors might be monaurally driven while the integrative detectors might operate subsequent to binaural fusion.

Sawusch (1977a) provided evidence for both of these levels in a series of interaural transfer experiments. Moreover, using this design he was able to obtain large adaptation effects even when adaptor and test series differed in frequency by at least a critical bandwidth. Based on these results, Sawusch argued that

the adaptation effects obtained with these speech stimuli were due to the operation of higher-level detectors that respond to more abstract relational properties of the stimuli in the absence of any spectral overlap. Thus, spectral commonality is not a necessary requirement for adaptation to be produced, at least for place of articulation, although the level of perceptual analysis obviously exerts a strong influence whether any adaptation will be observed.

Several other studies have also sought to determine the effects of nonspeech adaptors on the perception of nonspeech signals. Cutting, Rosner & Foard (1976) demonstrated selective adaptation effects for a nonspeech continuum varying in rise-time between "pluck" and "bow." Moreover, in cross-series tests involving variations in frequency and waveform they found that the largest postadaptation shifts occurred when the adapting and test stimuli shared both frequency and waveform. Progressively smaller cross-series shifts in identification were observed when adapting and test stimuli shared only frequency or only waveform or neither of these properties.

The experiments reported in the present paper differ from earlier adaptation experiments in several respects. First, the acoustic dimension under consideration here was a temporal variable closely related to one found to distinguish voicing in stop consonants. Previous nonspeech adaptation experiments have been concerned primarily with spectral differences. Secondly,

subjects could not readily categorize these nonspeech stimuli as they could in the "pluck" and "bow" experiments. Consequently some experience and training were required to be able to identify these stimuli consistently. Thus, it is difficult to argue that subjects had prior experience or familiarity with these particular perceptual categories. Finally, the present experiments represent one of the very few attempts reported in the literature to determine the effects of a relatively complex adapting stimulus such as speech on a simpler test series consisting of nonspeech signals. Other studies have examined the effects of speech-on-speech and nonspeech-on-speech (see Diehl, 1976; Samuel & Newport, 1979), but only one other study has attempted to assess the effects of speech adaptors on a nonspeech test series (Verbrugge and Liberman, 1975). In one condition in this study, listeners were asked to categorize a full series of isolated third formants from an /r/ to /l/ test series. In contrast to the results obtained in the present study, the nonspeech identification boundaries for the isolated /r/ to /l/ continuum were unstable and, moreover, no consistent shifts could be induced by either speech or nonspeech adapting stimuli.

The present series of experiments has been carried out more-or-less in the tradition of earlier adaptation experiments in speech perception aimed at establishing the existence and operating characteristics of feature detectors. However, a number of recent papers have raised serious questions about the

conclusions and implications of this earlier body of work (see Eimas and Miller 1978 for a recent review). For example, in a recent paper, Remez (1979) has raised several objections to the opponent-process conceptualization that has been explicitly assumed as the underlying organizational structure of the detectors. Remez found adaptation effects for a continuum differing between vowel (speech) and buzz (nonspeech) and argued that the presence of such effects implied the existence of a set of detectors organized as opponent-pairs for a speech-nonspeech distinction. Remez argued that such a feature opposition is hard to rationalize either linguistically, in terms of what is currently known about the inventory of distinctive features in language, or psychophysically, with regard to the structure and function of the auditory system. As a consequence, he concluded that selective adaptation of speech does not depend on, nor imply, the existence of feature detectors at all, but may result simply from perceptual sensitivity to higher-order values inherent in the stimulus.

In another recent paper, Simon and Studdert-Kennedy (1978) have argued that the physiological metaphor of feature detectors in speech perception implied by the selective adaptation experiments is unwarranted in the absence of converging evidence supporting their existence and operation; almost all of the empirical data used to support feature detector models of speech perception has, in fact, come from selective adaptation

experiments. Instead, these authors prefer to use a more descriptive or neutral term-- "channels of analysis," to characterize the way the auditory system responds selectively to properties of the stimulus input whether the signal is speech or nonspeech (see also Eimas and Miller, 1978). Such an account emphasizes the functional rather than structural nature of the observed selective adaptation effects. Hypotheses concerning the underlying structural organization of the mechanisms responsible for the effects can therefore be distinguished from descriptions of the phenomenon itself and the experimental variables that influence its magnitude.

Although the results of our first experiment demonstrated adaptation effects for temporal onset in nonspeech signals, the absence of any cross-series effects in the remaining two experiments is somewhat problematical with regard to the temporal order hypothesis of voicing perception. First, the within series adaptation effects were generally quite small in magnitude to begin with leaving the possiblity open that the absence of any adaptation effects in Experiments II and III might be due to the insensitivity of the expermental procedures to relatively small effects. Secondly, it is clear from earlier work by Haggard and Summerfield (1977) and Lisker (1975) that voicing perception is not based exclusively on the extraction of a single invariant dimension or property from the speech waveform. Rather the perception of voicing appears to involve a complex integration of

a number of somewhat disparate temporal and spectral cues that are highly context sensitive. Finally, there is the more general question raised earlier of precisely what is being revealed by the selective adaptation technique. The numerous studies carried out over the last few years all seem to support the conclusion that the channels of analysis revealed through selective adaptation are quite selective in their susceptibility to fatigue. Thus, the perceptual channels revealed through the use of this technique are not as general as investigators may have originally thought a few years ago (Eimas and Corbit, 1973). Based on these findings, selective adaptation may simply not reveal the operation of broadly tuned mechanisms that process temporal order information in the auditory system, if they exist at all.

Even if it is assumed that TOT and VOT are not mediated through the same perceptual mechanisms, they are similar in some sense if only because they each provide converging evidence about the fundamental limits of processing temporal order information in the auditory system. Such a view implies that the structure and function of the auditory system constrains the early processing of speech and nonspeech signals alike. Search for commonalities between the two perceptual classes is therefore an important goal although psychophysical and sensory-based accounts of speech perception have often been overlooked in favor of more abstract accounts dealing only with the perception of speech

signals. While there may be important differences in perception between speech and nonspeech signals, there may also be many similarities based on common psychophysical and perceptual processes. Such similarities might help to establish the underlying perceptual basis for the acoustic correlates of the distinctive features in speech. The present findings, taken together with the results of our earlier nonspeech experiments may therefore be thought of as examples of how the auditory system constrains the range of potential acoustic attributes that can be employed to signal phonetic distinctions in language. The inventory of distinctive acoustic attributes in speech appears to be limited by both the sound generating properties of the vocal tract and the sensory capabilities of the auditory system.

References

Ades, A. E. Adapting the property detectors for speech perception. In R. J. Wales and E. C. T. Walker (Eds.) New Approaches to Language Mechanisms. Amsterdam: North Holland, 1976, Pp. 000-000.

Cooper, W. E. Selective adaptation to speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, and D. B. Pisoni (Eds.) Cognitive Theory: Volume 1. Potomac, Maryland: Erlbaum Associates, 1975.

Cutting, J. E., Rosner, B.S. and Foard, C.F. Perceptual categories for musiclike sounds: Implications for theories of speech perception. Quarterly Journal of Experimental Psychology, 1976, 28, 361-378.

Diehl, R. L. Feature analyzers for the phonetic dimension stop vs. continuant. Perception & Psychophysics, 1976, 19, 267-272.

Diehl, R. L., Elman, J. L. and McCusker, S. B. Contrast effects on stop consonant identification. Journal of Experimental Psychology: Human Perception & Performance, 1978, 4, 599-609.

Divenyi, P. L. and Sachs, R. M. Discrimination of time intervals bounded by tone bursts. Perception & Psychophysics, 1978, 24, 5, 429-436.

Eimas, P. D. and Corbit, J. D. Selective adaptation of linguistic feature detectors. Cognitive Psychology, 1973, 4, 99-109.

Eimas, P. D., Cooper, W. E. and Corbit, J. D. Some properties of linguistic feature detectors. Perception and Psychophysics, 1973, 13, 247-252.

Eimas, P. D. and Miller, J. L. Effects of selective adaptation on the perception of speech and visual patterns: Evidence for feature detectors. In R. D. Walk and H. L. Pick (Eds.) Perception and Experience. New York: Plenum, 1978, Pp. 307-345.

Elman, J. L. Perceptual origins of the phoneme boundary effect and selective adaptation to speech: A signal detection theory analysis. Journal of the Acoustical Society of America, 1979, 65, 1, 190-207.

Hirsh, I. J. Auditory perception of temporal order. Journal of the Acoustical Society of America, 1959, 31, 759-767.

Hirsh, I. J. and Sherrick, C. E. Perceived order in different sense modalities. Journal of Experimental Psychology, 1961, 62, 423-432.

Liberman, A. M., Delattre, P. C. and Cooper, F. S. Some cues for the distinction between voiced and voiceless stops in initial position. Language and Speech, 1958, 1, 153-167.

Lisker, L. Is it VOT or a first-formant transition detector? Journal of the Acoustical Society of America, 1975, 57, 6, 1547-1551.

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

Lisker, L. and Abramson, A. S. The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences, Prague: Academia, 1967.

Miller, J. D., Engebretson, A. M., Spenner, B. F. and Cox, J. R. Preliminary analyses of speech sounds with a digital model of the ear. Journal of the Acoustical Society of America, 1977, 00, 000-000.

Miller, J. L. Properties of feature detectors for speech: Evidence from the effects of selective adaptation on dichotic listening. Perception and Psychophysics, 1975, 18, 389-397.

Miller, J. L. Properties of feature detectors for VOT: The voiceless channel of analysis. Journal of the Acoustical Society of America, 1977, 62, 641-648.

Pisoni, D. B. Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops. Journal of the Acoustical Society of America, 1977, 61, 1352-1361.

Remez, R. E. Adaptation of the category boundary between speech and nonspeech: A case against feature detectors. Cognitive Psychology, 1979, 11, 38-57.

Samuel, A.G. and Newport, E.L. Adaptation of speech by nonspeech: Evidence of complex acoustic cue detectors. Journal of Experimental Psychology: Human Perception & Performance, 1979, 00, 000-000.

Sawusch, J. R. Peripheral and central processes in selective adaptation of place of articulation in stop consonants. Journal of the Acoustical Society of America, 1977, 62, 738-750. (a).

Sawusch, J. R. Processing place information in stop consonants. Perception and Psychophysics, 1977, 22, 417-426. (b).

Summerfield, A. Q. and Haggard, M. P. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. Journal of the Acoustical Society of America, 62, 2, 435-448.

Searle, C. L., Jacobson, J. Z. and Rayment, S. G. Phoneme recognition based on human audition. Journal of the Acoustical Society of America, 1979, 00, 000-000.

Simon, H. J. and Studdert-Kennedy, M. Selective anchoring and adaptation of phonetic and nonphonetic continua. Journal of the Acoustical Society of America, 1978, 64, 1338-1357.

Stevens, K. N. The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David, Jr., and P. B. Denes (Eds.) Human Communication:on: A unified view. New York: McGraw-Hill, 1972.

Tartter, V.C. and Eimas, P.D. The role of auditory and phonetic feature detectors in the perception of speech. Perception & Psychophysics, 1975, 18, 293-298.

Verbrugge, R. R. and Liberman, A. M. Context-conditioned adaptation of liquids and their third formant components.

Journal of the Acoustical Society of America, 1975, 57, S1, 52-53.

Ward, W. D. Adaptation and fatigue. In J. Jerger (Ed.) Modern Developments in Audiology. New York: Academic Press, 1973, Pp. 301-344.

Footnotes

* This research was conducted at Indiana University in Bloomington and was supported, in part, by NIMH research grant MH-24027 and NINCDS research grant NS-12179. The final version of the manuscript was completed while the author held a Guggenheim fellowship during a sabbatical leave at the Speech Group, Research Laboratory of Electronics, M. I. T. I thank Jerry C. Forshee at Indiana University for his continued technical assistance and advice. Robert E. Remez and Michael Studdert-Kennedy offered critical and very helpful comments on an earlier draft of this paper and their efforts are greatly appreciated. Reprints may be obtained from the author who has now returned to the Department of Psychology, Indiana University, Bloomington, Indiana 47405.

Table 1

Individual and Mean Category Boundaries
(in ms.)

-50 ms. TOT Adaptor Group				+50 ms. TOT Adaptor Group			
S	Baseline	Adaptation	Difference (B-A)	S	Baseline	Adaptation	Difference (B-A)
7	11.7	5.3	+6.4	1	-14.8	-13.0	-1.8
8	9.3	6.6	+2.7	2	15.8	15.0	+ .8
9	15.2	14.6	+.6	5	3.2	3.3	- .1
11	15.8	14.4	+1.4	6	8.3	1.0	+7.3
12	7.7	4.7	+3.0	13	16.8	20.0	-3.2
21	7.9	6.0	+1.9	14	-21.2	-16.6	-4.6
22	18.1	15.0	+3.1	15	13.6	18.3	-4.7
23	18.8	15.9	+2.9	16	13.7	20.0	-6.3
24	20.0	12.0	+8.0	17	23.4	30.0	-6.6
				18	24.4	29.0	-4.6
Mean:	13.8	10.5	+3.3	Mean:	8.3	10.7	-2.4

Table 2

Individual and Mean Category Boundaries
(in ms.)

-50 ms. VOT Adaptor Group				+50 Ms. VOT Adaptor Group			
S	Baseline	Adaptation	Difference (B-A)	S	Baseline	Adaptation	Difference (B-A)
1	12.5	8.9	+3.6	8	8.2	8.2	0
2	20.6	17.1	+3.5	9	5.0	-2.0	+7.0
3	23.4	22.9	+.5	11	16.0	15.7	+ .3
6	15.3	16.3	-1.0	12	16.4	7.1	+9.3
20	19.2	21.0	-1.8	13	18.8	20.0	-1.2
21	16.7	14.3	+2.4	14	24.0	18.8	+5.2
24	17.5	16.1	+1.4	15	10.0	9.1	+ .9
				16	24.8	20.9	+3.9
Mean:	17.9	16.7	+1.2	17	-10.7	-12.0	+1.3
				18	4.7	3.9	+ .8
				19	13.9	27.3	-13.4
				Mean:	11.9	10.6	+1.3

Table 3

Individual and Mean Category Boundaries
(in msec)

-50 ms. TOT Adaptor Group				+50 ms. TOT Adaptor Group			
S	Baseline	Adaptation	Difference (B-A	S	Baseline	Adaptation	Difference (B-A)
26	30.1	24.4	+5.7	32	17.9	5.6	+12.3
27	3.6	3.0	+.6	33	34.6	34.4	+.2
28	16.9	23.2	-6.3	35	15.0	15.8	-.8
29	10.8	9.0	+1.8	36	20.6	10.7	+9.9
30	25.0	33.1	-8.1	42	13.9	13.9	0
31	17.6	14.2	+3.4	43	10.8	21.0	-10.2
38	20.6	20.6	0	44	26.3	25.4	+.9
39	15.6	9.4	+6.2	45	20.8	9.1	+11.7
41	16.1	11.9	+4.2	46	13.6	15.0	-1.4
				47	34.5	34.6	-.1
Mean:	17.4	16.5	-.9	Mean:	20.8	18.6	2.2

Figure Captions

Figure 1. Schematic representations of three stimuli differing in relative onset time: Leading (-50 msec.), Simultaneous (0 msec.) and Lagging (+50 msec.).

Figure 2. Average nonspeech (TOT) identification functions for baseline and after adaptation in Experiment I. The -50 msec. nonspeech TOT adaptor group is shown on the left; the +50 msec. nonspeech TOT adaptor group is shown on the right.

Figure 3. Average nonspeech (TOT) identification functions for baseline and after adaptation in Experiment II. The -50 msec. VOT speech adaptor is shown on the left; the +50 msec. VOT speech adaptor group is shown on the right.

Figure 4. Average speech (VOT) identification functions for baseline and after adaptation in Experiment III. The -50 msec nonspeech TOT adaptor group is shown on the left; the +50 msec TOT adaptor group is shown on the right.

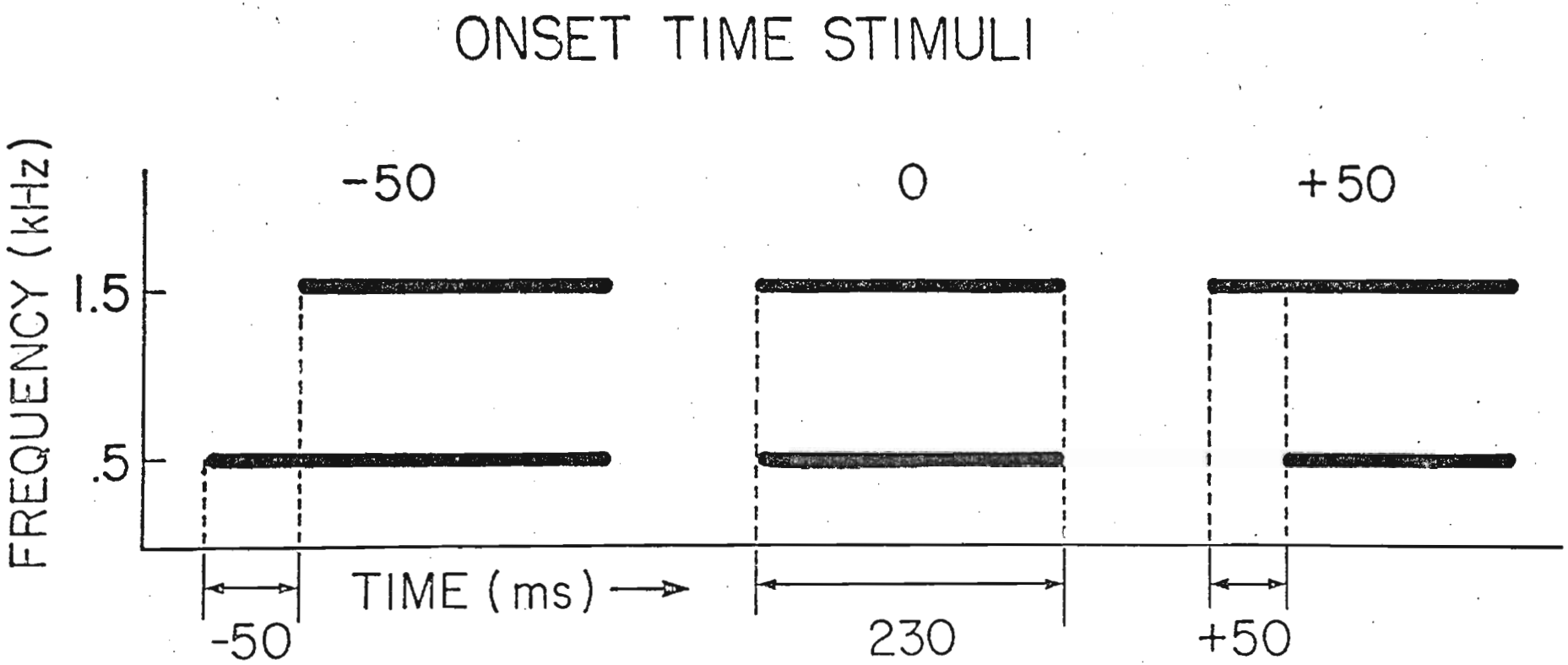


Figure 1.

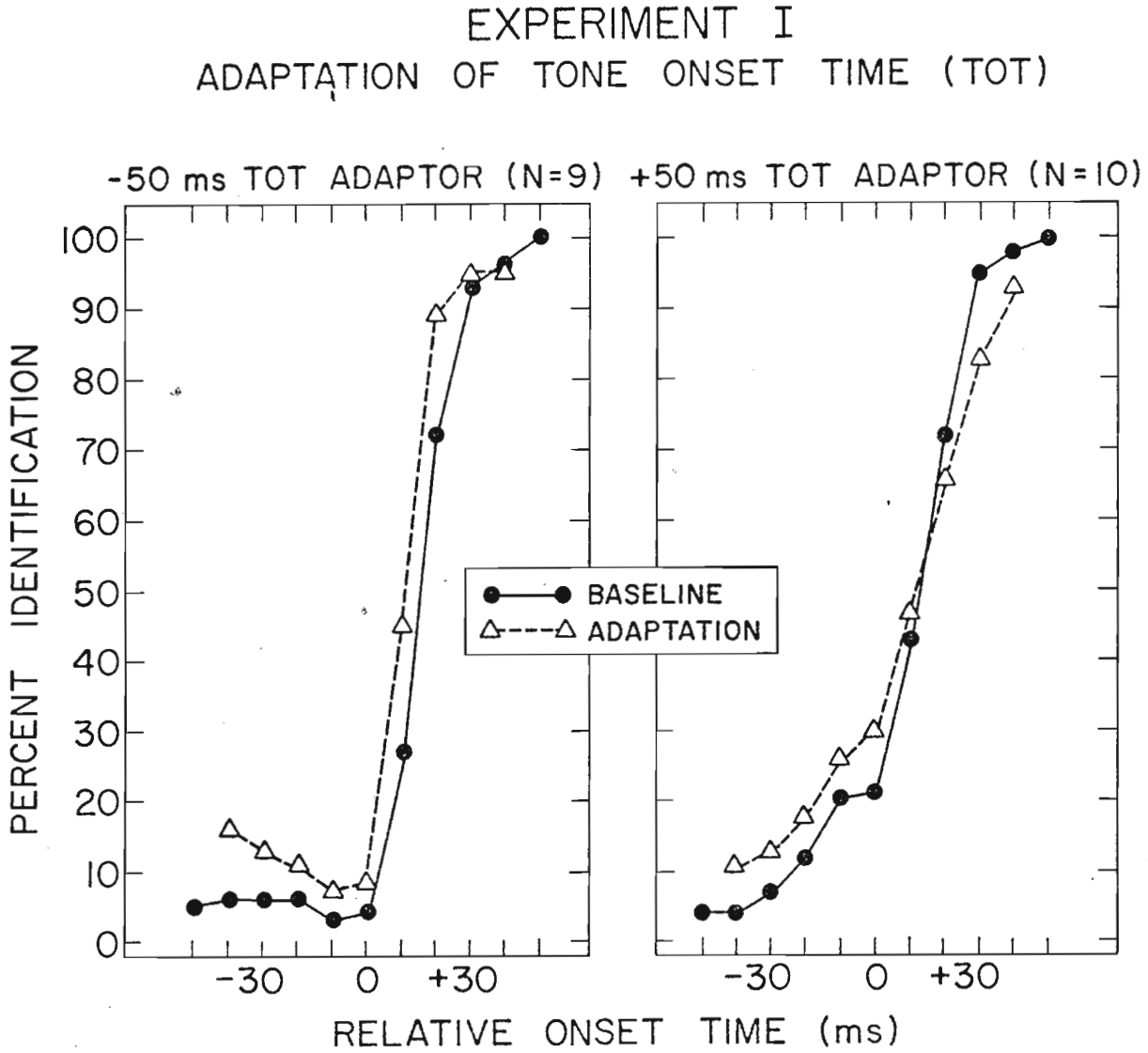

EXPERIMENT I
ADAPTATION OF TONE ONSET TIME (TOT)
-50 ms TOT ADAPTOR (N=9)
+50 ms TOT ADAPTOR (N=10)
PERCENT IDENTIFICATION
100
90
80
70
60
50
40
30
20
10
0
-30
0
+30
BASELINE
ADAPTATION
RELATIVE ONSET TIME (ms)

Figure 2.

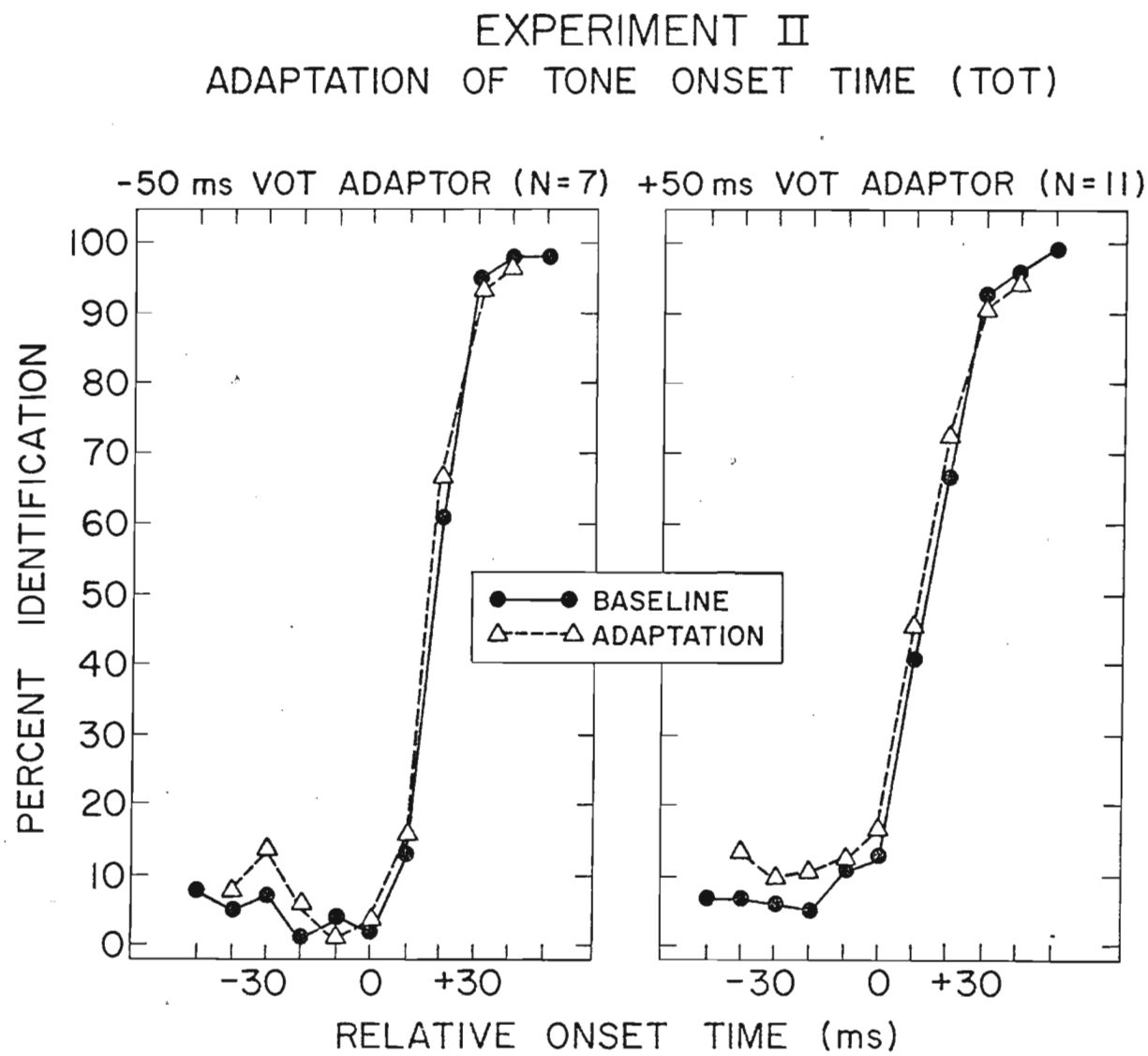

EXPERIMENT II
ADAPTATION OF TONE ONSET TIME (TOT)
-50 ms VOT ADAPTOR (N=7)
+50 ms VOT ADAPTOR (N=11)
PERCENT IDENTIFICATION
100
90
80
70
60
50
40
30
20
10
0
BASELINE
ADAPTATION
-30
0
+30
-30
0
+30
RELATIVE ONSET TIME (ms)

Figure 3.

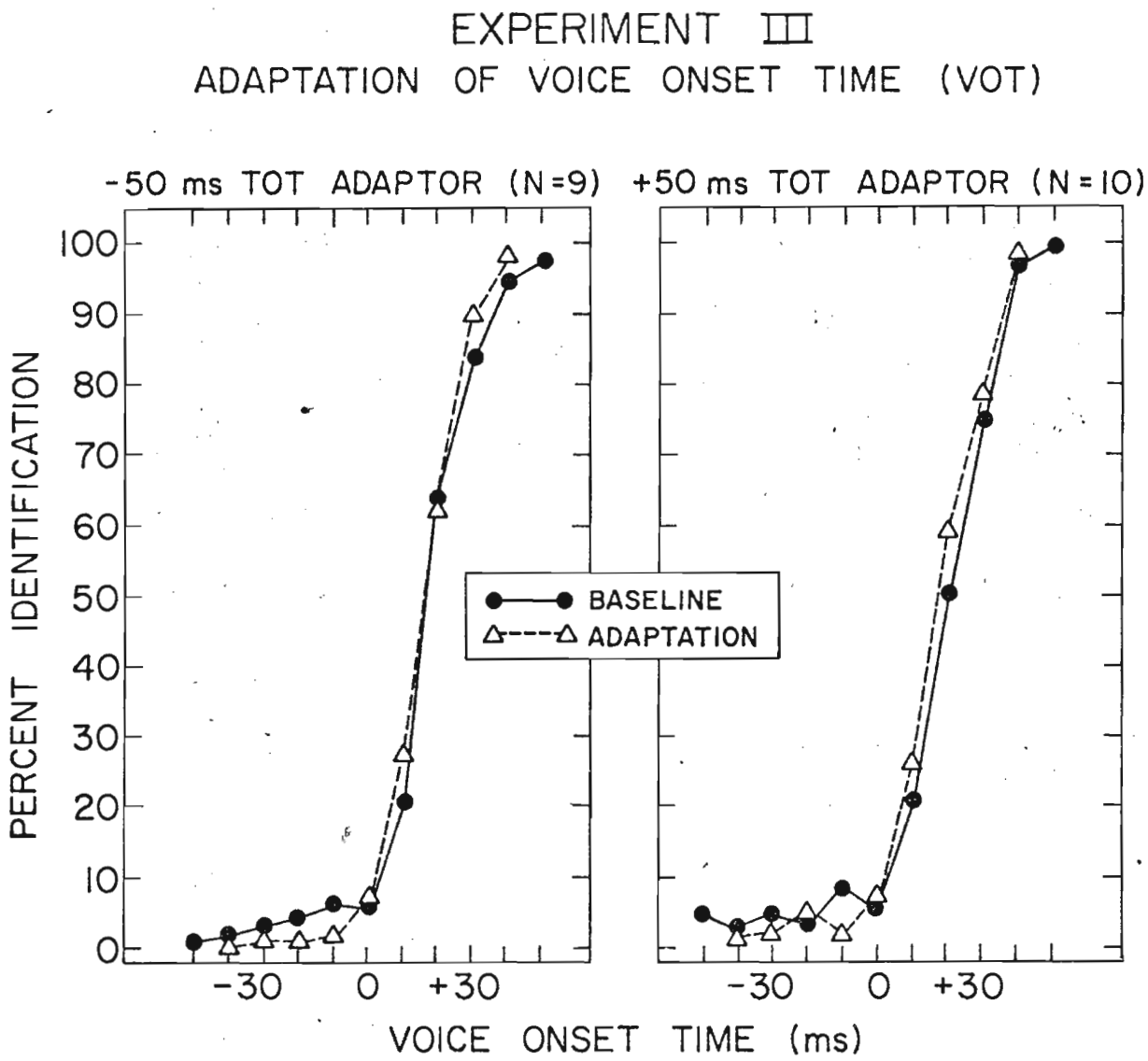

EXPERIMENT III
ADAPTATION OF VOICE ONSET TIME (VOT)
-50 ms TOT ADAPTOR (N=9)
+50 ms TOT ADAPTOR (N=10)
PERCENT IDENTIFICATION
100
90
80
70
60
50
40
30
20
10
0
BASELINE
ADAPTATION
-30
0
+30
-30
0
+30
VOICE ONSET TIME (ms)

Figure 4.

Identification and Discrimination of a New Linguistic Contrast: Some Effects of Laboratory Training on Speech Perception

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

Department of Psychology

Indiana University

Bloomington, Indiana 47405

Short Title: Effects of Laboratory Training on Speech Perception

Abstract

For many years there has been a consensus among investigators that the linguistic experience of an individual exerts a profound and quite often permanent effect on the perceptual abilities required to identify and discriminate between speech sounds. Based on the results of a small number of studies, it has been further assumed that selective modification of phonetic perception cannot be accomplished easily and quickly in the laboratory with simple discrimination training techniques involving only a few hours of practice. The present paper reports the results of four experiments that systematically examined the acquisition of a new linguistic contrast in voicing with naive monolingual speakers of English. Laboratory training procedures were implemented with a small computer in a real-time environment to establish a three-way contrast among voiced, voiceless unaspirated and voiceless aspirated stops. New perceptual categories emerged quite rapidly for most subjects after only a few minutes of exposure to the new contrast and subsequent perceptual tests revealed reliable and consistent labeling and categorical-like discrimination functions for all three categories. These new results differ quite substantially from previous studies in demonstrating that the adult perceptual system is quite plastic and can be modified or "retuned" selectively by environmental input through the use of simple laboratory training techniques.

Identification and Discrimination of a New Linguistics Contrast: Effects of Laboratory Training on Speech Perception*

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

Over the last fifteen years considerable empirical research has been devoted to the detailed study of the voicing feature in stop consonants, particularly as it has been conceptualized in terms of the dimension known as voice-onset-time (VOT). More recently numerous experiments employing synthetically produced speech stimuli have investigated VOT perception in human adults, human infants, chinchillas and monkeys. These developmental and cross-species comparisons have been undertaken in order to better understand the potential interactions between genetic predispositions and experiential factors in perceptual categorization of speech stimuli differing in VOT. The results of these rather diverse studies of VOT have shown the combined influence of two factors operating in speech perception. First, an extensive body of research has established a substantial effect of linguistic experience on speech perception, particularly in human adults exposed to different language-learning environments. Subjects identify and discriminate speech sounds with reference to the linguistic categories in their language. Secondly, another somewhat smaller although more recent series of studies on VOT with animals and young infants as well as a series of experiments with nonspeech

signals has called attention to the operation of several basic sensory and psychophysical constraints on auditory system function. For example, it appears that the perception of voicing in stop consonants requires the analysis of a temporal relations between laryngeal and supralaryngeal events. It has been suggested that these constraints on perception may play an important role in defining the inventory of the acoustic correlates of distinctive features in speech (Stevens, 1972). Thus, based on these considerations, both experiential and genetic factors appear to interact and contribute to the manner in which speech signals are perceived.

The results of the earliest cross-language experiments dealing with the perception and production of VOT carried out by Lisker and Abramson (1964, 1967) confirmed in a quantitative way what had been suspected by linguists for many years-- namely, that the linguistic environment exerts a profound influence on an individual's ability to produce and perceive the distinctive acoustic attributes of speech. In this report we are concerned quite specifically with the extent to which prior linguistic experience controls and maintains the perceptual analysis of VOT and how the strategies used for such perceptual analysis can be modified selectively through laboratory training procedures. Our work was motivated, in part, by the publication of a recent chapter on the role of linguistic experience (Strange & Jenkins, 1978) and the strong assertions that were made concerning the

effects of the linguistic environment on identification and discrimination of speech sounds. Strange & Jenkins (1978) argued that the linguistic environment strongly influences an adult's ability to discriminate between phonetic contrasts that are not distinctive in their native language. Moreover, they have concluded that laboratory training procedures are ineffective in facilitating the acquisition of new phonological contrasts in adult listeners. The results of laboratory training experiments such as those reviewed by Strange and Jenkins also raise very broad and important questions concerning the ontogenesis of the perceptual mechanisms that mediate speech discrimination in young infants as well as their structure and function in adults.

A proper understanding of the mechanisms underlying speech perception, and its susceptibility to environmental influence must begin, of course, with a consideration of subjects whose early linguistic input is known to differ markedly. In the case of voicing in stop consonants, Lisker and Abramson (1964) carried out an examination of the voicing and aspiration differences among stops in eleven diverse languages, and were able to identify three primary modes of voicing: (1) a lead mode in which voicing onset precedes the release from stop closure, (2) a short-lag mode in which voicing onset is more-or--less simultaneous with release from stop closure, and (3) a long-lag mode in which voicing onset occurs substantially after the release. In addition to measurements of VOT in production of

these stop contrasts, Lisker and Abramson also carried out perceptual experiments with synthetically produced speech stimuli differing in VOT. The results of these experiments demonstrated that, in general, subjects from different linguistic backgrounds identified and discriminated between these synthetic stimuli in terms of the distinctive phonological categories of their language. A summary of these findings is shown in Figure 1.

Insert Figure 1 about here

The cross-language identification functions obtained by Lisker and Abramson were very reliable and showed steep slopes at the boundaries separating one phonological category from another. The discrimination functions showed sharp discontinuities along the stimulus continuum with peaks at the cross-over points separating categories in identification. The correspondence of heightened discrimination at the category boundaries coupled with relatively poor discrmination within perceptual categories suggested that subjects could discriminate between stimuli only as well as they could identify them as different on an absolute basis. These results suggest that phonological categories in language are determined in large part by linguistic experience.

The subjects in these early perceptual experiments, as well as those in even more recent studies, appeared to have great

difficulty in identifying and subsequently discriminating between stimuli that were not distinctive in their language. The failure of adults to perceive non-native distinctions in voicing has been interpreted by some investigators as support for the view that linguistic experience exerts a very profound effect on an individual's ability to discriminate speech stimuli. Indeed, in a recent chapter Eimas (1978) has even suggested the possibility that the relevant perceptual mechanisms might atrophy if stimulation is not forthcoming.

These conclusions about the role of linguistic experience in speech discrimination have been widely accepted in the literature on speech perception despite the existence of several studies demonstrating that subjects can discriminate small differences between speech sounds that are identified as the same phonological category. When the experimental conditions are modified by reducing uncertainty or when the subjects are explicitly directed to attend to the acoustic differences between the signals rather than their phonetic qualities subjects can discriminate very small differences in VOT. Nevertheless, based on an extensive review of the literature on the effects of linguistic experience, Strange and Jenkins (1978) preferred to accept the same general conclusion that has been prevalent in the literature for over twenty years--namely, that subjects do not typically discriminate between speech sounds unless they are used distinctively in their native language. A strong conclusion such

as this has several broad implications not only for conceptions of perceptual development in young infants but also for the attitude it promotes toward questions surrounding reacquisition and perceptual learning in adults.

Although subjects typically fail to discriminate between speech sounds that are nondistinctive in their native language, a small number of laboratory training studies have been conducted in the past to determine the extent to which subjects can be made consciously aware of these distinctions. In discussing these studies in their review, Strange and Jenkins (1978) concluded that the use of laboratory training techniques with adult subjects was generally ineffective in promoting enhanced discrimination of phonetic contrasts that were not already present in the subject's language. Even with specific laboratory training experience adult subjects appeared to be incapable of acquiring (or reacquiring) a new linguistic contrast, at least under the conditions employed in these earlier studies.

These conclusions are also consistent with the recent views of Eimas based on his work with young infants. Eimas suggested that "the course of development of phonetic competence is one characterized by a loss of abilities over time if specific experience is not forthcoming." Thus, like the adult, if phonetic differences are not used distinctively in the language-learning environment of an infant, sensitivity to the relevant acoustic attributes of these speech sounds may be

attenuated and the developing child will therefore fail to develop the specific mechanisms needed to discriminate the differences between these sounds. Of course, one of the most interesting aspects of the recent work on infant speech perception is the extent to which environmental input determines the developmental course and final sensitivity of the perceptual mechanisms employed in processing speech signals (see Aslin & Pisoni, 1978). The extensive literature on the role of early experience in visual system development indicates that early environmental experience can modify the selectivity of cortical cells in kittens (see Blakemore (1974) and Daniels & Pettigrew (1976) for reviews). The analogy to the findings on visual system development has already been drawn by Eimas (1978) in reviewing the earlier infant work. He has argued that the lack of experience with specific phonetic contrasts in the local environment during language acquisition has the effect of modifying the appropriate feature detectors by reducing their sensitivity to specific acoustic cues in the speech signal. Thus, some detectors that were originally designed to process certain phonetic distinctions in speech may be "captured" or "subsumed" by other detectors after exposure to particular acoustic signals in the language learning environment. These detectors might, therefore, assume the specificity for only those attributes present in the stimuli to which they have been exposed. As a consequence, then, the poor discrimination

observed for some phonetic contrasts might actually be due to the modification and possible realignment of the low-level sensory mechanisms employed in discrimination of these acoustic attributes.

Our own recent findings have demonstrated that young infants from English speaking environments can discriminate both lead and lag contrasts along the VOT continuum (Aslin, Hennessy, Pisoni & Perey, 1979). Thus, we became quite interested in reexamining the numerous questions surrounding the ability of adults to identify and discriminate VOT contrasts that were not originally distinctive in their own language learning environment. Considering the previous attempts to use laboratory training procedures, we were also interested in determining why these earlier efforts seemed to be so uniformly unsuccessful in producing substantial changes in identification and discrimination of VOT (Strange, 1972). As part of the present study, we also wanted to determine precisely how much training and experience would be required for adult subjects to acquire a new linguistic contrast in voicing, whether it could be accomplished effectively in the laboratory in just a few hours, or whether it would require substantially more experience and training to produce changes in both identification and discrimination performance. Finally, we were also interested in determining how specific the training procedures would have to be to show evidence of the acquisition of a new linguistic contrast

and the nature of the variability among individual subjects in learning to perceive a new contrast in voicing.

Experiment I

The purpose of this experiment was to determine whether a large group of naive subjects could identify three perceptual categories along a VOT continuum without any formal training or systematic feedback. In earlier studies on VOT perception by Lisker and Abramson, the English subjects were only required to use two responses in identification corresponding to the phonological contrasts that occur in the language. As far as we know, subjects were never asked if they could identify stimuli into additional categories. The results of this study served as an initial benchmark for subsequent experiments in which more specific training procedures were implemented.

Method

Subjects. Twenty naive undergraduate students at Indiana University were recruited as paid subjects through an advertisement in the student newspaper. The group consisted of 17 females and 3 males with a mean age of 20.2 years. All subjects were monolingual right-handed speakers of English and reported no history of a hearing or speech disorder in a pretest questionaire. They were paid at a flat rate of $2.50 per hour for each testing session.

Stimuli. The stimuli for this and all subsequent experiments in this report consisted of a set of fifteen synthetic labial stop consonant-vowel syllables that were generated on the cascade-parallel software synthesizer originally designed by Klatt (1977). The fifteen stimuli differed in 10 msec steps of VOT from -70 msec to +70 msec. The values used for synthesis of these stimuli were chosen from measurements of natural speech made by Klatt (1978) as well as our own measurements from spectrograms of the speech of a male talker (RFP) producing various voicing contrasts between stops. 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=1200 Hz, BW2= 90 Hz; F3=2600 Hz, BW3=130 Hz; F4=3300 Hz, BW4=400 Hz; F5=3700 Hz, BW5= 500 Hz). The formant transitions into the vowel were 40 msec in duration and had starting frequencies appropriate for a bilabial stop in stressed syllable initial position (F1=438 Hz, F2=1025 Hz, F3=2425Hz). Voicing lead was simulated by passing the sinusoidal voicing source (ASV) through F1 which was set at 180 Hz with a bandwidth of 150 Hz. These values were chosen to match natural productions measured from broad-band spectrograms. A 10 msec release burst was generated by passing a turbulent noise source (AF) through the by-pass channel (AB) of the parallel branch of the synthesizer which has a broad-band (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 (AH) through the cascade branch to simulate the turbulence produced at the glottis. During the period of aspiration, the bandwidth of F1 was also widened to 300 Hz. To simulate breathiness the bandwidth of F1 was widened linearly at the end of the syllable from 90 Hz to 180 Hz and some aspiration noise was added to the final 35 msec. 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.

Procedure. All experimental events involving the presentation of stimuli and collection of responses were controlled on-line in real-time by a PDP-11 computer. Subjects were run in small groups in a quiet room equipped with six individual cubicles interfaced to the computer. The test stimuli were converted to analog form via a 12-bit D/A converter, low-pass filtered and then presented to subjects binaurally through Telephonics (TDH-39) matched and calibrated headphones. All stimuli were presented at a comfortable listening level of about 80 dB for the steady-state portion of the vowel. The same voltage levels were maintained throughout all of the experiments.

The present experiment consisted of two 1-hour sessions which were conducted on separate days. Subjects were required to carry out two labelling tasks. On each day, subjects identified

the test stimuli by using either two or three resonse categories. Half of the subjects carried out two-category identification on Day 1 followed by three-category identification on Day 2. The order was reversed for the remaining subjects.

In the two-category condition, subjects simply identified the stimuli into two categories corresponding to English /ba/ and /pa/ by depressing an appropriate response key. In the three-category condition, subjects were required to use three response categories, corresponding to [b], [p] and [p^h]. Immediately before testing began in the three-category condition, subjects listened to several tokens of the -70, 0 and +70 msec VOT stimuli presented in order to familiarize them with the stimulus contrasts and the appropriate responses. However, subjects only listened to these stimuli and were not required to respond to them overtly. No feedback was provided nor was any attempt made at this time to "train" subjects in any explicit way.

Testing on each day consisted of the presentation of two blocks of 150 trials each. Stimuli were presented one-at-a-time in a random order. All timing and sequencing of trials in the experiment was paced to the slowest subject in a given session.

Results and Discussion

The average labelling results for each condition are shown separately for each group in Figure 2.

Insert Figure 2 about here

As expected, subjects showed very reliable and consistent two-category identification functions when they were required to use the two perceptual categories that are distinctive in their native language. However, both groups of subjects were also able to make use of three categories reliably in identifying these same VOT stimuli despite the fact that differences in the voicing lead region of the VOT continuum are not phonologically distinctive in English. The group identification functions for the three-category conditions shown in Figure 2 do not accurately represent the consistency found in the data of individual subjects since two out of the twenty subjects failed to use three responses consistently at all in identification. In addition, the group data in this figure does not show the wide range of individual variability among subjects in this labelling condition. More importantly, however, is the fact that a substantial number of these subjects were able to identify the VOT stimuli into three categories reliably and consistently even in the absence of any explicit training or feedback. The individual subject data for two- and three-category identification are shown separately for each group in Figures 3 and 4.

Insert Figures 3 and 4 about here

Inspection of both figures reveals a great deal of consistency across individual subjects in the two-category identification condition, a fact that is not entirely surprising given the group data shown earlier and considering the distinctiveness of the voiced-voiceless contrast in English and the extensive experience these subjects have had in listening to and producing this voicing contrast. What is more interesting in these data, however, are the results for the three-category labelling task. Only two out of twenty subjects (S6 and S17) failed to use three responses at all. Although there is certainly a good deal of variability in the labelling data of the remaining subjects when looked at individually as shown in these two figures, there is also a surprising amount of consistency especially for some of the subjects. As a first approximation, then, these results suggest that naive and relatively unsophisticated listeners can identify phonetic contrasts differing in VOT that are not distinctive in their native language. Moreover, in contrast to earlier findings on the identification of VOT, these results suggest that such a task is quite easy and that relatively high levels of performance can be obtained without any formal or systematic discrimination training and with only a very minimal amount of experience listening to these stimuli.

In order to quantify these findings in more precise terms, we carried out a number of additional analyses on the identification data. (Relative H and Slopes)

The overall findings of the present study are in sharp contrast with the results of numerous earlier experiments in speech perception that demonstrated the strong influence of linguistic experience. In addition, these results also call into question the conclusions advanced by Strange and Jenkins (1978) on the apparent inability of adult listeners to identify and discriminate phonetic contrasts that are not distinctive in their native language. Our results demonstrate quite clearly that a relatively large group of unselected adult listeners have not lost the sensory abilities or the underlying perceptual mechanisms needed to discriminate the acoustic correlates of the voicing feature, particularly in the lead region of the VOT continuum. Subjects are capable of reliably identifying three categories of voicing easily and in a short period of time under laboratory conditions.

Of course, these initial results were obtained in an identification task and it could be argued that subjects might have adopted a set of specific perceptual strategies. In earlier speech perception experiments, a close relationship was observed between identification of speech stimuli into phonological categories and subsequent discrimination of differences between pairs of stimuli selected from the stimulus continuum. The

results of these experiments showed that, in many cases, discrimination could be predicted from identification under the strong categorical perception assumption--namely, that subjects could discriminate between two stimuli only to the extent that they could identify the stimuli differently on an absolute basis. If labelling or perceptual categorization is the primary factor responsible for the observed differences in discriminability found in previous speech perception experiments, it would be of some importance to determine if discrimination of VOT is also affected by prior labelling experience and whether such discrimination could also be modified selectively in the laboratory in a short period of time. In the next experiment, both identification and discrimination data were collected from the same group of subjects to resolve this question.

Experiment II

This experiment was carried out to assess the relationship between the perceptual categories employed in identification and subsequent discrimination of VOT. In one condition, subjects identified VOT stimuli into two categories as in the previous experiment and then carried out ABX discrmination. In the second condition, subjects were required to use three categories in identification which was then followed by ABX discrimination. We hoped to determine if discrimination could also be affected by prior labelling experience and, if so, whether it would be

constrained in the same manner as suggested by the extensive body of earlier research on speech perception. In other words, we wanted to determine whether we could predict discrimination from the identification data obtained in both labelling conditions.

Method

Subjects. Twenty-five additional subjects were recruited. The group consisted of 20 females and 5 males with a mean age of 20.1 years. The subjects met the same requirements as in the previous experiment and were paid at the same base rate.

Stimuli. The same fifteen synthetic stimuli differing in VOT that were generated for the previous study were also used in this experiment.

Procedure. The procedure was similar to that used in the previous experiment except that ABX discrimination was measured after identification. Ten subjects were assigned to the two-category identification condition and fifteen subjects were assigned to the three-category condition. On each of two separate days, subjects first identified the VOT stimuli into categories and then discriminated differences between pairs of stimuli in an ABX format.

As in the previous experiment, no formal training procedures were used in identification. On each day subjects first received one block of 150 trials for identification. This was followed, after a short break, by the ABX test which consisted of one block

of 208 trials. All thirteen two-step pairs of stimuli from the VOT continuum were presented four times each in all possible ABX arrangements. Stimuli within an ABX triad were separated by 500 msec. Subjects were instructed to determine whether the third sound in each triad was most like the first or second sound and to enter their response accordingly on the response box. No feedback was in effect during identification or ABX discrimination testing.

Subjects in the three-category group also received several tokens of the -70, 0 and +70 msec endpoint stimuli before identification testing began in order to familiarize them with the test stimuli and the appropriate response categories to be used in this condition. All timing and sequencing of trials in both identification and discrimination was paced to the slowest subject in a given session.

Results and Discussion

The average identification and ABX discrimination functions are shown separately for each group in Figure 5.

Insert Figure 5 about here

The two- and three-category labelling functions shown in the left-hand panel of each figure are similar to those obtained in

the previous experiment. Although the average two-category data shown here are very consistent and representative of individual subjects, the average three-category data are less consistent and do not reflect the general pattern of individual subjects. Two out of the original fifteen subjects in this group failed to use three response categories. Identification functions for individual subjects in both groups are shown in Figures 6 and 7.

Insert Figures 6 and 7 about here

Examination of the average ABX discrimination functions shown in the right-hand panels of Figure 5 reveals two distinct peaks in discrimination for both groups of subjects. The larger peak occurs in the voicing lag region of the continuum at roughtly 20 msec whereas a smaller peak can be observed in the voicing lead region at roughly -20 msec. Although the overall level of ABX discrimination is significantly higher for subjects in Group II, the shape of the two group discrimination functions is very nearly identical.

It is interesting to note here that subjects in Group I, the two-category labelling condition, showed reliable evidence of discrimination of stimuli in the voicing lead region of the stimulus continuum despite the fact that these stimuli were all classified into the same phonological category. Such a finding

is not entirely surprising given some of the recent demonstrations of within category discrimination in speech perception (Pisoni and Lazarus, 1974; Carney, Widin and Viemeister, 1977). However, it should be emphasized that no special efforts were made in this experiment to direct subjects attention to voicing differences between stimuli in this region of the continuum or to modify the discrimination task so as to improve its sensitivity in anyway. In our view, it is very likely that subjects in some of the earlier VOT experiments could have identified a new perceptual category in this region since these differences in voicing lead are discriminable to adults as well as infants. We suspect that the failure to demonstrate such effects in discrimination stems from the exclusive use of the oddity discrimination procedure which encourages subjects to encode these stimuli linguistically and then base their discrmination on these encoded representations in short-term memory (see Pisoni, 1973).

Many of the individual subjects in Group II also showed two peaks in their ABX discrimination functions corresponding to the boundaries separating the three voicing categories. The individual discrimination data shown for these subjects in Figure 7 is quite variable with some subjects showing a fairly close correspondence between identification and discrimination. Other subjects are considerably more variable. Despite the wide range of variability among individual subjects, however, the results

clearly indicate that naive subjects are capable of identifying and discriminating differences in VOT that represent a new linguistic contrast in voicing.

It is important to emphasize here that these results were obtained without the use of any formal training procedures and in the course of only two 1-hour testing sessions. There can be little doubt from these data that more formalized and systematic training procedures would substantially reduce subject variability and increase response consistency. We take up this matter in the last experiment in this report. However, before proceeding it is necessary to consider one aspect of the present results in somewhat greater detail. In both conditions of this experiment, subjects first identified the test stimuli into perceptual categories and then carried out the ABX discrimination task. It has been argued in the past by a number of investigators that the observed ABX discrimination data are a consequence of some covert labelling process brought about by requiring subjects to identify the test stimuli prior to measuring discrimination. The ABX data from the two-category condition in this experiment suggests that this may not be the case. Nevertheless, it is worthwhile examining discrimination in the absence of any explicit labelling experience to assess its contribution to discrmination. The next experiment is directed at this question.

Experiment III

This experiment was carried out to determine the extent to which prior labelling experience influences discrimination of VOT. The results of the previous experiment showed that the discrimination functions obtained after two- and three-category identification were quite similar suggesting the possibility that discrimination performance may be dissociated, to some extent, from identification. In this experiment we simply measured discrimination in the absence of any prior labelling. If the overall shape of the ABX discrimination function shows two peaks and is similar to the results found in the previous experiment, we can conclude that discrimination performance is not strongly controlled by prior labelling experience. Such a finding would suggest the possiblity that some sensory or psychophysical factor may be responsible for the observed discontinuities in discrimination found at selected regions along the VOT continuum.

Method

Subjects. Twenty additional naive subjects were recruited for this experiment. There were 13 females and 7 males with a mean age of 20.0 years. All subjects met the same requirements as in the previous experiment and were paid at the same base rate.

Stimuli. The same fifteen synthetic stimuli differing in VOT were also used in this experiment.

Procedure. The procedure for ABX discrimination was identical to that used in the previous experiment except that one group of subjects received feedback for correct responses after each trial. Ten subjects were assigned to each of two groups. Subjects in Group I completed testing without feedback while subjects in Group II received feedback. Subjects in each group were run in only a single testing session lasting about one hour. Two blocks of 208 trials were presented in each session.

Results and Discussion

The average ABX discrimination functions for both groups of subjects are shown in Figure 8.

Insert Figure 8 about here

Both groups showed two peaks in their ABX discrimination functions, a relatively large peak in the lag region and a somewhat smaller peak in the lead region of the continuum. The presence of two peaks in ABX discrimination is a general finding observed in the data for all subjects in both conditions of the experiments. The ABX data for individual subjects is shown in Figures 9 and 10.

Insert Figures 9 and 10 about here

Although there is individual variability from subject to subject in both of these conditions, almost all subjects showed two peaks in discrimination that were located in roughly the same region of the VOT continuum at about -20 and +20 msec. Vertical lines have been drawn through these two points to facilitate comparisons. These discrimination functions are quite similar to those obtained in the previous experiment in which the labelling task preceeded discrimination testing. Thus, the present findings demonstrate that the peaks in discrimination are not simply a consequence of prior labelling experience brought about by having subjects identify these stimuli before ABX discrimination. Instead, subjects apparently respond differentially to the psychophysical properties of the stimuli themselves without recourse to perceptual categorization. These discrimination data suggest the presence of relatively narrow regions along the VOT continuum characterized by high discriminability (i.e., boundaries) separated by somewhat broader regions of lower discrminability (i.e., perceptual categories).

In an earlier paper on the perception of nonspeech signals differing in temporal order, one of the authors (Pisoni, 1977), suggested the possibility that VOT perception might be accounted for in terms of a basic limitation on processing temporal order

in the auditory system. Subjects are unable to identify differences in temporal order between component events if their onsets occur with +/- 20 msec of each other. The results of the present experiment on discrimination of VOT are consistent with the general temporal order hypothesis of voicing perception. The overall shape of the ABX discrimination functions suggests the presence of three well-defined perceptual categories that are similar to those obtained in the previous nonspeech experiments (Pisoni, 1977).

It is interesting to note here that Lisker and Abramson (1964) were able to account for the voicing and aspiration differences among the stops they examined by proposing the existence of three basic modes of voicing. They suggested that these three modes of voicing in stops as realized by the dimension of VOT might reflect the operation of a linguistic universal. Thus, at least in the case of VOT, the perceptual categories do not seem to be selected in a random or arbitrary way but rather can be accounted for by considering the underlying sensory and psychophysical constraints on auditory system function. Acknowledging the presence of such underlying constraints in speech perception is of special relevance since it can provide one principled way of accounting for the VOT discrimination data from young infants and chinchillas who have shown a remarkable degree of consistency despite substantial differences in the experimental procedures.

In summary, the results of this experiment have provided additional support for a dissociation of labelling and discrimination of VOT in speech perception. At least under the conditions employed here, prior labelling experience does not appear to substantially affect the general shape and overall level of ABX discrimination. The presence of peaks in discrimination at roughly +/- 20 msec may be interpreted as further support for the operation of an underlying sensory or psychophysical constraint on the perception of temporal order information in speech.

Experiment IV

All of the previous experiments in this report involved testing of naive and unsophisticated listeners who received relatively little exposure or formal training experience with these stimuli. Although substantial individual variability was observed in both identification and discrimination, the overall results suggested that these listeners could easily learn to attend to the relevant acoustic correlates of voicing in the lead region of the VOT continuum. In this last experiment, we were interested in reducing the variability from subject to subject while at the same time increasing response consistency in categorization. To accomplish this in a relatively short period of time, we used a discrimination training procedure which made use of immediate feedback and the presentation of salient

exemplars of the three perceptual categories. After the training phase was completed, subjects who met a predetermined criterion were selected for subsequent testing in which identification and discrimination data were collected.

Method

Subjects. Twelve additional naive subjects were recruited for this experiment in the same way as in the previous experiments. The original group consisted of 10 females and 2 males with a mean age of 21.2 years. All subjects met the same requirements as in the previous experiments and were paid at a flat rate of $3.00 per hour for their services.

Stimuli. The same fifteen synthetic speech stimuli were also used in the present experiment.

Procedure. The experiment involved four 1-hour testing sessions that were conducted on consecutive days. On Day 1 all subjects were trained to identify the -70, 0 and +70 msec VOT stimuli into three categories. As in the previous experiments involving three-category identification, subjects first listened to the three category exemplars presented in sequential order 10 times (i.e., -70, 0, +70; -70, 0, +70; etc.) to familiarize them with the stimuli and required responses. When this phase was completed, subjects received a block of 240 trials for identification. This block of trials consisted of 80 replications of each exemplar stimulus presented in a random order. Immediate feedback was provided after each trial

indicating the correct response. After completing this phase of the experiment on Day 1, subjects who met a predetermined criterion of 85% correct for each stimulus were invited back for the remaining three days of testing.

On Day 2 the criterion subjects initially received a 75 trial warmup sequence containing 25 replications of each of the three training stimuli in a random order with feedback in effect for correct responses. When this was completed, subjects were presented with two blocks of 150 trials of the full stimulus continuum (i.e., -70 through +70 msec) in a random order for identification testing. No feedback was provided during identification testing of the full series in order to measure generalization.

Testing on Days 3 and 4 was identical. Subjects received one block of 150 trials for identification which was followed by one block of 208 trials for ABX discrimination. No feedback was in effect during identification testing although immediate feedback was provided for correct responses after each trial in the ABX test.

Results and Discussion

Of the original twelve subjects, six passed the 85% criterion on Day 1 and were invited back for the remaining sessions of the experiment. Subjects who failed to meet this criterion all responded to the three stimuli at levels well above

chance although they did not reach the required performance levels. Such performance was expected since the results of our earlier experiments demonstrated a wide range of variability among individual subjects in identification. By setting up an initial performance criterion for identification of the three category exemplars we hoped to increase the consistency among subjects on the full stimulus series. At this time, however, we did not attempt to assess how much further exemplar training would be needed to bring the remaining subjects up to comparable performance levels. Obviously, this is a problem to be explored in future studies specifically designed to assess various types of discrimination training procedures.

The average identification functions for the six subjects who met the criterion on Day 1 are shown in the left-hand panel of Figure 11.

Insert Figure 11 about here

These are the data collected on Day 2 of testing. As expected, these six subjects showed much greater consistency in labelling stimuli in the voicing lead region of the continuum after only a very modest number of training trials on the exemplar stimuli. The average data shown in this panel of the figure closely mirrors the performance of the six subjects as

shown by an examination of the individual data given in Figure 12.

Insert Figure 12 about here

The consistency of these data were expected since, after all, the subjects were required to meet an initial performance criterion with the category exemplars before any identification data were collected. Nevertheless, we feel that the results for the individual subjects are quite striking given the earlier reports indicating the apparent difficulty that English subjects have in identifying and subsequently discriminating differences in voicing lead.

The average identification and ABX discrimination data collected on Days 3 and 4 are shown in the right-hand panel of Figure 11. The corresponding individual subject data for Days 3 and 4 are shown in the right-hand panels of Figure 12. As observed earlier in Experiments II and III, the present ABX discrimination functions show peaks at category boundaries and troughs within perceptual categories, although the overall level of discrimination performance is slightly higher here than in the data obtained in the two previous experiments.

As expected, after a brief period of discrimination training involving the presentation of salient category exemplars combined

with immediate feedback, naive subjects were able to identify and subsequently discriminate between stimuli that constitute a new linguistic contrast in voicing. Performance of these subjects was very consistent and reliable suggesting that the underlying neural mechanisms and cognitive abilities subserving speech processing have not been lost or realigned from exposure to only English voicing contrasts during the long course of language-learning. These findings suggest, at the very least, that the adult perceptual system is quite plastic and can be "modified" or "retuned" selectively in a relatively short period of time by environmental intervention involving the use of simple laboratory training procedures. Taken together, the results of these experiments on VOT indicate that the adult perceptual system is considerably more flexible than earlier reports may have led investigators to believe.

General Discussion

The overall results of these four experiments demonstrate quite clearly that naive listeners can easily learn to perceive differences in the voicing lead region of the VOT continuum. The relative shape of the observed identification and discrimination functions were consistent with the acquisition of a new voicing category. The present findings are therefore in substantial conflict with the results reported in earlier investigations of VOT perception which suggested that prior linguistic experience

strongly controls voicing perception in adults. Moreover, our results differ quite substantially from previous investigations concerned with the use of laboratory training procedures in speech perception. Since our results have demonstrated rather clearly that voicing perception can be modified quite easily in the laboratory in a very short period of time, often amounting to only a few minutes, it seems appropriate to inquire into some of the reasons why almost all of the earlier studies were so uniformly unsuccessful in selectively modifying voicing perception in adults.

The one exception to this general pattern of results is a brief study by Lane and Moore (1962) who studied the identification and discrimination of voicing by an aphasic patient. Before training, the patient was unable to discriminate differences between /d/ and /t/ when presented in isolation or in the context of minimal pairs. Using synthetic stimuli differing in F1 cutback between /do/ and /to/, Lane and Moore (1962) measured identification and discrimination both before and after a brief training interval. Before training began, the aphasic subject was unable to reliably identify the /do/--/to/ stimuli into two well-defined perceptual categories. Moreover, he was unable to discriminate differences between pairs of these stimuli in an ABX format at levels above chance for all comparisons along the stimulus continuum. The identification training procedure that Lane and Moore (1962) used consisted of simply training two

discrete labelling responses to the two endpoint stimuli. In this case, the stimuli had F1 cutback values of 0 and +60 msec respectively. The entire training session lasted only about fifteen minutes and consisted of the presentation of an alternating arrangement of the two stimuli to enhance their salient differences. The aphasic subject received immediate feedback for correct responses to these two training stimuli after each trial. When the training was completed, the aphasic carried out identification and ABX discrimination tests again.

The results of these tests showed a very dramatic increase in differential labelling of the entire set of stimuli in the identification task as well as a marked improvement in discrimination performance, particularly for pairs of stimuli selected from different perceptual categories. We suspect that previous investigators were probably unaware of this very early training study by Lane and Moore (1962) and the specific procedures used in training identification. These same training procedures were, however, used in studies carried out by Cross and Lane (1962), Lane (1965) and more recently Pisoni (1976, 1977) who were all able to obtain very substantial changes in labelling of nonspeech signals after relatively short periods of training.

If we set aside for the moment the early results of Lane and Moore (1962) on the reacquisition of the voicing contrast in an aphasic and the previous nonspeech experiments, it appears that

all previous attempts to modify voicing perception in adult subjects have failed to produce very consistent or reliable effects on subsequent measures of identification and/or discrimination (see Strange & Jenkins, 1978).

Why have previous researchers been so uniformly unsuccessful at selectively modifying the perception of VOT in adults? Is it something peculiar about speech stimuli or might the differences be a consequence of the specific methodologies employed? When we went back and examined these earlier studies more carefully, we found a number of potentially important methodological and procedural differences that could, in all likelihood, account for the failure to modify the perception of VOT. Once these differences are taken into account, it is relatively easy to see how previous investigators came to such strong conclusions about the effects of linguistic experience on speech perception. Moreover, it is clear to us now why investigators such as Strange and Jenkins (1978) argued so strongly for the position that the adult perceptual system is extremely resistant to environmental modification brought about by laboratory training procedures.

Let us turn to the earliest cross-language experiments carried out by Lisker and Abramson (1967) and then proceed to several more recent investigations. In the perceptual experiments on VOT, Lisker and Abramson found that subjects could readily identify their synthetic stimuli into the phonological categories of their language. The procedure used in these

experiments typically involved having native speakers of the language under study name the initial stop consonant by identifying it with one or another words that occur in their language. As far as we know, no efforts were ever made by Lisker and Abramson to ascertain whether subjects in their experiments could reliably identify more perceptual categories than were used distinctively in their language. It is obvious, at least from the results of our experiments, that subjects can recognize more categories than are in their language. Moreover, as shown in Experiment I, a very large proportion of the subjects can recognize these categories reliably simply by being provided with an additional response category. It was not necessary to institute elaborate training procedures to obtain these results with a large number of naive subjects.

Although subjects in the Lisker and Abramson cross-language experiments may have been able to recognize additional categories by having more response alternatives available to them, the results of the oddity discrimination tests suggested that subjects could not discriminate these differences. The failure to discriminate these differences in VOT is ambiguous since subjects almost always received the labelling task before discrmination was measured. Moreover, discrimination performance was measured in the "oddity" paradigm which strongly encourages subjects to adopt a context-coding mode of response. That is, the stimuli are almost immediately recoded into phonological form

for maintenance in short-term memory to solve the discrimination problem (see Pisoni, 1973, 1975). Such a context-coding mode would also be favored by the high uncertainty conditions of the discrimination task brought about by the use of a roving standard from trial to trial which effectively mixes "easy" trials with "hard" trials. Finally, immediate feedback was not provided during testing. Under conditions such as these, untrained listeners have great difficulty in determining precisely what acoustic attributes they are supposed to attend to in carrying out the discriminaion task. Thus, the failure to discriminate fine phonetic differences within a perceptual category may be more a matter of subjects adopting a very lax criterion for detecting small differences between speech sounds than a true capacity limitation on processing sensory input. The voluminous body of research over the last few years on the underlying psychophysical basis of categorical perception provides very strong support for this account. Thus, the particular combination of tasks and the order of presentation may have been responsible for the observed relations between identification and discrimination found in these early cross-language investigations of voicing perception.

In an experiment designed specifically to study the learning of a new contrast in voicing, Lisker (1970) attempted to train native speakers of Russian to distinguish between voiceless unaspirated and voiceless aspirated stops, a voicing contrast

that is distinctive to English but not Russian speakers. Lisker used a training procedure that was superficially similar to the one used by Lane and Moore (1962) except that no feedback was provided to subjects after each trial during training. Although the Russian subjects could identify the endpoint stimuli (i.e., +10 and +60 msec VOT) in this task somewhat better than chance, their performance was not the same for both stimuli. A majority of the six subjects identified the +10 msec stimulus above 90 percent correct. However, the +60 msec stimulus was identified only better than 75 percent correct for five of the six subjects. Thus, while the majority of these subjects could differentiate the training stimuli and use two discrete labelling responses, their performance on this task was not always very consistent nor reliable. Since feedback was not provided after each trial, the subjects no doubt had difficulty in determining what criterial attributes they were to attend to selectively. Nevertheless, when the training phase was completed, all of these subjects were given the full series including all the intermediate stimuli in what appeared to be a scaling or magnitude estimation task.

Compared to English subjects who were also run in the same scaling procedure, the Russians did not show sharp or consistent identification functions for these stimuli. Instead, their responses showed a somewhat more gradual or continuous change from one stimulus to the next along the continuum from +10 to +60 msec. Lisker preferred to interpret the results as evidence that

these Russian listeners were not generally able to recognize the voicing boundary found in English. However, the outcome of this study is to some extent ambiguous since the Russian subjects may not have been able to selectively attend to the relevant acoustic attributes distinguishing aspirated and unaspirated voiceless stops in the absence of relevant feedback during the training phase. Taken together with the scaling procedure used in this study, subjects may have adopted a strategy of focusng on several different properties of the stimuli over the course of the experiment.

Another attempt to modify voicing perception in adults was carried out by Strange (1972). She tried to train a small number of college-age students to identify and discriminate differences in VOT in the lead region of the continuum where the Thai voiced/voiceless unaspriate boundary occurs. In one study, four subjects received training in the oddity discrimination paradigm with "right"-"wrong" feedback provided after each trial. Subjects were also required to provide confidence ratings after each response. When the training phase was completed, subjects carried out the oddity discrimination task again without feedback.

In comparison to the pretest data collected before any training experience, all four subjects showed improved overall discrimination performance on the VOT stimuli during the posttest. However, no improvement was observed for pairs of

stimuli straddling the Thai labelling boundary at -20 msec VOT. The greatest increase in performance occured for stimuli adjacent to the voicing boundary in English. Based on these results, Strange (1972) concluded that her subjects did not "learn" to discriminate the VOT dimension as native Thai-speaking subjects typically do. Moreover, she concluded that "there is no prepotency for adult native English speakers to discriminate differences in the region of the Thai prevoiced-voiced boundary that can be easily realized by mere practice with feedback."(p. 40).

In another study, Strange (1972) trained three subjects to identify the members of a truncated apical series of VOT stimuli (i.e., -100 to +10 msec) into two discrete categories. Initial training involved presentation of the endpoint stimuli in alternation without immediate feedback. Oddity testing was carried out, as usual, after labelling and the results showed some evidence for a peak in discrimination at the boundary between these two new perceptual categories. However, identification and discrimination tests with a labial VOT series failed to show strong evidence of transfer of training from one series to another. Nevertheless, subjects in this experiment were able to reliably identify members of the truncated series into two categories and, moreover, this labelling experience was carried over to discrimination.

Strange (1972) also carried out a third study on the modification of VOT perception using a scaling procedure. Subjects were required to rate each stimulus along a scale between two endpoint reference stimuli. This procedure was adopted as a way of training subjects to perceive the VOT dimension as an acoustic continuum rather than directing their attention to discrete labelling responses. After training in the scaling task, subjects also carried out oddity discrimination. The results of this study were complicated by subject variability in both tasks. However, there was some evidence that training with the scaling procedure did produce effects on perception of VOT. Posttest results for some subjects showed a shift in the scaling responses toward more gradual or continuous functions. The oddity discrimination results were more inconsistent. Some subjects showed overall improvement in discrimination whereas others did not. As in Strange's second experiment, no consistent transfer effects from one VOT series to another could be observed.

Based on the results of these three experiments Strange and Jenkins (1978) arrived at the following conclusions:

> "The results of these three studies show that, in general, changing the perception of VOT dimensions by adult English speakers is not easily accomplished by techniques that involved several hours of practice spread over several sessions. Although performance on each of the kinds of tests did change somewhat with experience, only the identification training task (which involved practice with general feedback only) produced

> categorical results approaching those found for native speakers of Thai." (p. 154).

Although the previous studies of Lisker (1970) and Strange (1972) indicated that modification of speech sound perception may not be obtained easily in a short period with simple laboratory procedures, several more recent investigations have been carried out that have provided more positive results. For example, Carney, Widin and Viemeister (1977) modified several aspects of the standard procedures used to measure VOT discrimination and observed very substantial improvements in within category performance. In addition, they also showed that with long-term practice and the use of immediate feedback, subjects could learn to identify various VOT stimuli into arbitrarily defined categories depending simply on the experimenter's prior criterion for category membership. Although the results of this study are important in demonstrating the existence of noncategorical perception of VOT, these findings were obtained by introducing substantial changes in the experimental procedures typically in most speech perception experiments (see also Edman, Soli and Widin, 1978). First, a "same"-"different" discrimination procedure was substituted for the more traditional oddity or ABX paradigm. From earlier work, it is likely that this reduced the memory load requirements and encouraged subjects to operate in a trace-coding mode. Second, the particular arrangement of the stimuli tested in the experiment represented a low-uncertainty discrimination task for the observers since the roving standard

procedure was eliminated. By using only one standard during a block of trials in the discrimination task, subjects' criterion can become much more stable from trial to trial. Finally, only a very small number of highly practiced subjects were used and these all had extensive experience in previous psychophysical experiments. Moreover, in the experiments described by Carney et al. their three subjects were run in over a dozen testing sessions distributed over a period of several weeks. Thus, it is not at all surprising, at least to us, that the discrimination performance of these three subjects was so good compared with the results obtained in earlier studies. Extensive experience in listening to these particular stimuli during training combined with immediate feedback and the use of a low uncertainty paradigm appear to be the major factors responsible for the very marked improvement in discrimination.

When the results of our experiments are considered in light of these previous findings, it is apparent that many factors contributed to the poor performance observed by earlier investigators. Nevertheless, it has generally been assumed that the failure to "learn" to perceive a new voicing contrast was somehow related to a permanent change in the perceptual or sensory mechanisms. We believe that such widely held conclusions are unjustified in light of the results reported in the present paper. There is little evidence that the underlying sensory or perceptual apparatus, whatever they turn out to be, has been

"retuned" or modified in any permanent sense. Our results suggest that the perceptual selectivity observed in almost all of the previous studies on VOT perception is primarily a consequence of attentive processes brought about by exposure to a specific subset of distinctive acoustic attributes as used in the phonolgical system of the native language under consideration.

While some investigators have tried to explain away findings showing that adults can discriminate very fine phonetic details by arguing that these subjects do not "typically" discriminate these differences, we find these arguments less than convincing. Statements about what subjects "typically" do in experiments versus claims about what subjects "can" do in experiments are simply vacuous exercises in rationalization since they are not independent of the requirments of the information processing task. Most experimenters carrying out research in speech perception have, in our view, simply failed to distinguish between statements about the sensitivity of the sensory system and criterion shifts brought about by a subject's long experience in the language-learning environment. Our results suggest that the adult perceptual system, at least as it is used in processing voicing information in stop consonants, is quite plastic and is far from being as rigid as one might have been lead to believe from past studies.

References

Aslin, R.N., Hennessy, B., Pisoni, D.B. and Perey, A.J. Individual infants' 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, March, 1979.

Aslin, R.N. and Pisoni, D.B. Some developmental processes in speech perception. Paper presented at the NICHD conference "Child Phonology: Perception, Production and Deviation," Bethesda, Maryland, May 28-31, 1978.

Attneave, F. Applications of Information Theory to Psychology. New York: Holt, Rinehart and Winston, 1959.

Blakemore, C. The conditions required for the maintenance of binocularity in the kitten's visual cortex. Journal of Physiology (London), 1976, 261, 423-444.

Carney, A.E., Widin, G.P. and Viemeister, N.F. Noncategorical perception of stop consonants differing in VOT. Journal of the Acoustical Society of America, 1977, 62, 961-970.

Cross, D. V. and Lane, H. L. On the discriminative control of concurrent responses: The relations among response frequency, latency and topography in auditory generalization. Journal of the Experimental Analysis of Behavior, 1962, 5, 487-496.

Daniels, J.D. and Pettigrew, J.D. Development of neuronal response in the visual system of cats. In G. Gottlieb (ed.) Neural and Behavioral Specificity Studies on the Development

of Behavior and the Nervous System, Vol. 3. New York: Academic Press, 1976, Pp. 195-232.

Donald, S.L. Discrimination of subphonemic phonetic distinctions. Status Report on Speech Research, SR-54 New Haven: Haskins Laboratories, 1978, Pp. 000-000.

Edman, T.R., Soli, S.D., and Widin, G.P. Learning and generalization of inra-phonemic VOT discrimination. Paper presened at the 95th meeting of the Acoustical Society of America, Providence, RI, May 1978.

Eimas, P.D. Developmental aspects of speech perception. In R. Held, H. Leibowitz, and H.L. Teuber (eds.), Handbook of Sensory Physiology, Volume VIII: Perception. New York: Springer-Verlag, 1978, Pp. 357-374.

Klatt, D.H. A cascade-parallel terminal analog speech synthesizer and a strategy for consonant-vowel synthesis, Journal of the Acoustical Society of America, 1977, 61, Suppl. 1, 68(A).

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

Lane, H. L. The motor theory of speech perception: A critical review. Psychological Review, 1965, 72, 275-309.

Lane, H.L. and Moore, D.J. Reconditioning a consonant discrimination in an aphasic: An experimental case history. Journal of Speech and Hearing Disorders, 1962, 27, 3, 232-243.

Lisker, L. On learning a new contrast. Status Report on Speech Research SR-24. New Haven: Haskins Laboratories, 1970. Pp.1-15.

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

Lisker, L. and Abramson, A.S. The voicing dimension: Some experiments in comparitive phonetics. Proceedings of the 6th International Congress of Phonetic Sciences, Prague, 1967.

Pisoni, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. Perception and Psychophysics, 1973, 13, (2), 253-260.

Pisoni, D.B. Auditory short-term memory and vowel perception. Memory and Cognition, 1975, 3, 7-18.

Pisoni, D. B. Some effects of discrimination training on the identification and discrimination of rapid spectral changes. Research on Speech Perception Progress Report, No. 3. Bloomington, Indiana: Indiana University, 1976. Pp. 121-143.

Pisoni, D. B. Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops. Journal of the Acoustical Society of America, 1977, 61, 1352-1361.

Pisoni, D. B. and Lazarus, J. H. Categorical and noncategorical modes of speech perception along the voicing continuum. Journal of the Acoustical Society of America, 1974, 55, 328-333.

Stevens, K.N. The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David, Jr., and P. B. Denes (Eds.), Human Communication: A Unified View. New York: McGraw-Hill, 1972.

Strange, W. The effects of training on the perception of synthetic speech sounds: Voice onset time. Unpublished doctoral dissertation, University of Minnesota, 1972.

Strange, W. and Jenkins, J. J. Role of linguistic experience in the perception of speech. In R. D. Walk and H. L. Pick (Eds.) Perception and Experience. New York: Plenum Press, 1978, Pp. 125-169.

Streeter, L. A. Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. Nature, 1976, 259, 39-41.

Streeter, L. A. Kikuyu labial and apical stop discrimination. Journal of Phonetics, 1976, 4, 43-49.

Streeter, L. A. and Landauer, T. K. Effects of learning English as a second language on the acquisition of a new phonemic contrast. Journal of the Acoustical Society of America, 1976, 59, 448-451.

Walker, H. M. and Lev, J. Statistical Inference. New York: Henry Holt, 1953.

Woodworth, R.S. Experimental Psychology. New York: Holt, 1938.

Footnotes

* The preparation of this paper was supported, in part, by NICHHD grant HD-11915-01 and NIMH grant MH-24027-05 to Indiana University at Bloomington. The paper was written while the first author was a Guggenheim Fellow at the Research Laboratory of Electronics, M. I. T. We thank Diane Kewley-Port for her help and assistance in preparing the synthetic stimuli, Wendy Crawford for help in running subjects and drawing graphs and especially Jerry C. Forshee for his expert advice with regard to the computer facilities used to carry out this work.

Figure Captions

Figure 1. Cross-language identification data from Lisker and Abramson (1967) for labial, apical and velar stops differing in Voice Onset Time from -150 msec to +150 msec.

Figure 2. Average identification functions for two and three category labelling of VOT in Experiment I.

Figure 3. Individual subject data for two and three category labelling from Group A of Experiment I.

Figure 4. Individual subject data for two and three category labelling from Group B of Experiment I.

Figure 5. Average identification and ABX discrimination functions for two category (Group I) and three category (Group II) labelling obtained in Experiment II.

Figure 6. Individual subject data for two category identification obtained from Group I in Experiment II.

Figure 7. Individual subject data for three category identification obtained from Group II in Experiment II.

Figure 8. Average ABX discrimination functions obtained for subjects in Experiment III. The no feedback group (NFB) is shown on the left, the feedback group (FB) is shown on the right.

Figure 9. Individual subject data obtained in ABX discrimination in Experiment III for the no feedback condition.

Figure 10. Individual subject data obtained in ABX discrimination in Experiment III for the feedback condition.

Figure 11. Average identification and ABX discrimination functions for subjects in Experiment IV. Data on the left shows the average function on Day 2. The data on the right shows both average identification and ABX discrimination combined for Days 3 and 4 of the experiment.

Figure 12. Individual subject data for identification obtained on Day 2 and identification and discrimination combined for Days 3 and 4.

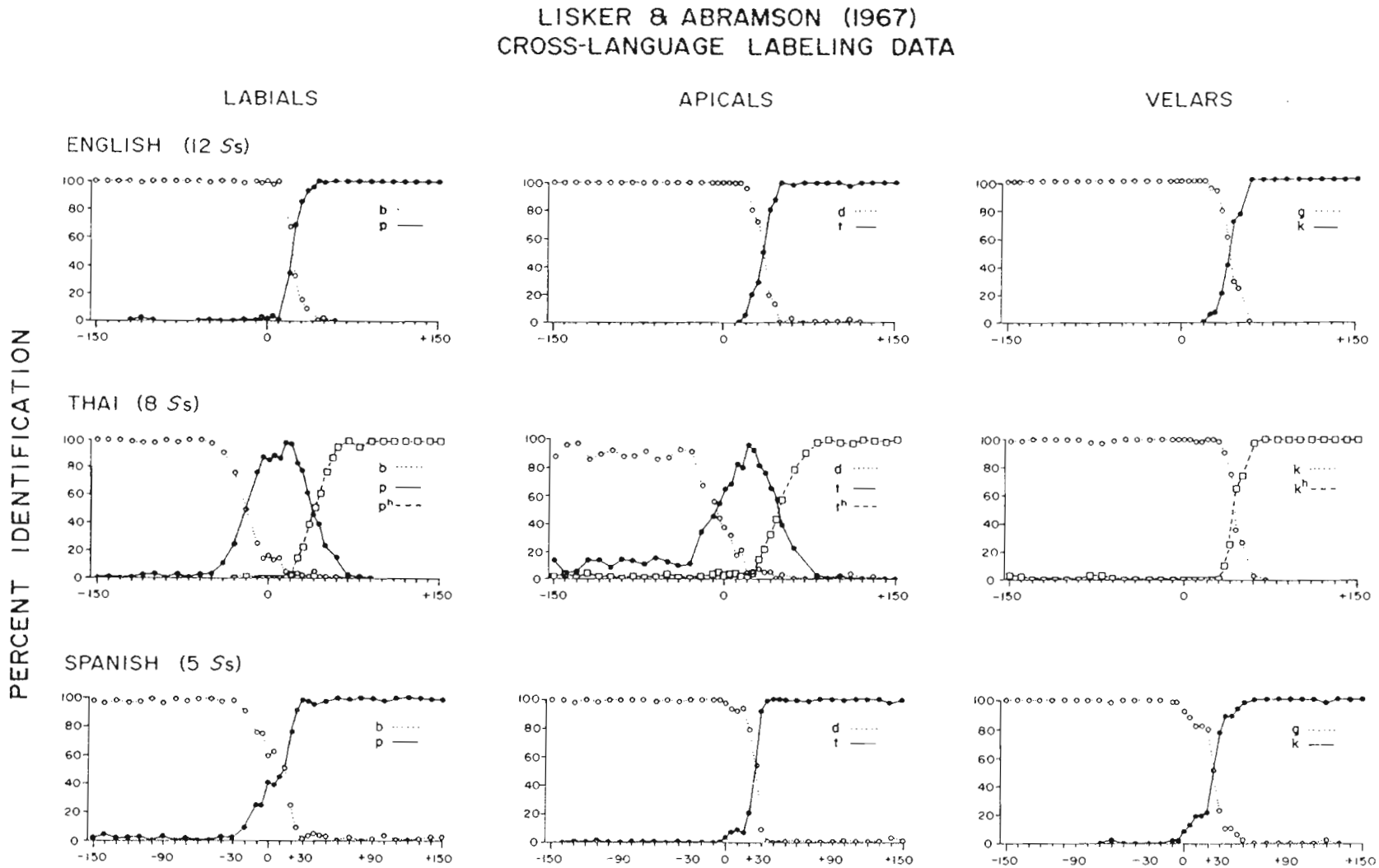

LISKER & ABRAMSON (1967)
CROSS-LANGUAGE LABELING DATA
LABIALS
APICALS
VELARS
ENGLISH (12 Ss)
THAI (8 Ss)
SPANISH (5 Ss)
b
p
d
t
g
k
p^h
t^h
k^h
PERCENT IDENTIFICATION
VOICE ONSET TIME IN MSEC

Figure 1.

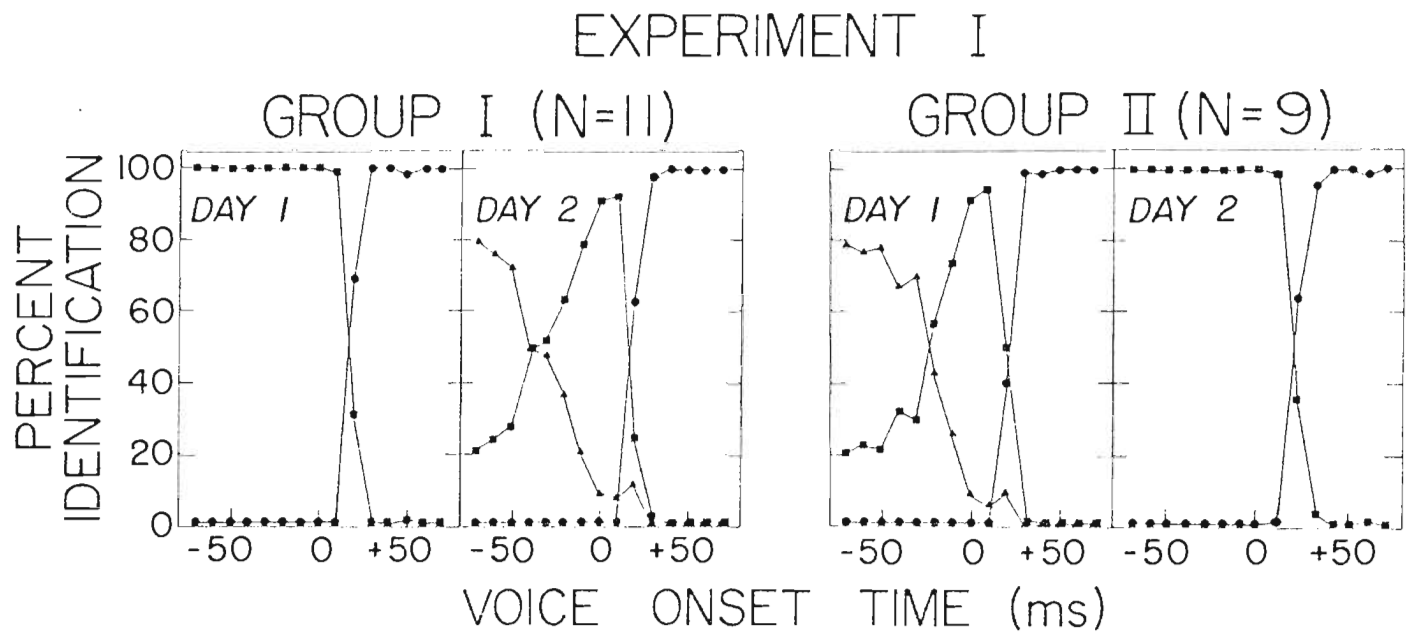

EXPERIMENT I
GROUP I (N=11)
GROUP II (N=9)
DAY 1
DAY 2
DAY 1
DAY 2
PERCENT IDENTIFICATION
100
80
60
40
20
0
-50
0
+50
-50
0
+50
-50
0
+50
-50
0
+50
VOICE ONSET TIME (ms)

Figure 2.

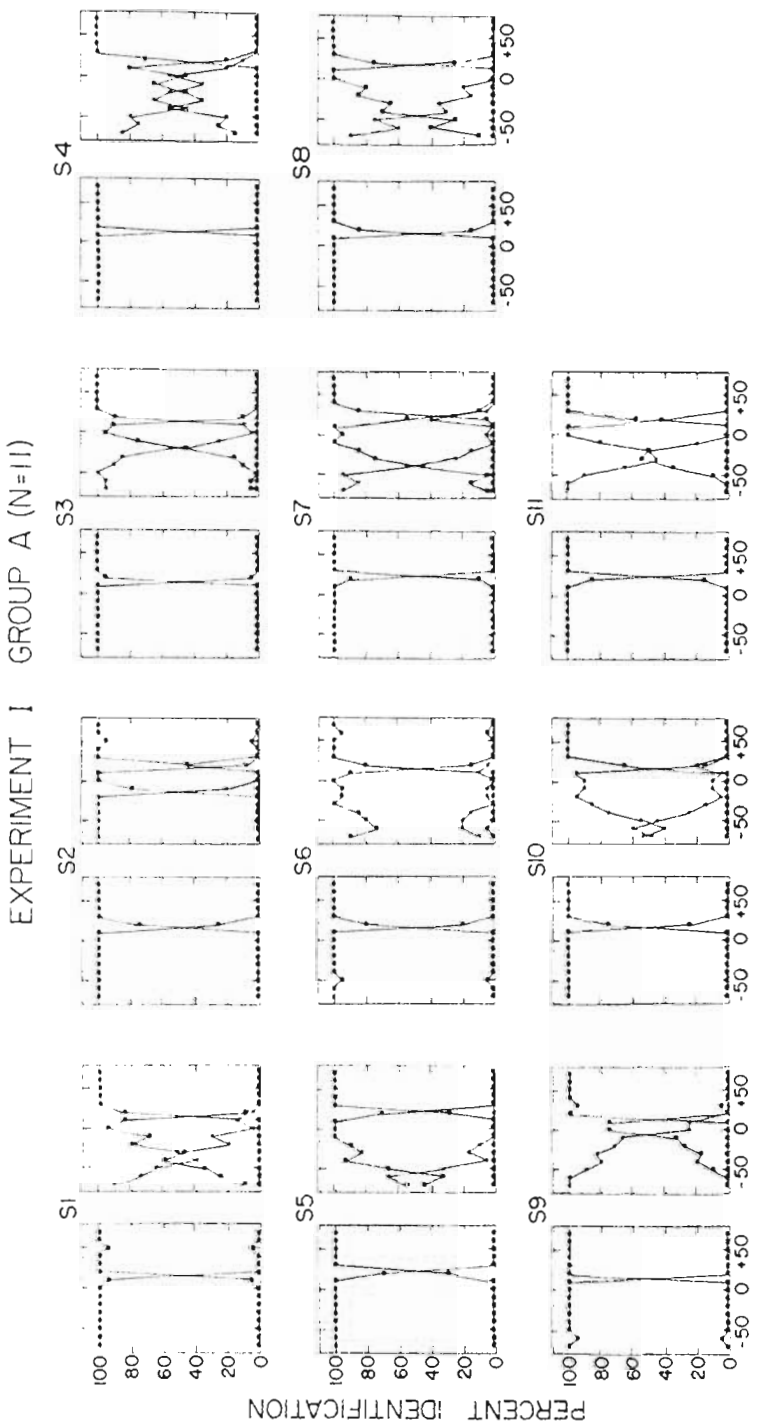

EXPERIMENT I GROUP A (N=11)
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
PERCENT IDENTIFICATION
VOICE ONSET TIME (ms)
100
80
60
40
20
0
-50 0 +50

Figure 3.

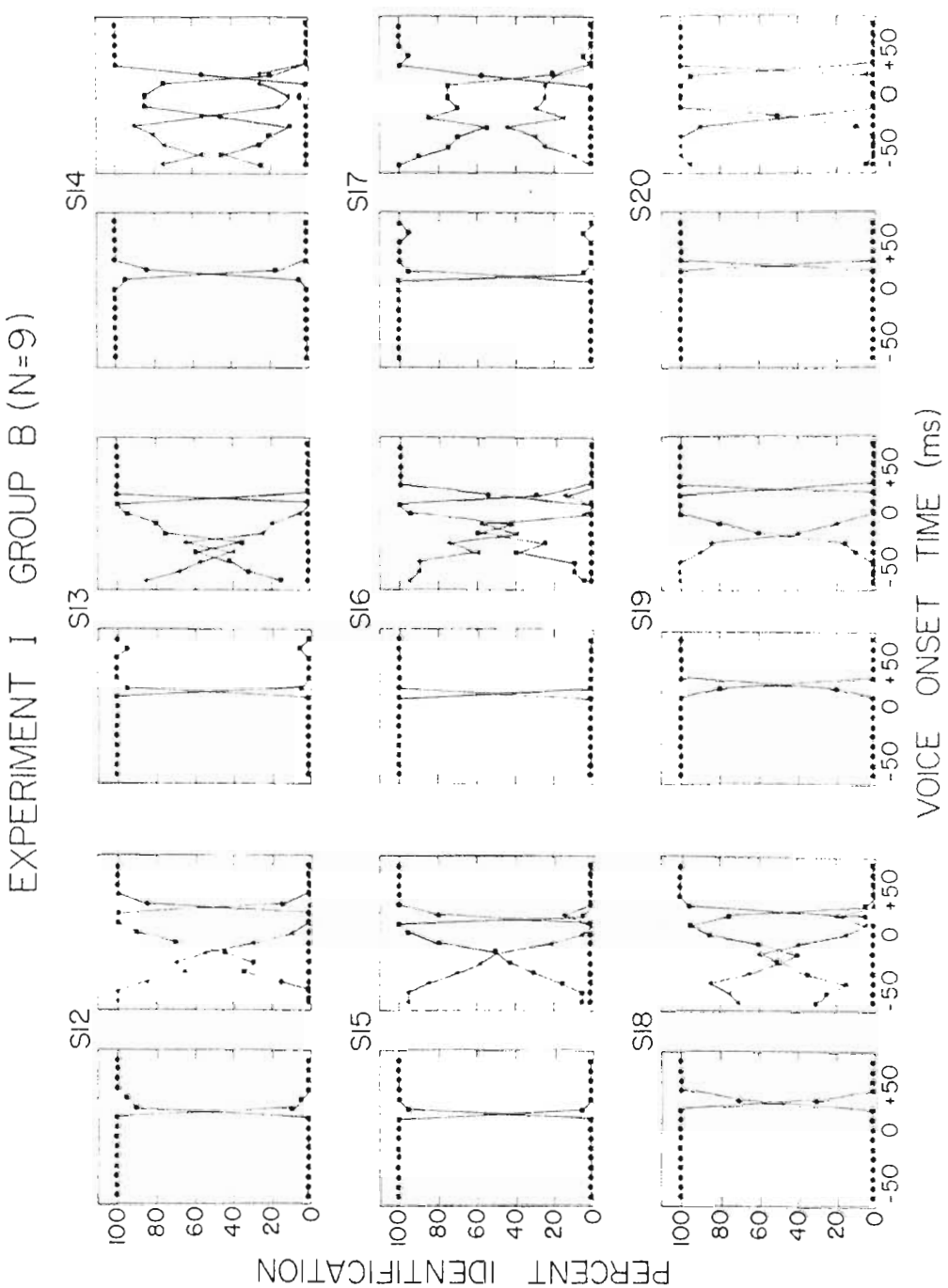

EXPERIMENT I GROUP B (N=9)
S12
S13
S14
S15
S16
S17
S18
S19
S20
PERCENT IDENTIFICATION
100
80
60
40
20
0
-50 0 +50
VOICE ONSET TIME (ms)

Figure 4.

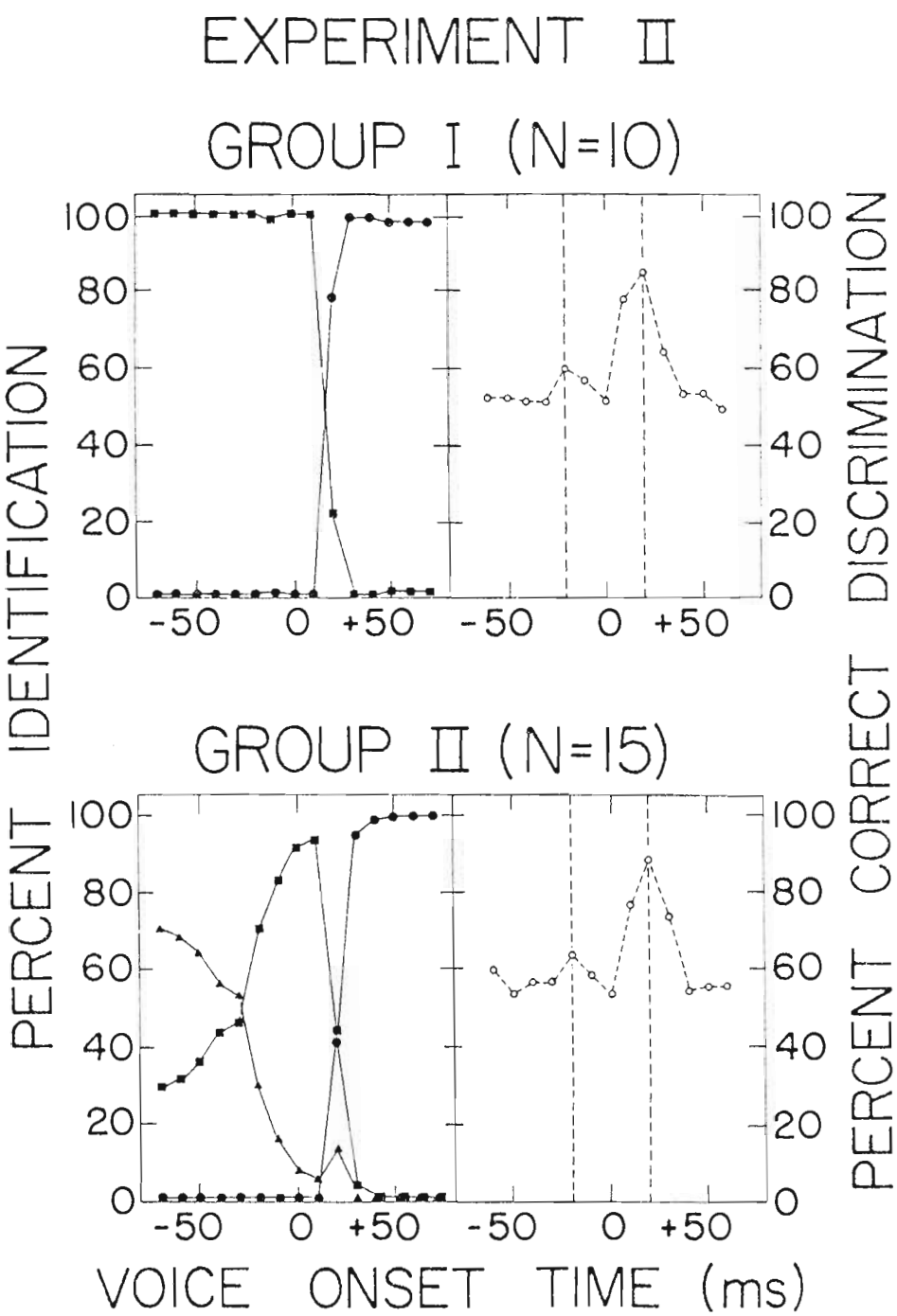

EXPERIMENT II
GROUP I (N=10)
GROUP II (N=15)
PERCENT IDENTIFICATION
PERCENT CORRECT DISCRIMINATION
100
80
60
40
20
0
-50
0
+50
VOICE ONSET TIME (ms)

Figure 5.

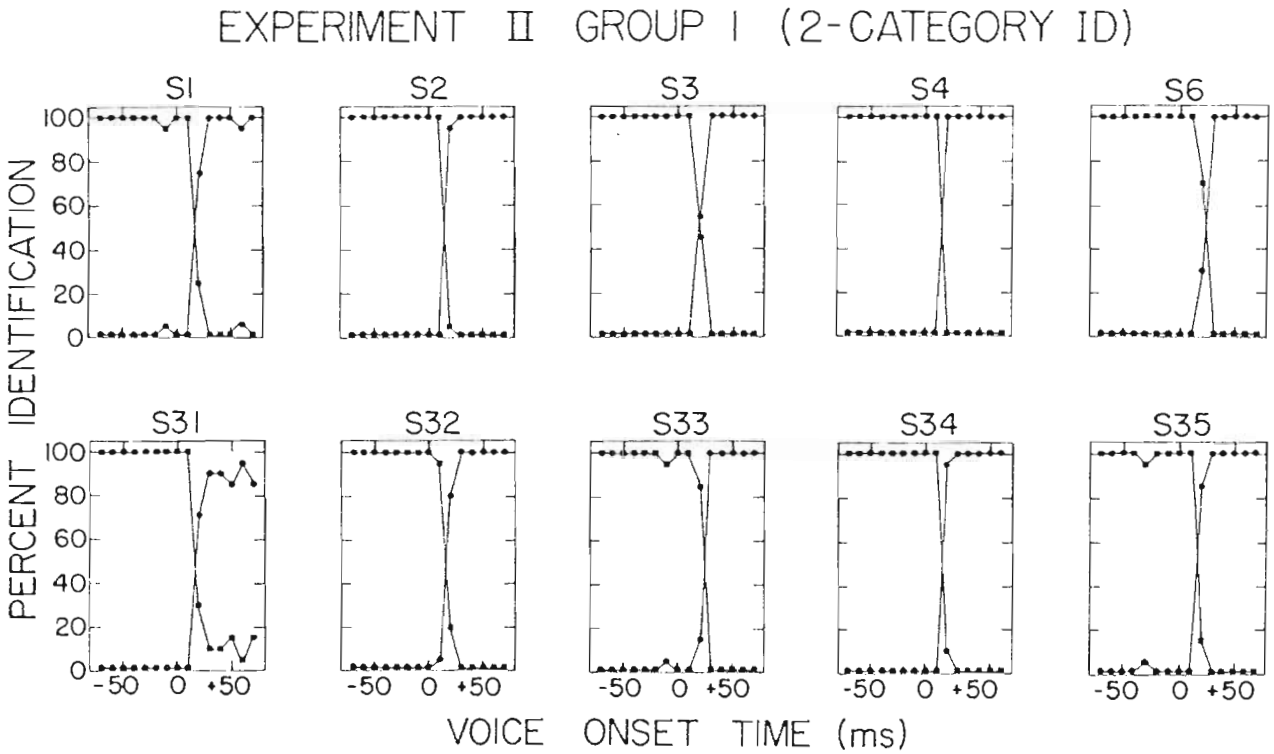

EXPERIMENT II GROUP I (2-CATEGORY ID)
S1
S2
S3
S4
S6
S31
S32
S33
S34
S35
PERCENT IDENTIFICATION
100
80
60
40
20
0
-50
0
+50
VOICE ONSET TIME (ms)

Figure 6.

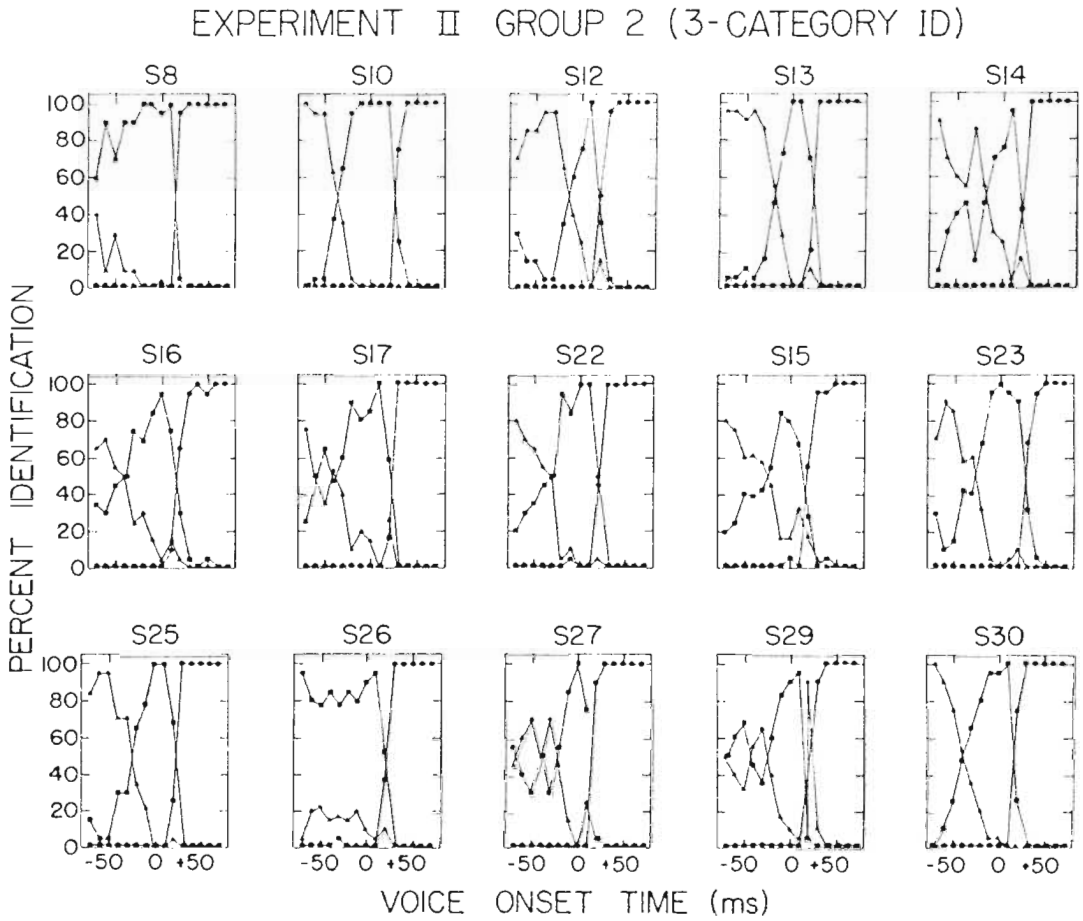

EXPERIMENT II GROUP 2 (3-CATEGORY ID)
S8
S10
S12
S13
S14
S16
S17
S22
S15
S23
S25
S26
S27
S29
S30
PERCENT IDENTIFICATION
100
80
60
40
20
0
-50 0 +50
VOICE ONSET TIME (ms)

Figure 7.

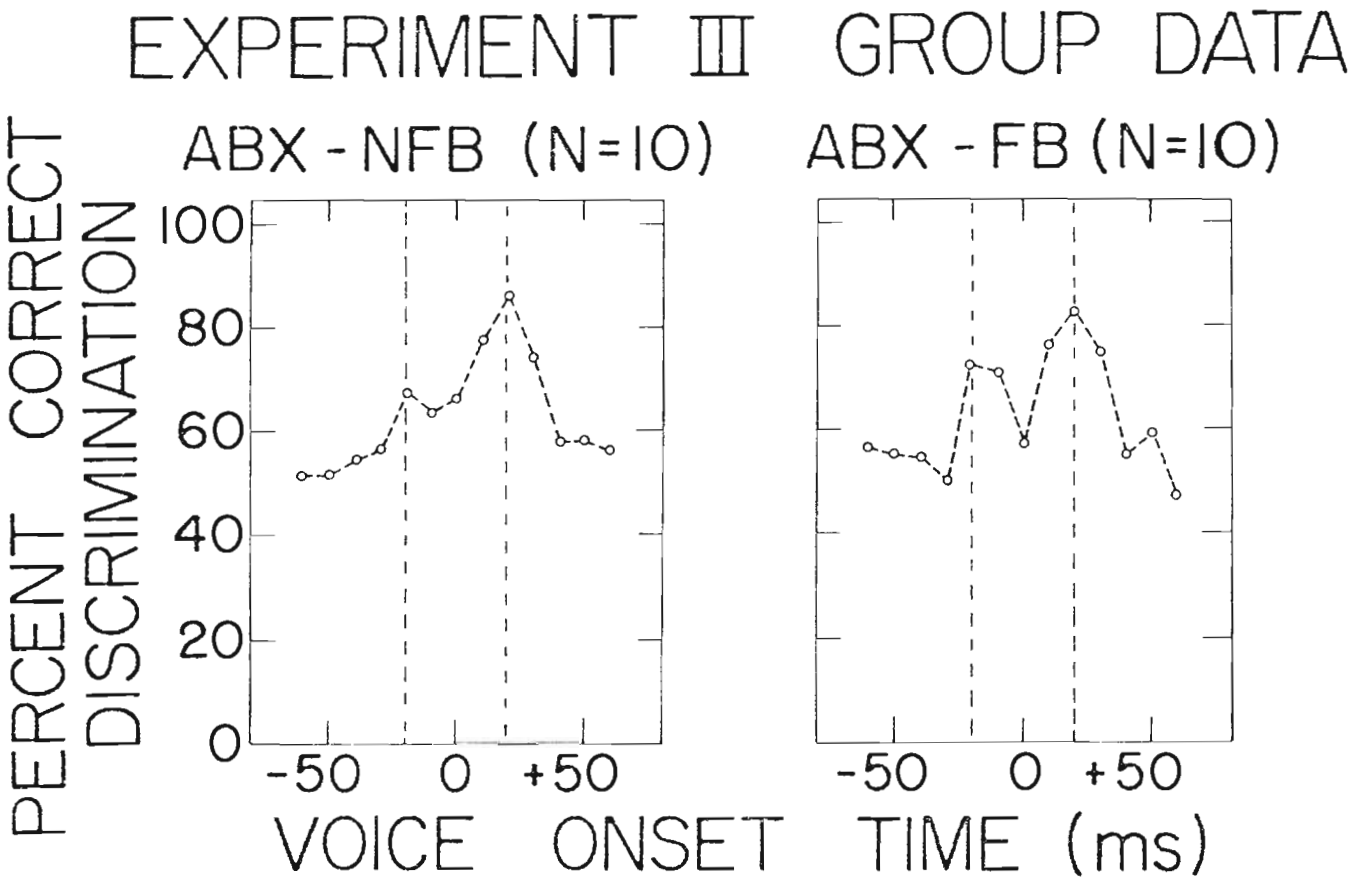

EXPERIMENT III GROUP DATA
ABX - NFB (N=10)
ABX - FB (N=10)
PERCENT CORRECT DISCRIMINATION
100
80
60
40
20
0
-50
0
+50
-50
0
+50
VOICE ONSET TIME (ms)

Figure 8.

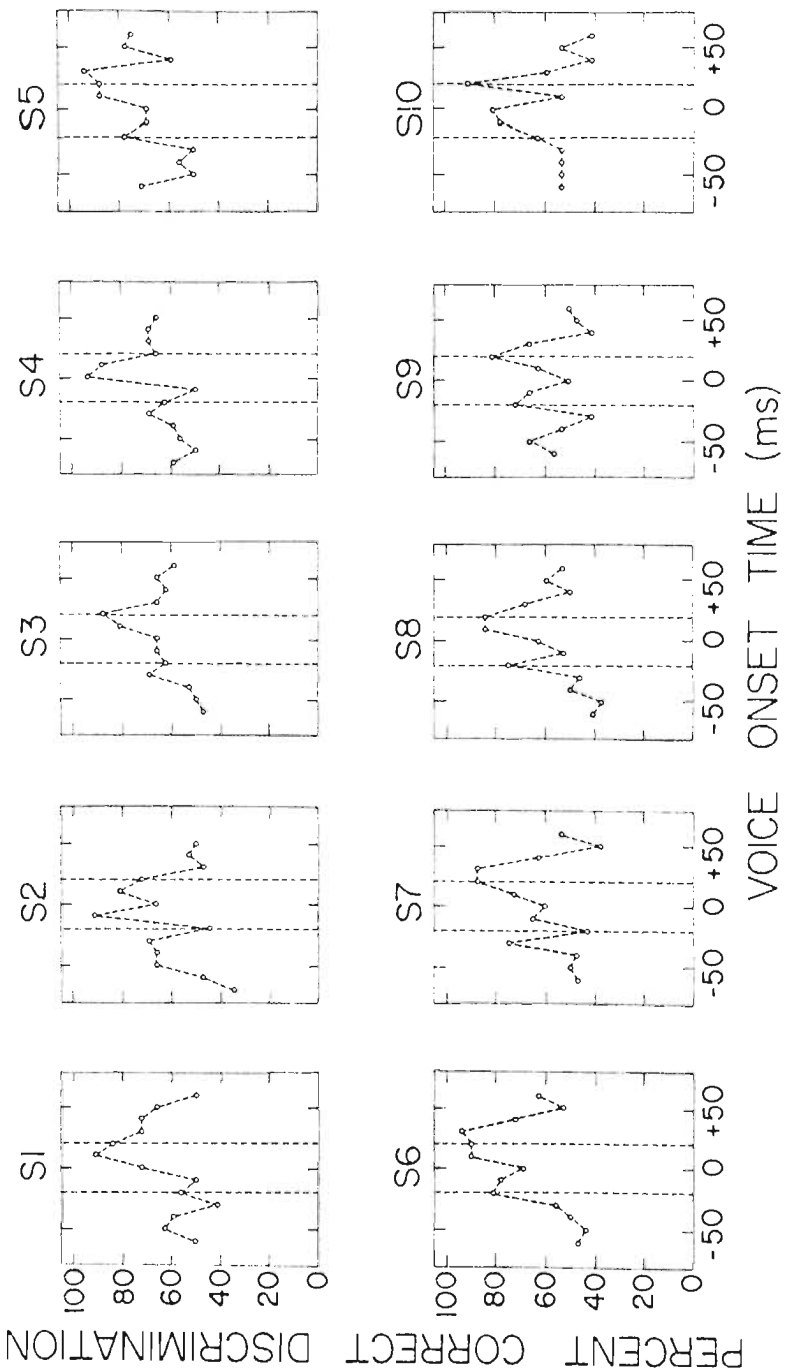

EXPERIMENT III (ABX NFB)
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
PERCENT CORRECT DISCRIMINATION
100
80
60
40
20
0
-50
0
+50
VOICE ONSET TIME (ms)

Figure 9.

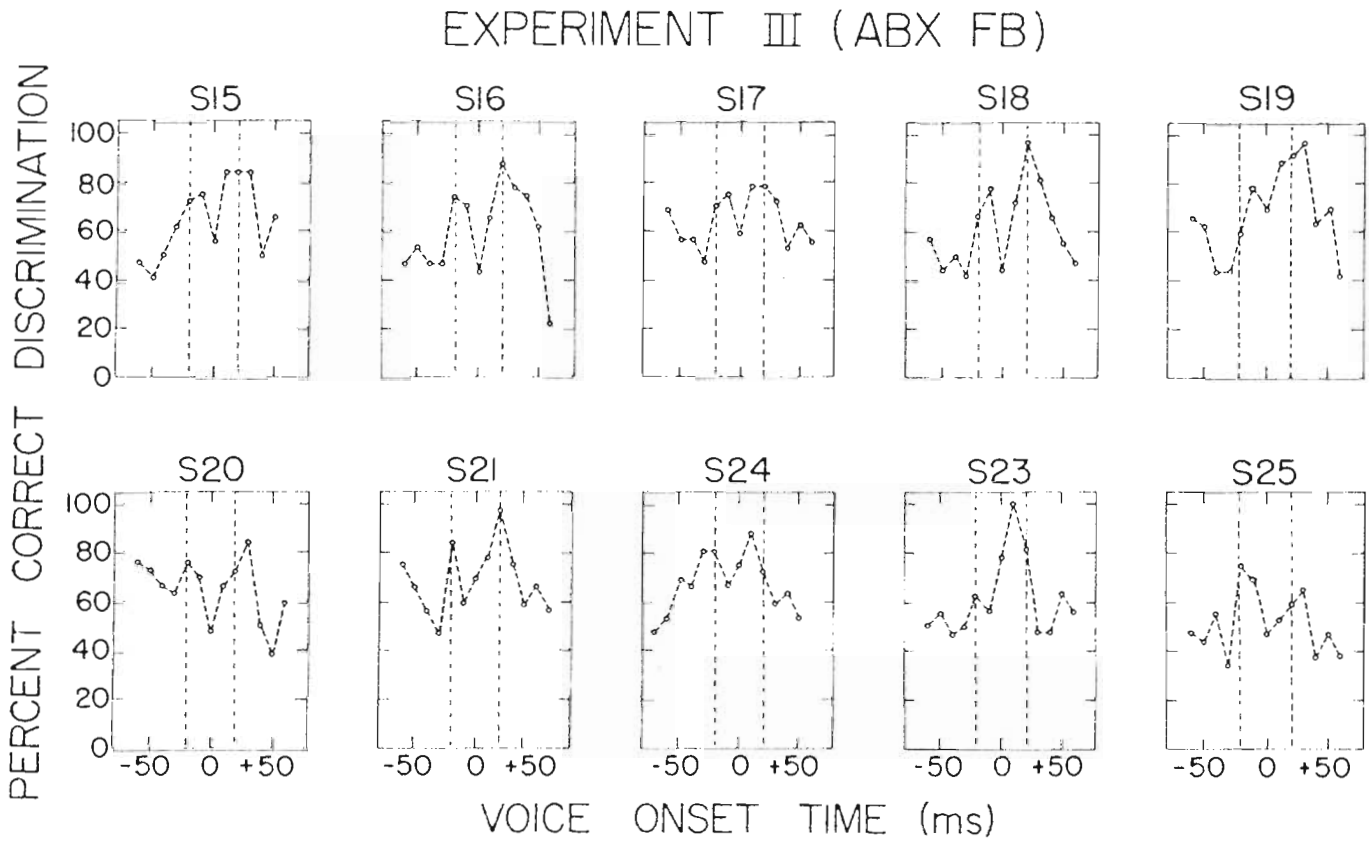

EXPERIMENT III (ABX FB)
PERCENT CORRECT DISCRIMINATION
S15
S16
S17
S18
S19
S20
S21
S24
S23
S25
100
80
60
40
20
0
-50
0
+50
VOICE ONSET TIME (ms)

Figure 10.

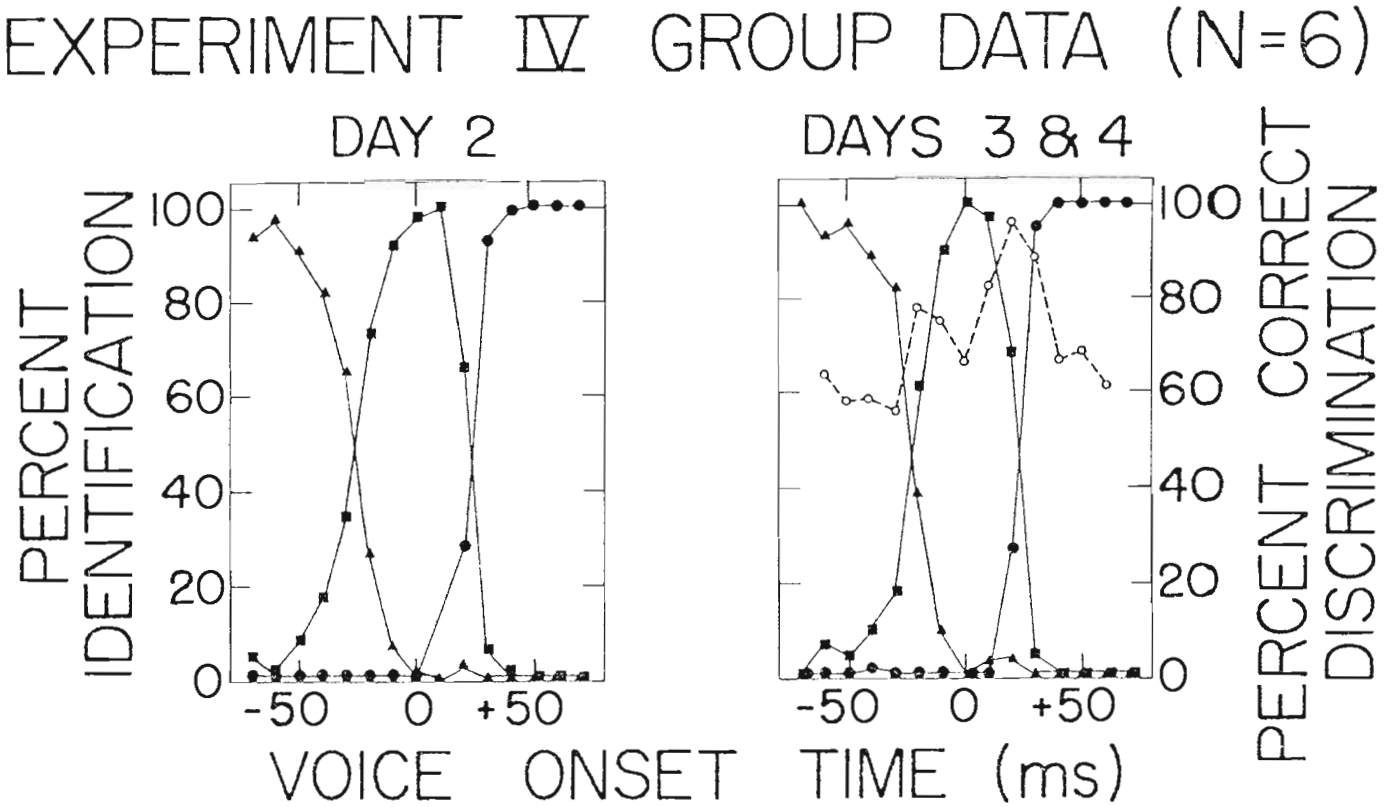

EXPERIMENT IV GROUP DATA (N=6)
DAY 2
DAYS 3 & 4
PERCENT IDENTIFICATION
PERCENT CORRECT DISCRIMINATION
100
80
60
40
20
0
-50
0
+50
VOICE ONSET TIME (ms)

Figure 11.

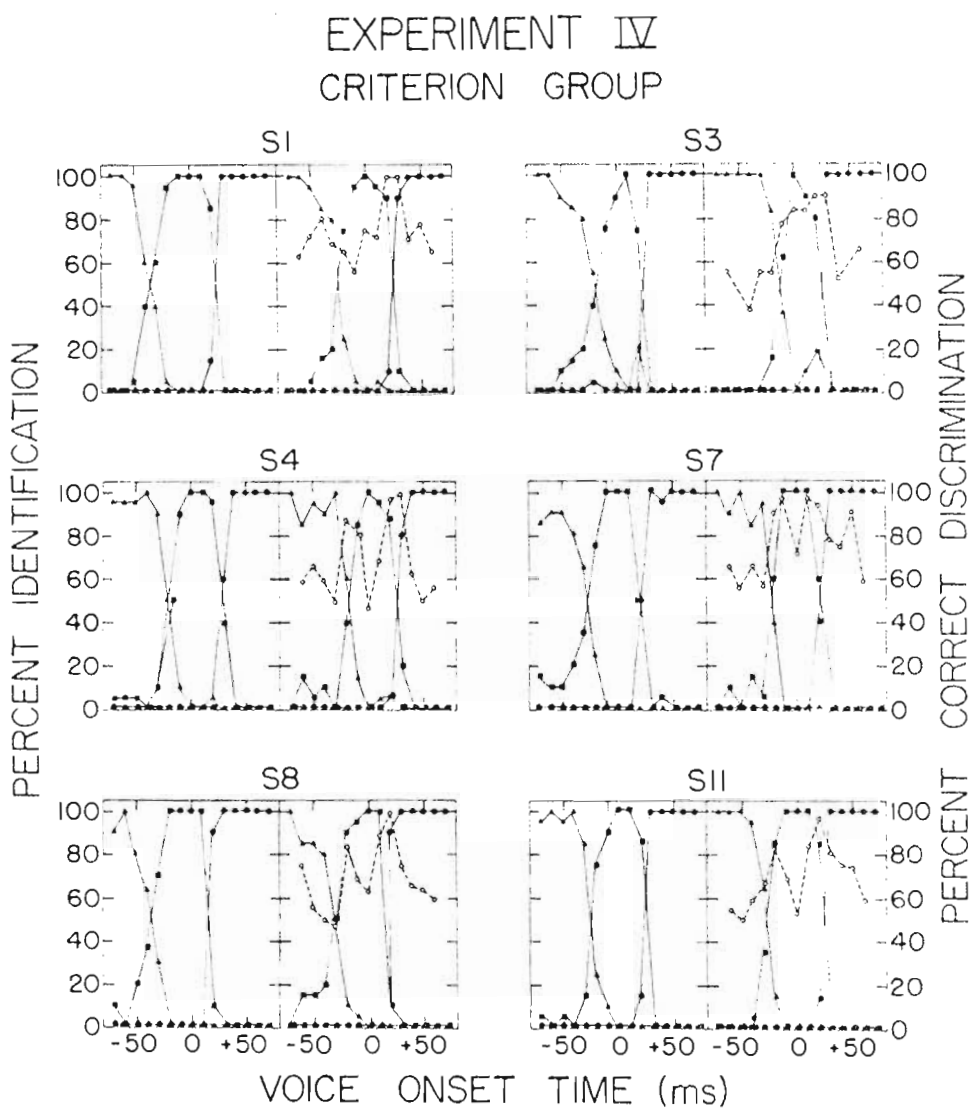

EXPERIMENT IV
CRITERION GROUP
S1
S3
S4
S7
S8
S11
PERCENT IDENTIFICATION
PERCENT CORRECT DISCRIMINATION
100
80
60
40
20
0
-50 0 +50 -50 0 +50
VOICE ONSET TIME (ms)

Figure 12.

Some Developmental Processes in Speech Perception*

Richard N. Aslin and David B. Pisoni

Department of Psychology

Indiana University

Bloomington, Indiana 47401

* This is the final draft of a paper that was presented at the NICHD conference "Child Phonology: Perception, Production and Deviation" which was held in Bethesda, Maryland on May 28-31, 1978. The preparation of this paper was supported by grants from NIMH (MH-24027-04 and MH-30424-01) and NIH (NS-12179-03 and HD-11915-01) to Indiana University at Bloomington. We wish to thank Alan Perey, Beth Hennessy, Natalie Olinger and Wendy Crawford for their help in carrying out the research reported here and Jerry C. Forshee for his efforts and expertise in developing and maintaining the computer facilities used in our laboratories. Finally, we thank Dr. Peter Jusczyk for his advice on an earlier draft of the manuscript.

Some Developmental Processes in Speech Perception

Richard N. Aslin and David B. Pisoni

Indiana University

Bloomington, Indiana 47401

INTRODUCTION

We would like to begin by summarizing the main points that we hope to cover in this presentation. First, we will briefly review several theories that have attempted to explain the processes and mechanisms underlying the development of speech perception in young infants. Second, we will present a conceptual framework from which one can evaluate these theories and the numerous alternative explanations that can be applied to the available empirical findings already reported in the literature. Third, we will describe an account of the processes underlying the development of several segmental contrasts with a special emphasis on the perception of voicing in stop consonants, a distinction that has received considerable attention in the speech perception literature. And finally, we will summarize some recent methodological work from our own laboratory which has been aimed at developing new techniques to investigate the various levels of perceptual analysis that may underlie the perceptual behavior of infants, particularly as these are related to issues surrounding perceptual constancy, feature extraction, and the role of early experience.

The present contribution is not an empirical paper and we do not plan to report the details of new experimental findings at this meeting. Instead, we will focus on some of the general theoretical issues surrounding the development of speech perception within an organized

conceptual framework. In our view, such theoretical efforts are desperately needed at this time and such an undertaking would seem to be especially appropriate at a conference such as this one which is properly concerned with questions about phonological development. Due to time constraints, we will concern ourselves primarily with the most extensively studied class of speech sounds -- stop consonants, particularly those varying in voice-onset-time, although other examples will be referred to from time to time. We believe that our general approach to the problems of perceptual development can be applied to other classes of speech sounds and to other aspects of the phonology of natural languages that must be acquired by children learning language.

BACKGROUND

The now classic experiment reported by Eimas and his colleagues aroused a great deal of interest in the development of speech perception not only because it demonstrated that prelinguistic infants could discriminate synthetic speech sounds categorically, but also because it provided support for the inference that the perceptual categories found in young infants closely matched the phonetic categories of adults. As Eimas, et al. (1971) put it:

> The implication of these findings is that the means by which the categorical perception of speech, that is perception in a linguistic mode, is accomplished may well be part of the biological makeup of the organism... (p. 306)

This conclusion strongly implied, at the time, that categorical perception was not only unique to the perception of speech signals but also that the discriminative behavior of infants was a consequence of perceptual mechanisms that are innately determined. This view of the development

of speech perception, therefore, acknowledged little, if any, influence by non-genetic or experiential factors that might be operative in the early environment of prelinguistic infants.

Although one could argue, as we will shortly, that categorical-like discrimination performance by infants does not necessarily imply the operation of a linguistic mode of processing, there can be little doubt from the available experimental evidence that for certain classes of speech sounds discrimination performance is discontinuous. For example, in the case of speech sounds differing in voice onset time (VOT), there is convincing evidence now for the existence of at least one region of heightened discriminability along the VOT continuum. However, the presence of a discontinuity in discrimination performance does not of necessity imply that discrimination is based on a linguistic level of analysis, since at least one major sensory (nonlinguistic) factor, namely, the discrimination of temporal order,has been proposed to account for VOT discrimination performance (see Hirsh, 1959, Pisoni, 1977). Some seven years later, it is obvious, at least to us, that the implications drawn from Eimas' early experiments were premature.

For example, with regard to the claim that categorical perception is unique to speech signals, several recent reports by Cutting and Rosner (1974), Miller, Wier, Pastore, Kelly & Dooling (1976), and Pisoni (1977) have demonstrated quite conclusively that several classes of non-speech signals which contain "speech-like" acoustic attributes can be perceived categorically by adults. Moreover, Jusczyk,Rosner, Cutting, Foard & Smith (1977) have also demonstrated that young infants show categorical-like discrimination when presented with nonspeech signals differing in rise-time. Thus, the strong contention that categorical perception is unique

to the perception of speech sounds because these signals are perceived as linguistic segments does not appear to have the conclusive support that it once had. It follows, then, that the demonstration of categorical-like discrimination of speech signals in infants is not sufficient evidence either to support the view that speech is perceived by "specialized" perceptual mechanisms or to argue that the infant's discriminative behavior is constrained in a principled way by the phonological structure of any particular natural language. It may simply be the case that infants respond primarily to the psychophysical or sensory properties of speech signals without any subsequent interpretation of these signals as linguistic entities.

The second implication of Eimas, et al.'s conclusion was that the perceptual categories for at least the stop consonants in initial position are innately specified. The basis for this nativistic claim was twofold. First, Lisker & Abramson (1964, 1967) had previously shown in both analysis and synthesis studies that VOT is a sufficient measure to characterize the voicing and aspiration differences among the stop consonants that exist in a number of diverse languages. From the study of stops in 11 languages Lisker & Abramson (1964) found that the various tokens were distributed at one of three modal values along the VOT continuum corresponding to long lead, short lag and long lag distinctions in voicing. Although the precise locations of the boundaries between phonological categories for stops differ somewhat from language to language, Lisker & Abramson (1964) suggested that the dimension of VOT was, in all likelihood, a universal and therefore closely tied to the biological basis of language and speech. Second, in their infant study Eimas

et al. had found evidence of categorical perception of stops differing in VOT in the short lag region of the continuum--- a region used by all languages to signal voicing differences, including English. The unusually close correspondence between the infants' discrimination performance on synthetic VOT stimuli and the Lisker & Abramson adult English data permitted Eimas et al (1971) to argue that prelinguistic infants are preattuned to process speech sounds in a linguistically relevant manner, a manner approximating the categorical perception of stops observed in adults. More recently Eimas (1975) has summarized the earlier findings as follows:

> Given the considerable overall similarity of the adult and infant discriminability data, infants, apparently, also have access to a phonetic feature code for purposes of deciding whether two speech events are the same or different. (p. 341)

However, despite the apparently sound reasoning behind these nativistic views of Eimas (1975), several important empirical findings were apparently overlooked. First, as noted earlier, the precise location of the voicing boundaries described by Lisker & Abramson differ somewhat from language to language suggesting that some fine tuning or alignment will take place during perceptual development. Second, Lasky, Syrdal-Lasky & Klein (1975) tested infants raised in a Spanish-speaking environment and found evidence of discrimination of voicing contrasts that are <u>not</u> discriminated by Spanish-speaking adults. They also failed to find evidence of discrimination of the contrasts that cross the Spanish adult voicing boundary. These latter findings therefore led Eimas (1975) to conclude that "the ability to perceive voicing distinctions in accord with a phonetic feature code during early infancy is independent of the

infant's linguistic environment (p. 341)."

Thus, on the one hand, Eimas used the correspondence of the English infant-adult discrimination data to support his claim that the voicing categories are linguistically relevant and therefore genetically specified; but, on the other hand, Eimas did not appear to see a conflict in the lack of correspondence between the Spanish infant-adult discrimination data. This line of reasoning raises the interesting question, in our minds, of how past accounts of infant speech perception might have evolved if the Spanish infant discrimination data had been published first. Although the nativistic account of speech perception proposed by Eimas has been a significant advance over previous views of perceptual development which have assumed that speech production preceeds or parallels perception (e.g., Fry, 1966), we still see numerous problems in the logic underlying current conceptualizations of the development of speech perception. Recent empirical findings from a number of sensory and cognitive domains have no doubt broadened the perspective from which researchers can conceptualize the process of perceptual development. However, there is still a very strong tendency toward theoretical simplification of issues in infant speech perception,either in vague terms of learning or by recourse to nativistic accounts of perceptual development. Our discussion below has been motivated chiefly by these theoretical considerations.

The last implication of Eimas' study was that a genetic specification of a perceptual category and its boundaries, although not modifiable during early infancy, may be modified during later childhood. Eimas (1975) obviously realized this modification must occur and even discussed it in light of his earlier results:

> ...a strong genetic determination of phonetic categories and boundaries...does not actually preclude modifications of the mechanisms underlying this categorization of speech. Indeed, the data of Lasky et al. demand that modifications in the loci of the phonetic boundary of infants from Spanish environments occur if there is to be effective communication. (p. 342)

What has remained unclear even to this time is why the perceptual categories for English infants are specified so accurately by genetics? One possible answer to this question comes from the previously mentioned nonspeech categorical perception findings obtained with adults and from recent research using animal models to study the perception of species-specific acoustic signals.

Kuhl and Miller (1975, 1978) have shown that chinchillas, who obviously do not make use of a human voicing distinction in their own vocal repertoire, can be trained to respond consistently to synthetic labial stop consonant stimuli. Moreover, the perceptual boundaries of the chinchilla correspond quite closely to the category boundaries found for voiced-voiceless stops in English adults. These results have raised the question of what process or level of perceptual analysis might be responsible for the categorical-discrimination performance of both human adults and infants, and whether these results can be accounted for in a principled way by recourse to a linguistic mode of processing.

At the present time it is equally unclear whether the boundaries in human infants and adults undergo a selective modification developmentally as a result of particular linguistic input in their environment and what perceptual mechanisms are responsible for this modification. Clearly, there must be some selectivity in the course of phonological development as evidenced by the fact that different languages have different phonologies,

and by the apparent difficulty adults have in recognizing phonetic contrasts which are phonologically irrelevant in their native language. The now classic cross-language work of Lisker and Abramson (1967) has supported the contention that only phonologically distinctive perceptual categories are perceived by adults. A summary of their results for English, Thai and Spanish subjects is shown in Figure 1 to illustrate the significant role that linguistic experience plays in the categorization of speech signals.

Insert Figure 1 About Here

More recent cross-language research by Miyawaki, Strange, Verbrugge, Liberman, Jenkins & Fujimura (1975) also provides support for the view that the phonologically distinctive [r]-[l] contrast in English is perceived by English adults but not by Japanese adults who do not have the [r]-[l] contrast in their language. Training studies such as those summarized recently by Strange and Jenkins (1977) suggest further that, although the ability to perceive a phonologically irrelevant contrast may have been present at birth, adults who have lost or failed to develop that contrast are probably incapable of acquiring or reacquiring it. Such findings could be interpreted as evidence that a neural substrate for the perception of phonologically irrelevant contrasts either failed to be formed during a critical or sensitive period or atrophied as a result of the absence of experience with that contrast. This neural theory of phonological development is passive in the sense that it assumes that little or no involvement, either attentional or productive, is required to maintain or create a particular perceptual ability. As we shall see in a later section, such views of sensory development have not received strong support in recent years.

However, an alternative to this passive or strictly receptive account of the role of early experience in speech perception is the view that the failure to actively engage an attentional or productive system in the use of a particular phonetic contrast only depresses or attenuates subsequent performance on that linguistic contrast. The difficulty shown by adults in discrimination, then, may not be due to any neural process per se but may be simply a consequence of an attentional deficit similar to the process of acquired equivalence -- a perceptual mode that involves learning to ignore distinctive differences among stimuli. This view assumes that perception of the relevant distinctive contrasts is so automatic as a result of previous processing strategies acquired by the subject that reacquisition of a phonologically-irrelevant contrast is difficult to obtain reliably in untrained adults (see Shiffrin and Schneider, 1977).

To study these questions in more detail, we recently collected some preliminary data from adult subjects that supports the predictions of the attentional deficit model outlined above. Figure 2 shows labeling data from four adult subjects who were given two repetitions of a synthetic

---
Insert Figure 2 About Here
---

prevoiced /ba/ with a VOT value of -70 msec prior to a forced-choice identification task. In one condition of the experiment, the subjects had three response buttons corresponding to [ba], [pa], and prevoiced [ba]. Note that all subjects were highly consistent in labeling three perceptual categories despite the absence of highly prevoiced stops in initial position in English and the very limited exposure and training experience that preceded the labeling task. These findings are particularly striking when compared with

the more traditional two-alternative forced-choice labeling results shown in the right-hand panel for each of these subjects. Note the classic two-category identification functions obtained in this task for /ba/ and /pa/ responses. Our recent findings suggest, therefore, that phonologically-irrelevant categories can be consistently categorized by adults even without very extensive training and presumably without significant neural loss or atrophy of the feature detectors which have been assumed to underlie phonetic categorization (Eimas, 1975). Such findings call into question the recent conclusions of Strange and Jenkins (1977) concerning the effects of laboratory training studies in speech perception. Moreover, given that subjects could use three responses consistently and without feedback in this task, it is difficult to argue that there was any "selective loss" in perceptual sensitivity of these subjects in processing voicing information. The performance decrements observed in earlier studies on voicing discrimination may simply be the result of criteria shifts and response constraints resulting from the use of different subject strategies in these tasks (Pisoni & Lazarus, 1974; Carney, Widin & Viemeister, 1977).

ROLE OF EARLY EXPERIENCE IN PERCEPTUAL DEVELOPMENT

The need for a coherent framework from which to view the course of perceptual development is of the utmost importance to our understanding of the processes underlying the development of speech perception, especially in light of the many seemingly diverse and contradictory empirical findings that have appeared in the infant perception literature in recent years. Several researchers working in the area of visual system development have begun to appreciate the many potential and seemingly diverse roles that genetic and experiential factors can play in the development of sensory

and perceptual systems. For example, some of the neural mechanisms underlying visual functioning are not present at birth. Moreover, they do not unfold during development as a simple result of a genetically controlled plan or schedule. In other words, early visual experience influences to some extent the course of visual system development. Yet early experience does not totally control the outcome of visual system development since some genetically specified limits are clearly placed on how much early experience can influence the course of visual system development (see Blakemore, 1976; and Grobstein & Chow, 1976 for general reviews).

The research of Hubel and Wiesel (1965, 1970) provides a good example to illustrate how complex the interactions are between genetic and experiential factors in visual system development. They have shown that kittens who have been selectively deprived of certain types of early visual experiences fail to develop the normal neural mechanisms subserving binocular vision. Moreover, they have found evidence for a relatively well-defined sensitive period in binocular development, as shown in the top panel of Figure 3, during which early visual deprivation exerts its most significant and permanent effects. Yet Hubel & Wiesel (1963) also showed that at least part of the neural mechanism underlying binocular vision <u>is</u> present at birth, and this mechanism does not deteriorate if the kitten is reared in total darkness. That is, the absence of visual experience (dark rearing) does not eliminate binocular function, whereas the presence of a binocular imbalance (monocular occlusion) does eliminate binocular function by creating competition between the inputs from the two eyes.

These findings from the animal literature have been extended and generalized to the study of humans who were deprived of certain visual

---

Insert Figure 3 About Here

---

experiences in early life. Banks, Aslin and Letson (1975) have reported, as displayed in the lower panel of Figure 3, that the development of the human visual system is also characterized by a sensitive period during which selective binocular deprivation can lead to permanent and irreversible deficits in binocular functioning, in particular, depth perception. In contrast, other studies (Creel, Witkop & King, 1974; Banks & Aslin, 1975) have shown that some humans have a genetic anomaly in their visual system associated with a condition of albinism. These individuals fail to develop normal binocular function irrespective of the presence or absence of binocular deprivation during early life.

Thus, it is clear from the study of visual system function and its development that a simple dichotomy between nativistic and empiricist accounts of the process of development is inadequate to capture the multiple and seemingly complex genetic and environmental interactions that underlie normal perceptual development. In the next section, we turn to a discussion of some of the intricacies of the genetic-environmental interactions that must be accounted for by any theory of perceptual development, including theories of the development of speech perception.

Recently Gilbert Gottlieb (1976a,b), a behavioral embryologist, has provided an account of some of the possible roles that early experience can play in the development of sensory systems. His conceptualization of these processes seemed to us to be particularly relevant to discussions of the development of speech perception. According to Gottlieb (1976a,b) there are four basic ways in which early experience could influence the development

of a perceptual ability. These alternatives are illustrated in Figure 4.

Insert Figure 4 About Here

First, a perceptual ability may be present at birth but require certain specific types of early experience to maintain the integrity of that ability. The absence or degradation of the required early experience can result in either a partial or a complete loss of the perceptual ability, a loss which may be irreversible despite subsequent experience. For example, as mentioned previously, the work of Hubel and Wiesel (1965, 1970) on the visual system of the kitten showed, among other things, that the full complement of neural cells responsible for binocular vision was present at birth, although they lost their function if the kittens were deprived of binocular vision during the sensitive period. Thus, early experience in this case served to maintain the functional integrity of the mechanisms underlying binocular vision.

Secondly, an ability may be only partially developed at birth, requiring specific types of early experience to facilitate or "attune" the further development of that perceptual ability. The lack of early experience with these stimuli which serve a facilitating function could result either in the absence of any further development or a loss of that ability when compared to its level at birth. As an example of a facilitating function, we can cite the work of Gottlieb himself who has shown that ducklings modify their subsequent preference and recognition of species-specific calls by their own vocalizations prior to and shortly after hatching (Gottlieb, 1976a). If these self-produced vocalizations are prevented from occurring (through devocalization techniques) while still in the early stages of development,

the developmental rate of preference for species-specific calls declines and the ability to discriminate and recognize particular calls is substantially reduced (Gottlieb, 1975).

Third, a perceptual ability may be absent at birth, and its development may depend upon a process of induction based on specific early experiences by the organism. The presence of a particular ability, then, would depend to a large extent upon the presence of a particular type of early experience. For example, specific early experiences presented to young ducklings leads to imprinting to a particular stimulus object and can be taken as an instance of inducing a behavioral preference (Hess, 1972). Thus, in this case the presence of a particular early experience is necessary for the subsequent development of a particular perceptual preference or tendency.

Finally, of course, early experience may exert no role at all in the development of a particular perceptual ability. That is, the ability may be either present or absent at birth and it may remain, decline or improve in the absence of any particular type of early experience. Absence of an experiential effect is particularly difficult to identify and often leads to unwarranted conclusions, particularly conclusions that assume that an induction process might be operative. For example, it is quite common for investigators to argue that if an ability is absent at birth, but then observed to be present sometime after birth, the ability must have been learned. In terms of the conceptual framework outlined above, this could be an example of induction. Yet it is quite possible that the ability simply "unfolded" developmentally according to a genetically specified maturational schedule, a schedule that required no particular type of early

experience in the environment. This unfolding of an ability may be thought of as an example of the general class of maturational theories of development. As an example, although general motor activity is necessary to prevent the atrophy of particular muscle systems, many of the classic studies by Gesell in the 1930s demonstrated that no particular training experience was necessary for infants to acquire the ability to walk (Gesell & Ames, 1940). Thus, as we have tried to show, the complexity of these numerous possible alternatives -- maintenance, facilitation, induction, and maturation -- and their possible interactions should caution any rash or premature conclusions regarding the developmental course of specific perceptual abilities.

But what then is the specific relevance of Gottlieb's scheme of the roles of early experience to the development of speech perception? We would like to outline four general classes of theories of perceptual development that are, in our view, appropriate to discussions of phonological development. After we describe the assumptions underlying these four classes of theories, we will then select several examples from the available literature on infant speech perception to illustrate the usefulness of this conceptualization. The classes of theories of perceptual development we consider below are what we have called Universal theory, Attunement theory, Perceptual Learning theory and Maturational theory.

Universal Theory assumes that, at birth, infants are capable of discriminating all the possible phonetic contrasts that may be used phonologically in any natural language. According to this view, early experience functions to <u>maintain</u> the ability to discriminate phonologically relevant distinctions, those actually presented to the infant in the environment. However, the absence of phonologically-irrelevant contrasts, which are

obviously not presented to the infant, results in a selective "loss" of the abilities to discriminate those specific contrasts. The mechanisms responsible for this loss of sensitivity may be either neural or attentional or both. These two alternatives also make several specific predictions concerning the possible reacquisition of the lost discriminative abilities in adults, a topic of some interest in its own right, as we mentioned in an earlier section of this paper.

Attunement Theory assumes that at birth all infants are capable of discriminating at least some of the possible phonetic contrasts contained in the world's languages, but that the infant's discriminative capacities are incompletely developed and/or possibly quite broadly tuned. Early experience therefore functions to align and/or sharpen these partially developed discriminative abilities. Phonologically-relevant contrasts in the language-learning environment would then become more finely tuned with experience and phonologically-irrelevant contrasts would either remain broadly tuned or become attenuated in the absence of specific environmental stimulation.

In contrast with the other two views, Perceptual Learning Theory assumes that the ability to discriminate any particular phonetic contrast is dependent upon specific early experience with that contrast in the language-learning environment. The rate of development could be very fast or very slow depending on the frequency of occurrence of the phonetic contrasts during early life, the relative acoustic or psychophysical discriminability of the contrast compared with other contrasts, and the attentional state of the infant. According to this view, however, phonologically-irrelevant contrasts would <u>never</u> be discriminated better than the phonologically-relevant contrasts present in the language-learning environment.

Finally, Maturational Theory assumes that the ability to discriminate a particular phonetic contrast is independent of any specific early experience and simply unfolds according to a predetermined developmental schedule. All possible phonetic contrasts would be discriminated equally well irrespective of the language environment, although the age at which specific phonetic contrasts could be discriminated would be dependent on the developmental level of the underlying sensory mechanism.

These classes of theories of perceptual development make rather specific predictions concerning the developmental course of speech perception in infants and young children, predictions that we think are of special importance to the participants of this conference. It is important to note here that we are not claiming that only one of these classes of theories will uniquely account for the development of all speech contrasts. Rather, it may be the case that some hybrid of the theories provides the best description of the development of specific classes of speech sound discrimination. In fact, this view of parallel developmental processes appears to be supported by current empirical findings, as we hope to show below. In the remainder of this section, we will first summarize several of the empirical findings already available in the literature and then attempt to provide an account of these findings within the context of the framework outlined above. However, before proceeding to the empirical findings, it is appropriate to state rather explicitly what our goals are in trying to account for the data in this manner.

One of the key issues involved in a proper understanding of the development of speech perception is the level of analysis presumed to be operative in the processing of speech signals. In our view, the level of analysis

issue can ultimately be reduced to two basic alternatives, a sensory or psychophysical level and a phonetic or interpretive level. In the past, a phonetic level of analysis was strongly implicated as the basis for the infant's discrimination of various classes of speech signals, particularly stop consonants. As we noted earlier, the recent findings with adult subjects using nonspeech signals and the findings from non-human animals have raised the strong possibility that the discriminative behavior observed in the earlier adult experiments as well as the infant speech perception experiments may be based, to a large extent, on a sensory level of analysis which involves responding to the psychophysical attributes of the speech signals. Although both approaches -- the use of complex nonspeech signals with adults and infants and the cross-species comparisons -- have helped to broaden our understanding of the adult and infant literature on speech perception, additional evidence for deciding on the particular level or levels of analysis has come from studies of the discrimination of phonetic contrasts that are phonologically-irrelevant for a particular group of language-learning infants -- that is, the cross-language infant speech perception studies. By comparing an infant's discrimination of both phonologically-relevant and phonologically-irrelevant phonetic contrasts, we can gain information regarding not only the specific level of coding of the sensory input but also have an opportunity to examine several of the issues surrounding the role of early experience and the processes involved in perceptual development.

Voicing in Stop Consonants. In the more than seven years since Eimas, et al's study, over two dozen VOT contrasts have been studied in infants. Positive evidence of discrimination has been obtained for all contrasts that

crossed the English voiced-voiceless boundary. However, for contrasts that crossed a prevoiced-voiced boundary, the only positive evidence of discrimination was obtained with infants whose native language environment contained a phonological contrast between pre-voiced and voiced stop consonants.

At first glance these results on the discrimination of voicing contrasts by infants might appear to provide strong support for a Perceptual Learning explanation, although certain key findings are clearly in conflict with the predictions of that theory. For example, several contrasts were tested on infants whose native language environment was not English. Discrimination performance on the majority of these contrasts was observed despite the fact that these contrasts were phonologically-inappropriate and unlikely to occur in the language learning environment. That is, infants discriminated a contrast that their parents presumably never use in spoken language. However, other studies of VOT have failed to show evidence that infants discriminate contrasts that are present in their language learning environment. How can we reconcile these seemingly contradictory results? Within the conceptual framework outlined earlier, we think it is possible to offer a systematic account of these findings in terms of what is currently known about the psychophysical properties of these speech signals and the underlying developmental process responsible for realizing the discrimination.

We propose the following account of the development of the perception of voicing contrasts as cued by VOT in stop consonants. First, there is now sufficient evidence to suggest that the basis for VOT discrimination by infants is probably <u>not</u> directly related to phonetic categorization or a linguistic mode of analysis (see also Stevens & Klatt, 1974). Rather, we

would argue that the discrimination performance of infants tested on VOT contrasts is based on the detection of the relative onsets between two acoustic events; that is, in the case of VOT, the detection of the onset of the first formant relative to the onset of higher formants (Pisoni, 1977). We would suggest further that the discrimination of the relative order between these two events is more highly discriminable at certain regions along the VOT stimulus continuum corresponding roughly to the location of the threshold for resolving these differences psychophysically. In the case of temporal order processing this falls roughly near the region surrounding ±20 msec, a value corresponding to the threshold for temporal order processing (Hirsh, 1959).

The findings from Pisoni's (1977) study with nonspeech stimuli as shown in Figure 5 may be cited as additional support for the claim that a sensory or psychophysical process is probably responsible for the categorical-like

---
Insert Figure 5 About Here
---

discrimination performance found in adults and infants using synthetic speech stimuli differing in VOT. These results show that adult subjects are able to parse the TOT nonspeech continuum into three discrete perceptual categories corresponding to leading, lagging, and simultaneous onsets.

Figure 6 shows the ABX discrimination data from another nonspeech experiment by Pisoni (1977) with the same stimuli. Note that two distinct

---
Insert Figure 6 About Here
---

regions of high discriminability are present in the discrimination functions.

Thus, it is our contention that evidence of discrimination of VOT contrasts that straddle the -20 and +20 msec regions of the stimulus continuum probably results from general sensory constraints on the mammalian auditory system to resolve small differences in temporal order and not from phonetic categorization. However, two questions are immediately apparent from this analysis. First, why is there so little evidence of discrimination of VOT in the -20 msec region of voicing lead in the infant literature? And second, what role does the environment play in tuning the perceptual mechanism responsible for processing temporal order information?

The first question can be dealt with by a closer examination of the discrimination data shown in Figure 6. Note that even in this figure with nonspeech signals differing in their relative onset time, discrimination of TOT differences is greater in the positive region of the stimulus continuum than in the negative region. These findings are not unique to these particular nonspeech signals since the same relation can be observed in the original Abramson and Lisker discrimination data obtained with Thai subjects (1967). As shown in the top panel of Figure 7, the relative discriminability in the +20 msec region of voicing lag is greater than in the -20 msec region of voicing lead despite the fact that the slopes of the labeling

---
Insert Figure 7 About Here
---

functions for the Thai subjects in these regions are very nearly identical, as shown earlier in Figure 1. We propose, therefore, that the smaller incidence of discrimination of VOT differences in the minus region of voicing lead values is probably due to the generally poorer ability of the auditory system to resolve temporal differences in which a lower frequency component precedes a higher frequency component.

Lower discriminability of stimuli in the minus region of the VOT continuum cannot account entirely for infants' overall performance since all three positive instances of discrimination reported in the literature involved infants from linguistic environments that used contrasts between prevoiced and voiced stops distinctively. Thus, we would further argue that early linguistic experience does play some role in modifying the discriminability of speech stimuli depending on the relative predominance of certain VOT values in the productions of adults.

Differences in the relative discriminability of VOT contrasts along the stimulus continuum may be cited as additional support for the role of early environmental experience since there is evidence of two regions of high discriminability even in the discrimination functions obtained with English subjects as shown in the lower panel of Figure 7. However, the peak in the minus region is substantially reduced when compared with the Thai discrimination data shown in the top panel. A very similar result can also be observed in the discrimination data of Williams (1974) for Spanish and English subjects which is shown in Figure 8. Note that the Spanish subjects show a broad region of heightened discriminability extend-

---

Insert Figure 8 About Here

---

ing into the area encompassing the location of the English perceptual boundary.

The evidence on the development of voicing perception therefore appears to provide good support for the Attunement Theory described earlier. That is, a partially specified ability to process temporal order information is assumed to be present at birth. Perceptual sensitivity to temporal order

differences such as those present in synthetic VOT stimuli are, however, susceptible to the influence of early experience, thereby selectively modifying the strength and location of the regions of high sensitivity along a stimulus continuum such as VOT.

The course of perceptual development for other classes of speech sounds can also be accounted for in terms of the conceptual framework outlined earlier. To demonstrate the usefulness of our theoretical approach, we will briefly summarize some of the recent work that has appeared on the discrimination of fricatives, vowels, and liquids by young infants.

Fricatives. In a recent study, Eilers, Wilson, and Moore (1977) have reported that infants appear to have great difficulty discriminating between some of the acoustic attributes that differentiate the class of fricative sounds. These perceptual findings parallel the well-documented lag in the articulatory control of fricatives in speech production and suggest that infants probably must undergo a rather long period of perceptual learning to begin to isolate the appropriate acoustic cues for different fricatives. However, an equally plausible explanation of the developmental lag in the perception of fricatives is that the neural mechanisms underlying the perception of fricatives must unfold over a rather long postnatal period before a child is capable of discriminating the acoustic cues for different fricatives. We might suppose, then, that early experience either induces the abilities to discriminate fricatives, or that early experience plays no particular role in the development of the perceptual mechanisms required for discrimination of fricatives. If the Induction or Perceptual Learning Theory account of fricative development is correct, we might expect relatively poor discrimination of fricatives by infants whose native language does not employ specific fricative contrasts distinctively. Alternatively,

Maturational Theory would predict that at some postnatal age, infants from all language-learning environments would be able to discriminate differences between fricatives, but that early experience reduces the discriminability of the cues for some fricatives if they were absent from the language environment after the neural mechanisms necessary to process these acoustic cues had already reached maturity.

In contrast to these more traditional accounts of the development of fricative perception, an alternative account has been raised in a recent study conducted by our colleague Peter Jusczyk (personal communication). Jusczyk has demonstrated that 2-month-old infants discriminate a /fa/-/θa/ contrast categorically, but that the category boundary is located at the adult /ba/-/da/ boundary. In other words, for infants the region of heightened discriminability along both the /fa/-/θa/ and /ba/-/da/ continua is approximately coincident. Presumably, an experiential or maturational factor contributes to the developmental shift of the /fa/-/θa/ category boundary away from the /ba/-/da/ boundary. We would hypothesize on the basis of these data that infants' discrimination of fricatives is initially determined by the psychophysical attributes of the signal, but that a postnatal shift occurs in the manner in which fricatives are categorized. Thus, Attunement Theory may also provide a reasonable account of the development of fricative perception in infants.

Vowels. The work of Trehub (1976) provides another example of a cross-language comparison that helps to illuminate the possible roles of early experience in phonological development. She reported that infants from an English-speaking environment can discriminate a French vowel contrast, a contrast that was not discriminated reliably by English-speaking adults. Thus,

it would appear that this vowel contrast is discriminated at birth but that the original discriminative abilities decline postnatally as a result of the absence of particular kinds of early language experiences. This course of perceptual development may follow what we have characterized as the Universal Theory or alternatively may be best described by Attunement Theory. In the case of vowel perception in adults, there are specific regions of the vowel space that are generally associated with individual vowel classes, although these differ from language to language. The arrangement of the vowel space may, however, initially conform to the acoustic attributes of vowels that are processed most efficiently by the newborn's auditory system and then only during postnatal development, with environmental experience as input, will the vowel space be rearranged so as to conform more closely to the phonological categories present in the language-learning environment (see Liljencrants & Lindblom, 1972).

Liquids. Finally, the data on the discrimination of the liquids [r] and [l] as reported by Miyawaki et al. (1975) for adults and by Eimas (1975) for infants suggest that the ability to discriminate differences in the F3 transitions between the liquids is present at birth. However, the absence of an [r]-[l] contrast in the early postnatal environment of Japanese infants may have prevented the adults in the Miyawaki et al. study from discriminating what is now a phonologically-irrelevant speech contrast. The Universal Theory would then appear to be the best candidate to account for this selective loss of a perceptual ability that was initially present at birth.

The perceptual findings briefly reviewed here indicate that early experience may significantly modify the sensory-based perceptual categories

presumed to be present at birth. There are several ways in which early experience could selectively modify the perceptual mechanisms and therefore influence the discriminability of phonetic contrasts found in natural language environments. Figure 9 outlines several forms that this selective modification might take with reference to the shape and level of schematized discrimination functions.

Insert Figure 9 About Here

First, stimuli in the region of a boundary between two perceptual categories may become either more discriminable or less discriminable, processes we have called enhancement and attenuation. The process of enhancement may account for the heightened discriminability of stops in the pre-voiced region of the VOT continuum by Thai speakers. In contrast, the process of attenuation may account for the poor discriminability of VOT differences in the pre-voiced region by English speakers, and the apparent decrease in discriminability of the [r]-[l] contrast by Japanese speakers.

Perception of stimuli in the region of a perceptual boundary may also become more finely tuned or more poorly tuned, processes we have called sharpening and broadening. On the one hand, sharpening may account for the discrete and very well-defined cross-over points in adult labeling functions obtained for synthetic stops. On the other hand, broadening may account for the perception of vowels being somewhat more discriminable overall than consonants as well as the wider region of heightened discriminability of VOT found in Spanish subjects.

Finally, the perceptual boundary may undergo a shift, a process we have called realignment that may account for the shifts in the voiced-

voiceless boundary observed between English and Spanish stops as reported in studies by Lisker and Abramson (1967), Williams (1974), and Eilers, Wilson, and Gavin (1977).

From this brief summary of the perception of voicing contrasts in stop consonants, fricatives, vowels, and liquids it should be apparent that the major roles of early experience that we outlined earlier cannot be uniformly invoked to account for the development of the abilities needed to discriminate all speech contrasts found in spoken language. Clearly there are numerous variables and factors that will determine which particular ontogenetic function in Figure 4 best characterizes the developmental course of a particular phonetic contrast. For example, the auditory system of humans may well be specialized for processing certain very specific types of acoustic attributes at an early age. If some phonetic contrasts in language happen to have these distinctive acoustic properties in common, the infant should then be able to discriminate these speech signals with practically no experience in the language learning environment short of sensory deprivation. On the other hand, if a certain amount of neural maturation or specific early experience is required for discrimination, then we might anticipate a delay or developmental lag in observing discrimination of these contrasts, assuming, of course, that all other things remained constant. This hybrid or parallel view of the role of early experience in the development of speech perception that we are proposing here is not entirely without precedent or empirical support as shown by recent work on visual system development. For example, most visual cortical neurons are characterized by their simultaneous responsiveness to several aspects of stimulus structure: direction of movement, orientation, and retinal disparity (among others). Yet,

the mechanism underlying the development of each of these types of stimulus specificity is quite different. The property of directional selectivity is present at birth and undergoes little improvement or loss, unless the animal is deprived of stimulus movement (see Olson & Pettigrew, 1974 on stroboscopic rearing conditions). The property of orientational selectivity is also present at birth, but the sharpness of each neuron's orientational specificity is dependent upon the quality of early experience (Sherk & Stryker, 1975). And the property of disparity selectivity appears to be nearly absent at birth and the neurons acquire (within broad limits) the range of disparity values provided during early life (Pettigrew, 1974). Thus, the three general theory classes in the speech domain -- Universal, Attunement, and Perceptual Learning -- could be thought of as being analogous to three general mechanisms by which early experience influences visual system development. Moreover, the foregoing analogy emphasizes the fact that parallel developmental mechanisms can operate upon different aspects of the same sensory input and underly the perceptual abilities observed in the adult. It should be obvious now that only a very detailed description of the development of these discriminative abilities will enable us to distinguish between the various types of complex interactions caused by genetic and experiential factors and their contribution to the development of the normal speech processing mechanisms.

METHODOLOGICAL CONSIDERATIONS

Many of the questions raised in the previous sections about the role of early experience in perceptual development cannot be answered in a completely satisfying manner without new and improved experimental procedures for measuring discrimination and other aspects of perceptual analysis. For example, in order to answer questions about the discriminative abilities

of infants to resolve small differences between speech signals that exist at a purely sensory level, many more data points need to be obtained from each subject so that the shape of the discrimination functions can be examined in greater detail. To study questions surrounding perceptual constancy and categorization of speech signals some measure analogous to the adult labeling or identification function is needed.

For the past year we have been working on several techniques for use with infants which would provide us with new and more efficient ways to measure both discrimination and identification of speech and speech-like acoustic signals. The procedures we developed to measure discrimination and perceptual categorization of speech sounds by infants involved several modifications of the operant head-turning paradigm described in earlier papers by Eilers and Kuhl. An assistant attracts the gaze of the infant while various signals are presented to the infant who is seated on the mother's lap in a sound-attenuated booth. The mother and assistant both wear tight-fitting headphones during the entire experimental session and listen to a continuous recording of popular music played at a level sufficient to mask the background and target stimuli. A PDP-11 computer is used to present all experimental signals, record responses and deliver visual reinforcement to the infant. An experimenter,who is unaware of the specific stimulus conditions on each trial as well as the reinforcement contingencies,is located outside the booth and records headturns directly into the computer which has been programmed to present the stimuli and provide the visual reinforcement. In contrast to the head-turning procedure described earlier by Eilers and Kuhl to measure discrimination,in our procedures, the assistant located in the room with the infant does <u>not</u> have any control over the

stimulus or reinforcement conditions of the experiment and therefore is not in a position to provide subtle cues to the infant since she is unaware of the exact stimuli due to the masking stimulus in her earphones. Moreover, because the experiment is completely automated there is no chance of experimenter bias influencing any aspect of the procedure which could affect the infants behavior in the experiment.

In one of our studies we have collected discrimination data from infants using a modified staircase or up-down procedure (Levitt, 1971). Infants are initially shaped to detect a difference between a repeating background signal and a target signal by turning toward the direction of the change. A correct headturn to the left of midline is reinforced by the presentation of an animated toy monkey which is located next to the loudspeaker. This aspect of the headturning procedure is similar to the method described in earlier chapters by Eilers and Kuhl. However, after this initial shaping phase and a training phase is completed to a criterion level of performance, a series of trials is presented in which the characteristics of the target signal are modified systematically depending upon the infants' performance on the previous trials. An example of the results of this interactive staircase procedure when used to measure VOT discrimination is shown in Figure 10.

Insert Figure 10 About Here

This particular infant, KS, was trained on an initial stimulus pair consisting of a +70 msec VOT background signal and a -70 msec VOT target stimulus. After reaching a training criterion of 4/5 correct responses on experimental or change trials and 4/5 correct responses on control trials,

the infant was shifted immediately into the staircase discrimination procedure. As currently implemented, the staircase procedure operates according to a very simple algorithm. If the infant discriminates correctly on two consecutive experimental trials (i.e. a hit), the level of the target stimulus is changed on the next trial by some value of VOT so as to make it less discriminable than the previous level of the target. However, if a miss occurs on any trial, the level of the target is then adjusted on the next trial by some value of VOT to be more discriminable. The results of this up-down or staircase procedure typically produces a pattern of oscillation along the stimulus continuum at a value of VOT that can be correctly discriminated 70% of the time from the background stimulus or standard. The results of testing subject KS with the full VOT continuum from -70 to +70 msec is shown in the left panel where a ΔVOT value of 20 msec was computed. This same subject was re-tested a day later on only the 0 through +70 msec range of this continuum and a ΔVOT value of 25 msec was computed with a smaller step size. Both values agree remarkably well as estimates of the minimum value of VOT that can be discriminated reliably from the +70 msec background stimulus despite the change in the range of stimuli used.

The staircase discrimination procedure has a number of advantages over more traditional psychophysical methods used to measure discrimination, especially when used with young infants. First, it is very efficient, requiring substantially fewer trials than more conventional methods. Second, it offers greater flexibility, thus permitting the infant to continue responding in order to receive reinforcement and serves to maintain interest and attention to the experimental situation for longer periods of time. Finally, it provides a rapid and quantitative measure of discriminability

of the dimension under study. Such a measure is also less subject to biases resulting from peculiar response strategies than more traditional methods used with infants which often promote a high false alarm rate or alternatively a reduction in responding entirely.

Of course, establishing that infants can discriminate differences between various kinds of speech signals is only one aspect of the speech perception process since there are numerous acoustic attributes of speech signals that can be used for a discriminative response. Moreover, there is some difficulty in interpreting the results from discrimination experiments alone since it is often unclear from these studies which specific acoustic attribute or property was used by the infant for discrimination. In complex acoustic signals such as speech there are typically a very large number of redundant and irrelevant acoustic cues which may underlie the acoustic realization of a particular phonetic distinction and which may be sufficient for an infant or adult to discriminate this distinction from others.

To examine more directly some of the general questions surrounding perceptual constancy in speech perception and how constancy is maintained despite the wide diversity of physical variability in the sensory input to infants, it seemed necessary to develop a procedure which would provide information on how infants categorize or identify speech signals -- signals that are very likely to be acoustically different although in some cases phonetically or phonologically equivalent. Indentification or labeling functions secured from young infants in this manner could then be directly compared to data obtained with human adults and other species in order to determine, for example, how early experience with these stimuli

in the environment may have tuned or modified the perceptual system to respond selectively to certain criterial attributes and to ignore other non-criterial properties.

We have also been working on several techniques for use with infants which would provide an analog to the labeling or identification task used in adult speech perception experiments (see Aslin, Perey, Hennessy & Pisoni, 1977).

We call our procedure a "two-alternative go/no-go" categorization task because "Go" trials consist of presenting a stimulus, S+, and reinforcing an appropriate headturn toward the left whereas "No-Go" trials consist of presenting stimulus S- and not reinforcing the headturn response which was reinforced during S+ trials. In essence, we are training the infant to respond only to the criterial properties of S+ which serve as the discriminative stimulus for a headturn response and not to respond to S-, the stimulus for which a head-turning response is not reinforced.

Table 1 summarizes the performance of four infants tested on the go/no-go categorization task. In each case, the background stimulus, S_B was the syn-

---
Insert Table 1 About Here
---

thetic vowel /u/ which was 350 msec in duration and repeated once every second. Stimuli S+ and S- were the synthetic vowels /a/ and /i/. As shown in Table 1, all infants rapidly learned to turn toward a change from the background stimulus to the target stimulus S+. Summed over all four infants, the level of appropriate correct headturns to S+ was 88%. Correct performance is not as high initially when the infants are required to withhold headturning to a change from the background stimulus to S-. However,

the fact that they do turn their head on the first few trials indicates that they are capable of discriminating S- from the background signal -- a fact that is clearly necessary for this procedure to be a meaningful analog of an identification paradigm. Despite the infants' difficulties in initially withholding a headturn response on S- trials, the data shown in Table 1 indicate that infants can learn to respond correctly and consistently and that they can do this quite rapidly in the span of a small number of trials. Although discrimination paradigms such as those described earlier can provide useful and important information about the abilities of young infants to discriminate or detect small and very subtle differences between various kinds of speech signals, additional information is needed about infants early abilities to classify or categorize acoustically different sounds that adults perceive as phonologically the same.

We are currently using this categorization procedure in our laboratory to collect labeling data from young infants that should be directly comparable to the data collected with adults and chinchillas. Because all of our experiments are run "on-line" in real-time under computer control, we are able to present a wide range of different experimental signals for generalization testing, thus permitting us to ask a number of different questions concerning perceptual constancy, feature analysis, and the role of early experience in perceptual development. These categorization results should be particularly relevant to questions of the kind we raised earlier surrounding the existence of species-specific perceptual mechanisms for processing biologically significant acoustic signals such as speech.

CONCLUDING REMARKS

Our goal in this presentation has been to organize some of the scattered

literature in the area of infant speech perception under a uniform conceptual framework that recognizes the important roles early experience plays in perceptual development. While there is currently a great deal of empirical data available on infant speech perception, the same cannot be said for theories of development. Previous accounts of speech perception have been very naive, in our view, in their treatment of the role of early experience and the possible interactions between genetic and experiential factors in speech perception. We hope that we have been successful in at least pointing out how complex these interactions can be in the development of speech perception abilities and in suggesting some possible new ways to examine these interactions. Finally, it is our hope that with new experimental techniques to study infant perception and a more sophisticated conceptual framework, many of the general issues surrounding the developmental course of speech perception can be pursued in a more systematic way than has been true in the past.

References

Abramson, A.S. & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. In Proceedings of the 6th International Congress of Phonetic Sciences. Prague: Academia, Pp. 569-573.

Aslin, R.N., Perey, A.J., Hennessy, B. and Pisoni, D.B. Perceptual analysis of speech sounds by prelinguistic infants: A first report. Paper presented at the Acoustical Society of America meetings, Miami Beach, December, 1977.

Banks, M.S. and Aslin, R.N. Binocular development in human albinos and congenital esotropes. Paper presented at the Midwest Regional Association for Research in Vision and Ophthalmology, Ann Arbor, October, 1975.

Banks, M.S., Aslin, R.N. and Letson, R.D. (1975). Sensitive period for the development of human binocular vision. Science, 190, 675-677.

Blakemore, C. (1976). The conditions required for the maintenance of binocularity in kitten's visual cortex. Journal of Physiology, (London), 261, 423-444.

Carney, A.E., Widin, G.P. & Viemeister, N.F. (1977). Noncategorical perception of stop consonants differing in VOT. Journal of the Acoustical Society of America, 62, 961-970.

Creel, D., Witkop, C.J. and King, R.A. (1974). Asymmetric visually evoked potentials in human albinos: evidence for visual system anomalies. Investigative Ophthalmology, 13, 430-440.

Cutting, J.E., & Rosner, B.S. (1974). Categories and boundaries in speech and music. Perception and Psychophysics, 16, 564-570.

Eilers, R.E., Wilson, W.R., and Gavin, W.J. Perception of VOT by Spanish-learning infants. Paper presented at the 93rd Meeting of the Acoustical Society of America, December, 1977.

Eilers, R.E., Wilson, W.R., and Moore, J.M. (1977). Developmental changes in speech discrimination in infants. Journal of Speech and Hearing Research, 20, 766-780.

Eimas, P.D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the r-l distinction by young infants. Perception and Psychophysics, 18, 341-347.

Eimas, P.D., Siqueland, E.R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. Science, 171, 303-306.

Fry, D.B. (1966). The development of the phonological system in the normal and deaf child. In F. Smith & G.A. Miller (Eds.) The Genesis of Language. Cambridge: M.I.T. Press, Pp. 187-216.

Gesell, A.L., and Ames, L.B. (1940). The ontogenetic organization of prone behavior in human infancy. Journal of Genetic Psychology, 56, 247-263.

Gottlieb, G. (1975). Development of species identification in ducklings: I. Nature of perceptual deficit caused by embryonic auditory deprivation. Journal of Comparative and Physiological Psychology, 89, 387-399.

Gottlieb, G. (1976a). Conceptions of prenatal development: Behavioral embryology. Psychological Review, 83, 215-234.

Gottlieb, G. (1976b). The roles of experience in the development of behavior and the nervous system. In G. Gottlieb (Ed.), Neural and Behavioral Specificity. New York: Academic Press.

Grobstein, P., and Chow, K. (1976). Receptive field organization in the mammalian visual cortex: The role of individual experience in development. In G. Gottlieb (Ed.), Neural and Behavioral Specificity. New York: Academic Press.

Hess, E.H. (1972). "Imprinting" in a natural laboratory. Scientific American, August, 227, 24-31.

Hirsch, I.J. (1959). Auditory perception of temporal order. Journal of the Acoustical Society of America, 31 (6), 759-767.

Hubel, D.H., and Wiesel, T.N. (1963). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. Journal of Neurophysiology, 26, 994-1002.

Hubel, D.H., and Wiesel, T.N. (1965). Binocular interaction in striate cortex of kittens reared with artificial squint. Journal of Neurophysiology, 28, 1041-1059.

Hubel, D.H., and Wiesel, T.N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. Journal of Physiology (London), 206, 419-436.

Jusczyk, P.W. (1978). Personal communication.

Jusczyk, P.W., Rosner, B.S., Cutting, J.E., Foard, C.F., and Smith, L.B. (1977). Categorical perception of non-speech sounds by two-month old infants. Perception and Psychophysics, 21, 50-54.

Kuhl, P.K., and Miller, J.D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. Science, 190, 69-72.

Kuhl, P.K., and Miller, J.D. (1978). Speech perception by the chinchilla: Indentification functions for synthetic VOT stimuli. Journal of the Acoustical Society of America, 63, 905-917.

Lasky, R.E., Syrdal-Lasky, A., and Klein, R.E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. Journal of Experimental Child Psychology, 20, 215-225.

Levitt, H. (1970). Transformed up-down methods in psychoacoustics. Journal of the Acoustical Society of America, 49, 467-477.

Liljencrants, J., and Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. Language, 48, 839-862.

Lisker, L., and Abramson, A.S. (1967). The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences, Prague.

Miller, J.D., Wier, C.C., Pastore, R., Kelly, W.J., and Dooling, R.J. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. Journal of the Acoustical Society of America, 60 (2), 410-417.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J. and Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. Perception and Psychophysics, 18, 331-340.

Olson, C.R., and Pettigrew, J.D. (1974). Single units in visual cortex of kittens reared in stroboscopic illumination. Brain Research, 70, 189-204.

Pettigrew, J.D. (1974). The effect of visual experience on the development of stimulus specificity by kitten cortical neurons. Journal of Physiology (London), 237, 49-74.

Pisoni, D.B. (1977). Identification and discrimination of the relative onset of two component tones: Implications for the perception of voicing in stops. Journal of the Acoustical Society of America, 61, 1352-1361.

Sherk, H., and Stryker, M.P. (1977). Quantitative study of cortical orientation selectivity in visually inexperienced kitten. Journal of Neurophysiology, 39 (1), 63-70.

Shiffrin, R.M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychological Review, 84, 127-190.

Stevens, K.N., and Klatt, D.H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. Journal of the Acoustical Society of America, 55, 653-659.

Strange, W., and Jenkins, J.J. (1978). The role of linguistic experience in the perception of speech. In H.L. Pick, Jr., & R.D. Walk (Eds.), Perception and experience. New York: Plenum Publishing Corp.

Trehub, S.E. (1976). The discrimination of foreign speech contrasts by infants and adults. Child Development, 47, 466-472.

Williams, C.L. Speech perception and production as a function of exposure to a second language. Doctoral Dissertation, Harvard University, 1974.

Table 1

Individual subject data using the Two-Alternative Go/No-Go Procedure (from Aslin, Perey, Hennessy & Pisoni, 1977).

Subject	Trials	S+ = GO	S- = NO GO	Percent Correct
RF (6.5 months)	1-4	——— Shaping ———		
S_B = /u/	5-9	3/3	0/2	60%
S+ = /i/	10-19	3/3	3/7	60%
S- = /a/	20-29	5/5	3/5	80%
JH (9 months)	1-5	——— Shaping ———		
S_B = /u/	6-15	6/6	1/4	70%
S+ = /i/	16-25	5/5	2/5	70%
S- = /a/	26-35	4/5	2/5	60%
	36-45	4/5	2/5	60%
	46-52	4/4	2/3	85%
MB (6 months)	1-18	——— Shaping ———		
S_B = /u/	19-28	6/6	1/4	70%
S+ = /a/	29-38	4/5	3/5	70%
S- = /i/	39-48	4/5	2/5	60%
CM (8 months)	1-10	——— Shaping ———		
S_B = /u/	11-20	5/5	3/5	80%
S+ = /i/	21-30	4/7	3/3	70%
S- = /a/	31-40	5/7	1/3	60%
	41-50	4/5	3/5	70%
	51-60	4/4	1/6	50%

Figure Captions

Figure 1. Adult labeling functions for synthetic labial, apical and velar stop consonants varying in VOT obtained from native speakers of English, Thai and Spanish (Adapted from Lisker and Abramson, 1967).

Figure 2. Three-alternative and two-alternative labeling functions obtained from four (naive) native speakers of English. The synthetic stimuli were bilabial stop consonants varying in VOT (Data collected by Beth Hennessy).

Figure 3. Sensitive periods for visual system binocularity in cats and humans as estimated by data from Hubel & Wiesel (1965, 1970) and Banks, Aslin & Letson (1975), respectively.

Figure 4. Illustration of the major roles that early postnatal experience can play in modifying the relative discriminability of speech sounds. Three general classes of theories are shown here to account for the development of speech sound discrimination: Universal theory, Attunement theory and Perceptual Learning theory.

Figure 5. Adult labeling functions for tone-onset-time (TOT) stimuli showing the presence of three labeling categories in adult subjects (From Pisoni, 1977).

Figure 6. Adult ABX discrimination data for tone-onset-time (TOT) stimuli showing two regions of heightened discriminability (From Pisoni, 1977).

Figure 7. Oddity discrimination data obtained from adult speakers of Thai and English for synthetic bilabial stop consonants varying in

VOT (Redrawn from Abramson and Lisker, 1967).

Figure 8. A-X discrimination data from adult speakers of Spanish and English for synthetic bilabial stop consonants varying in VOT (Redrawn from Williams, 1974).

Figure 9. Five processes by which early experience in a particular language environment could modify the relative discriminability of speech sounds lying along a particular synthetic continuum.

Figure 10. Discrimination data from infant KS who was tested with synthetic VOT stimuli using a modified staircase procedure. The lefthand panel shows trial-by-trial performance when the full range of stimuli from -70 msec through +70 msec was used whereas the righthand panel shows performance with a reduced range from 0 msec through +70 msec VOT. A + indicates a correct response (hit) whereas a - indicates a incorrect response (miss). The ΔVOT needed to discriminate the +70 msec background stimulus from the target is quite consistent across both sessions lying roughly at a value of 25 msec VOT as shown by the horizontal line with filled circles in both panels.

LISKER & ABRAMSON (1967)
CROSS-LANGUAGE LABELING DATA

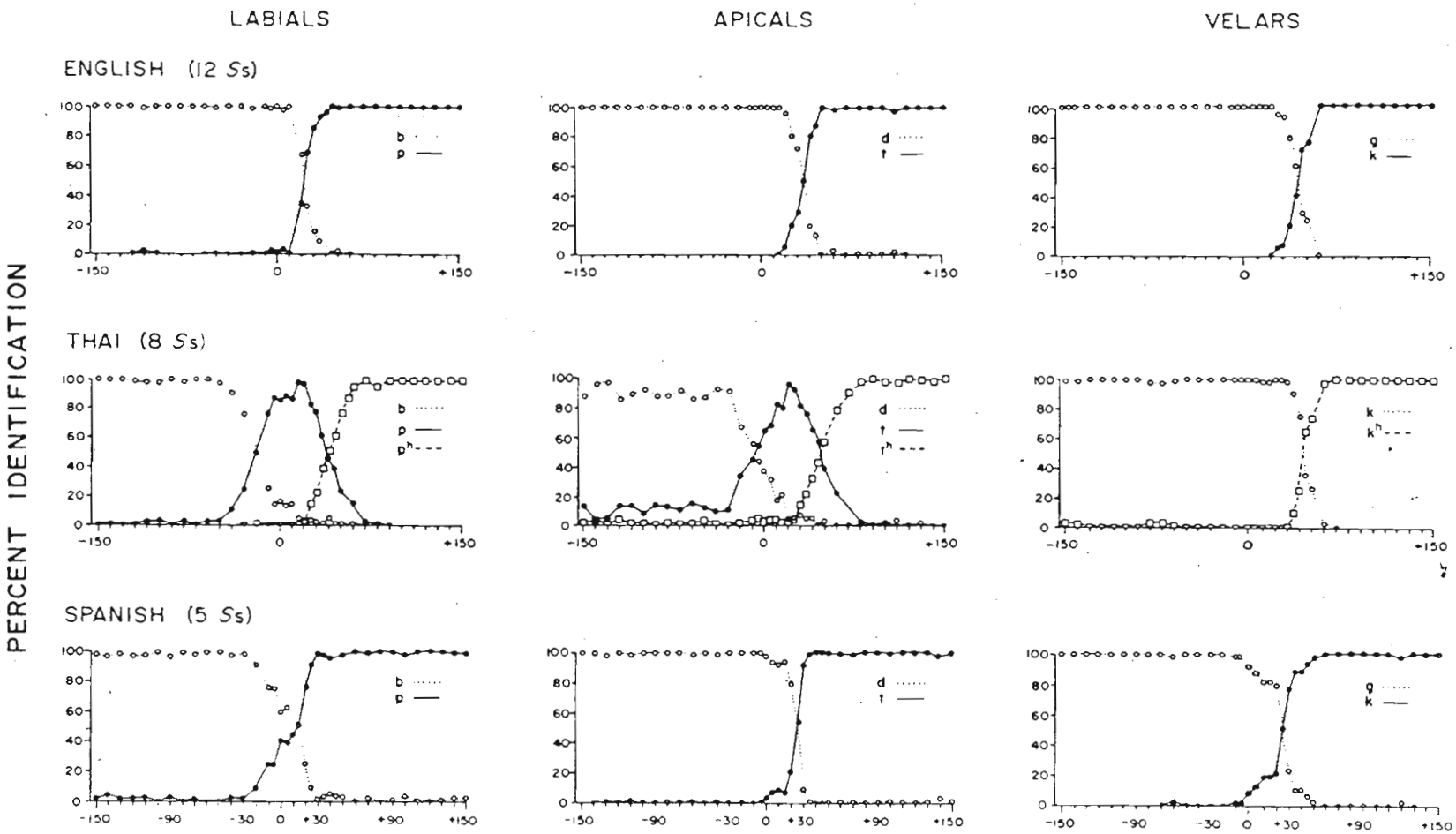


FIGURE 1.

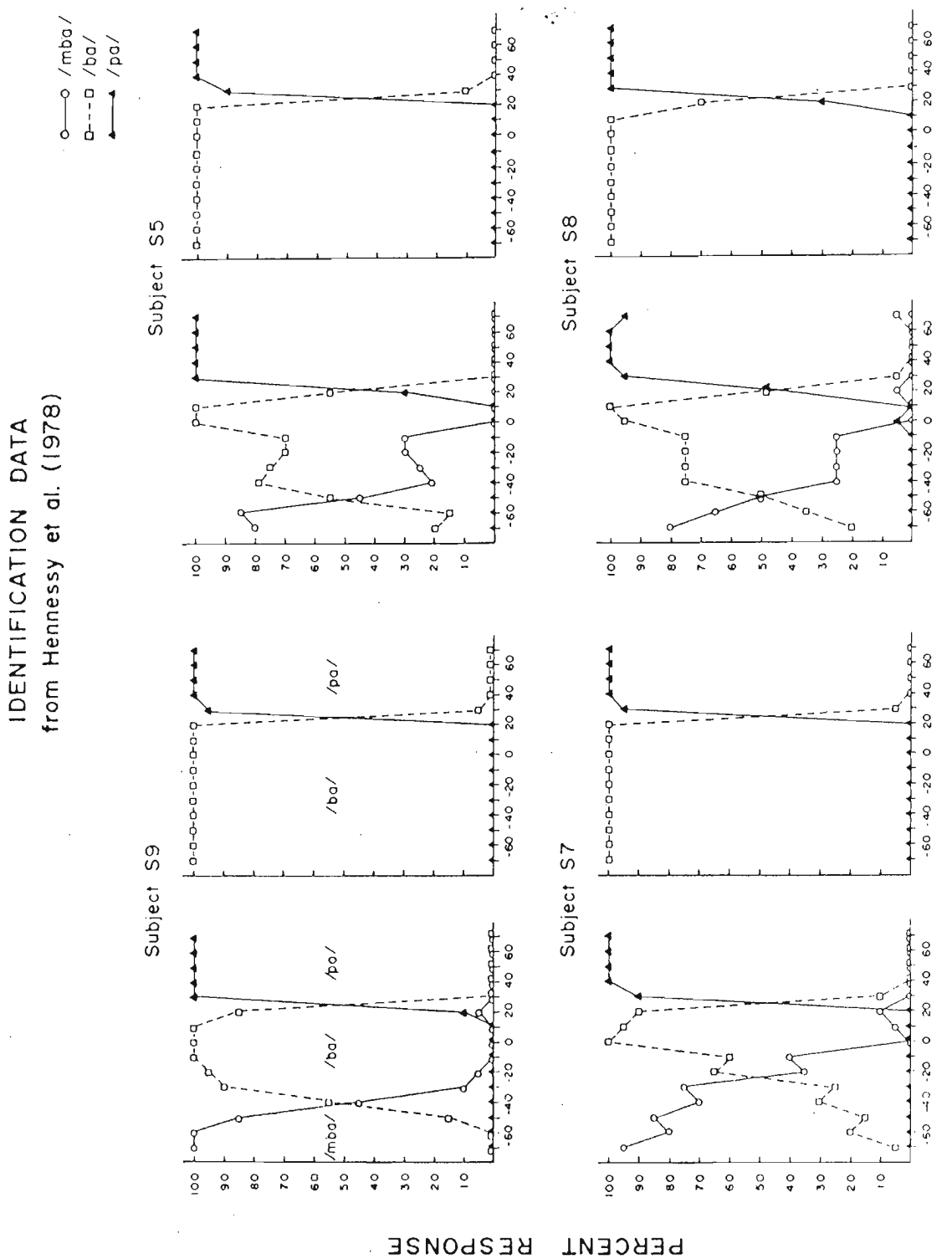
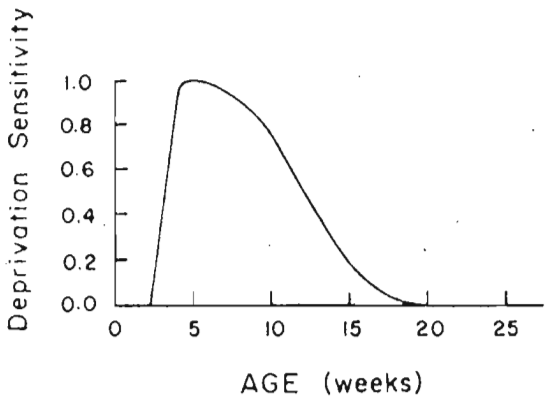

IDENTIFICATION DATA
from Hennessy et al. (1978)
/mba/
/ba/
/pa/
Subject S9
Subject S5
Subject S7
Subject S8
/mba/
/ba/
/pa/
PERCENT RESPONSE
VOICE ONSET TIME (msec)

FIGURE 2.

SENSITIVE PERIOD FOR BINOCULARITY IN KITTENS
(Hubel & Wiesel, 1965, 1970)



SENSITIVE PERIOD FOR BINOCULARITY IN HUMANS
(Banks, Aslin & Letson, 1975)

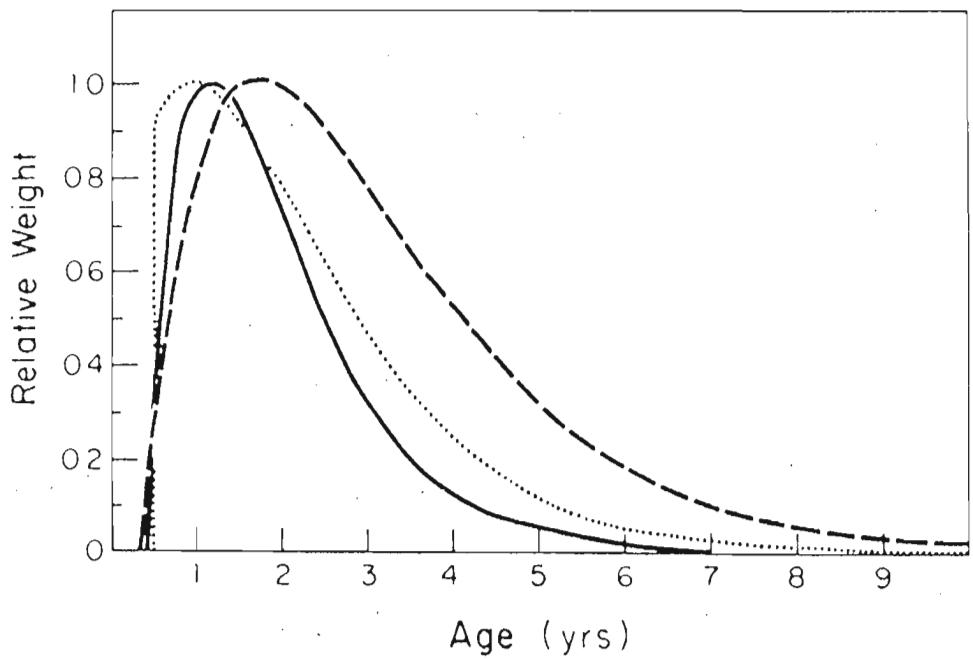


FIGURE 3.

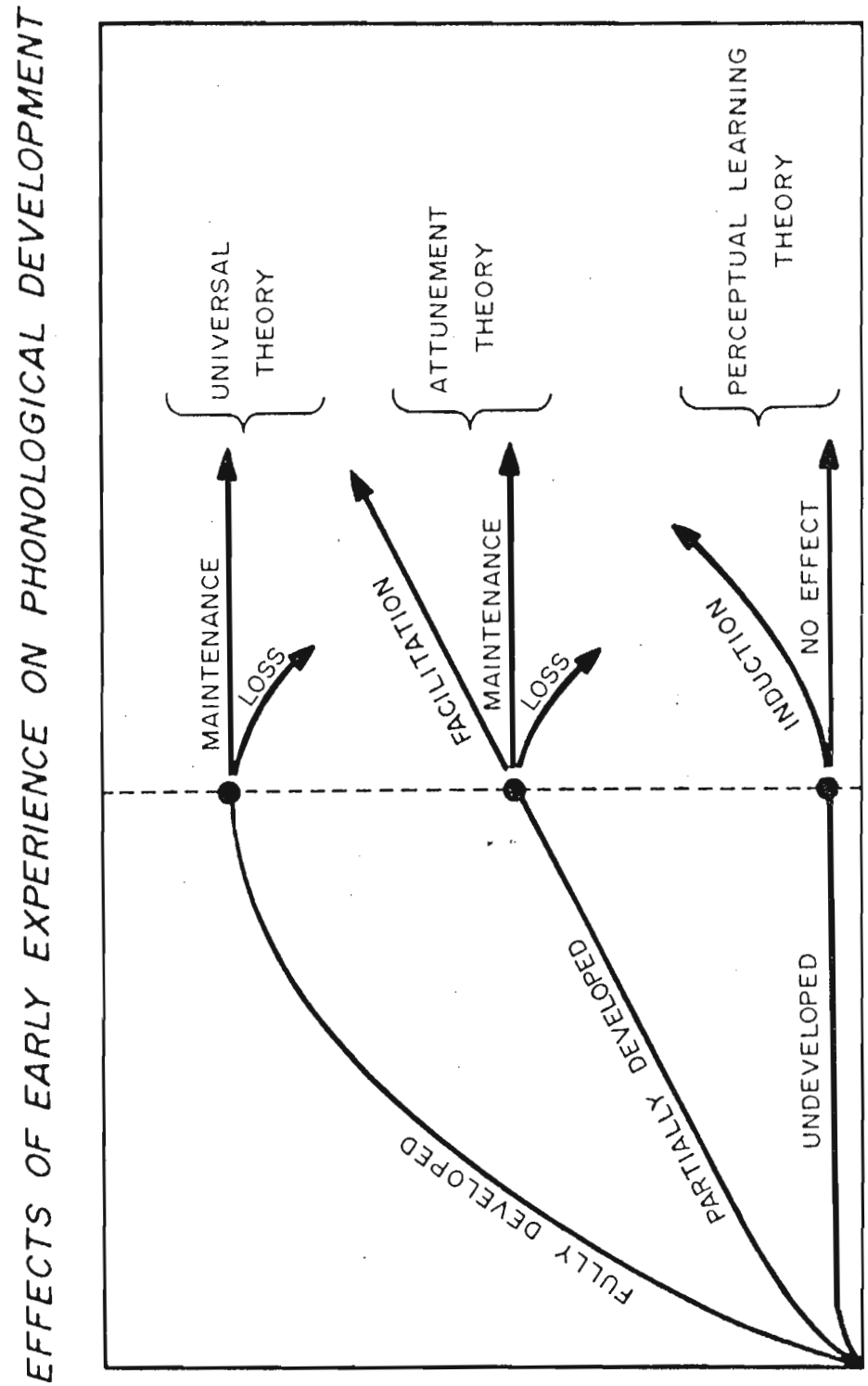

EFFECTS OF EARLY EXPERIENCE ON PHONOLOGICAL DEVELOPMENT
RELATIVE DISCRIMINABILITY
UNIVERSAL THEORY
ATTUNEMENT THEORY
PERCEPTUAL LEARNING THEORY
MAINTENANCE
LOSS
FACILITATION
MAINTENANCE
LOSS
INDUCTION
NO EFFECT
FULLY DEVELOPED
PARTIALLY DEVELOPED
UNDEVELOPED
PRENATAL
POSTNATAL
AGE

FIGURE 4.

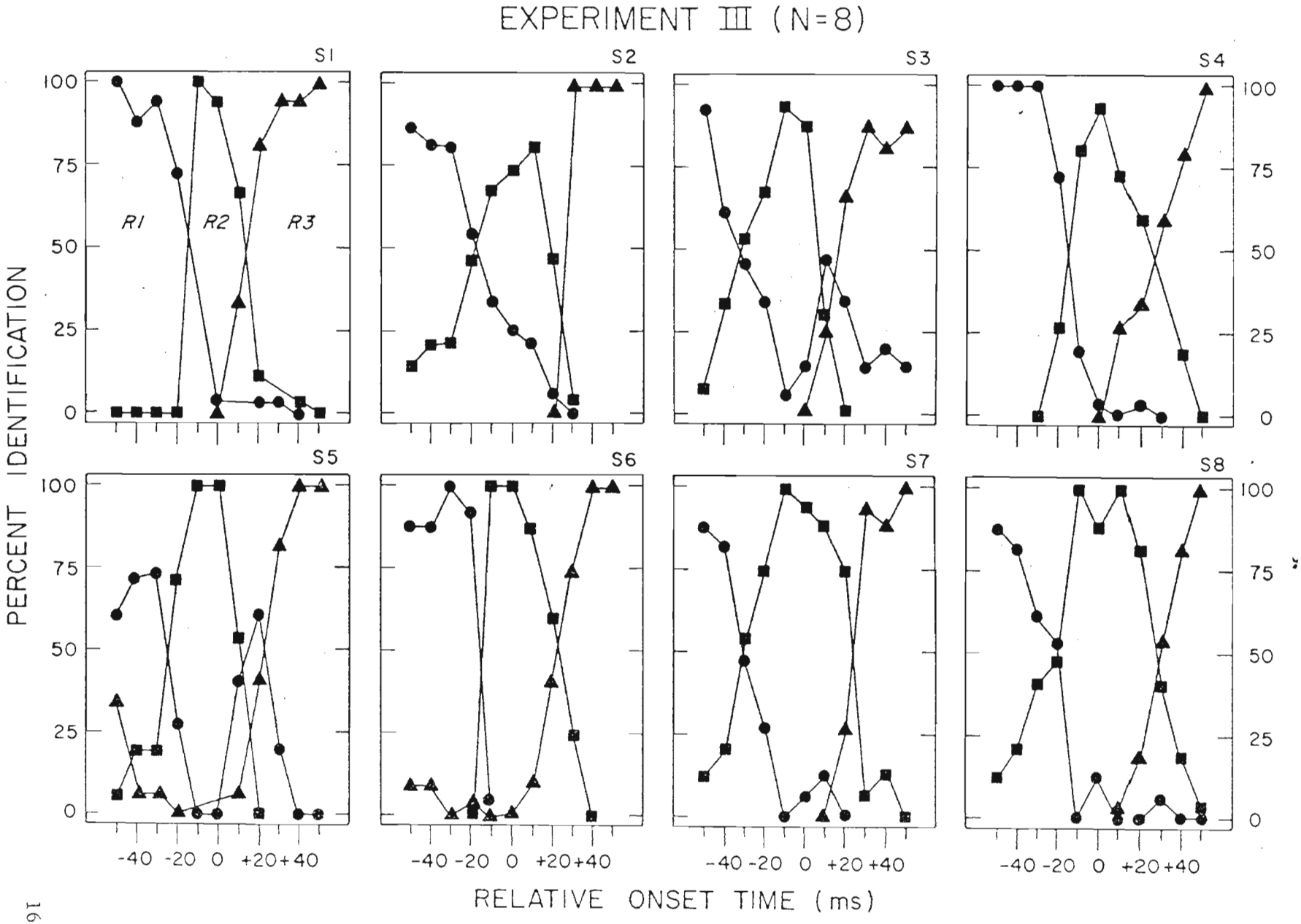

EXPERIMENT III (N=8)
S1
S2
S3
S4
S5
S6
S7
S8
R1
R2
R3
100
75
50
25
0
PERCENT IDENTIFICATION
-40 -20 0 +20 +40
RELATIVE ONSET TIME (ms)

FIGURE 5.

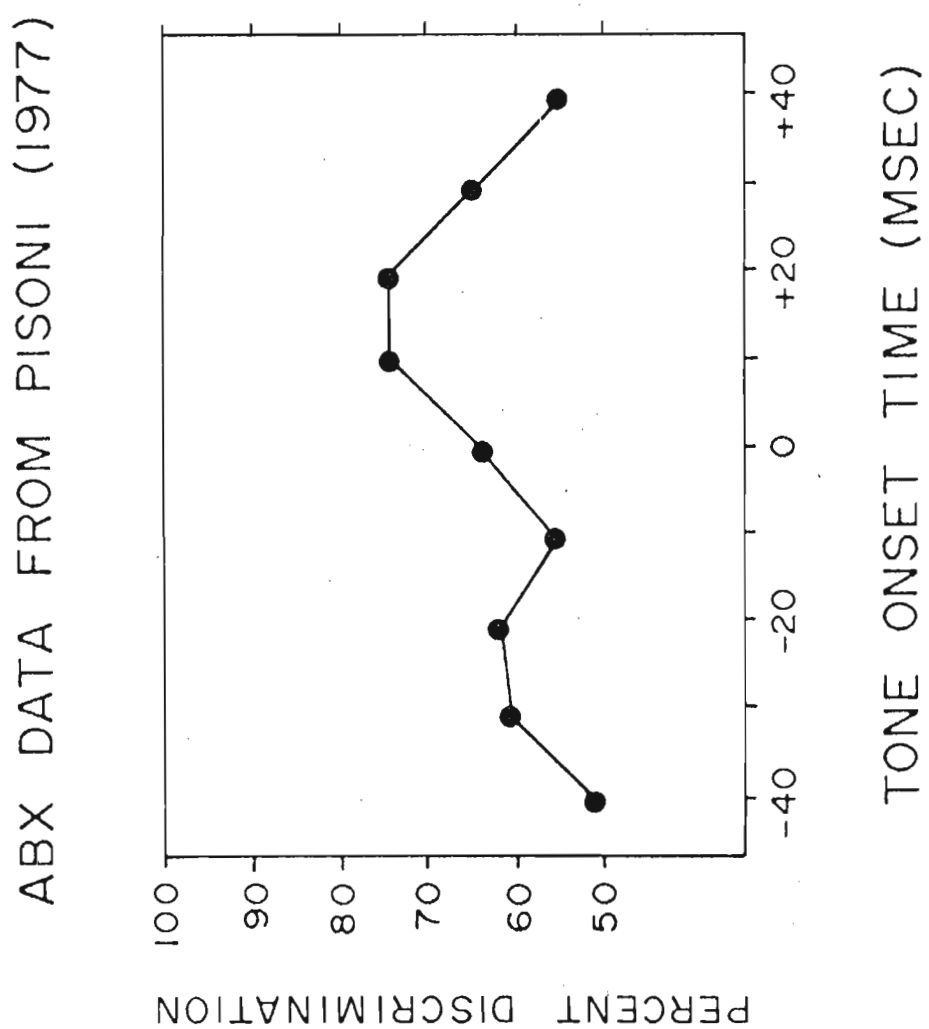

ABX DATA FROM PISONI (1977)
PERCENT DISCRIMINATION
100
90
80
70
60
50
-40
-20
0
+20
+40
TONE ONSET TIME (MSEC)

FIGURE 6.

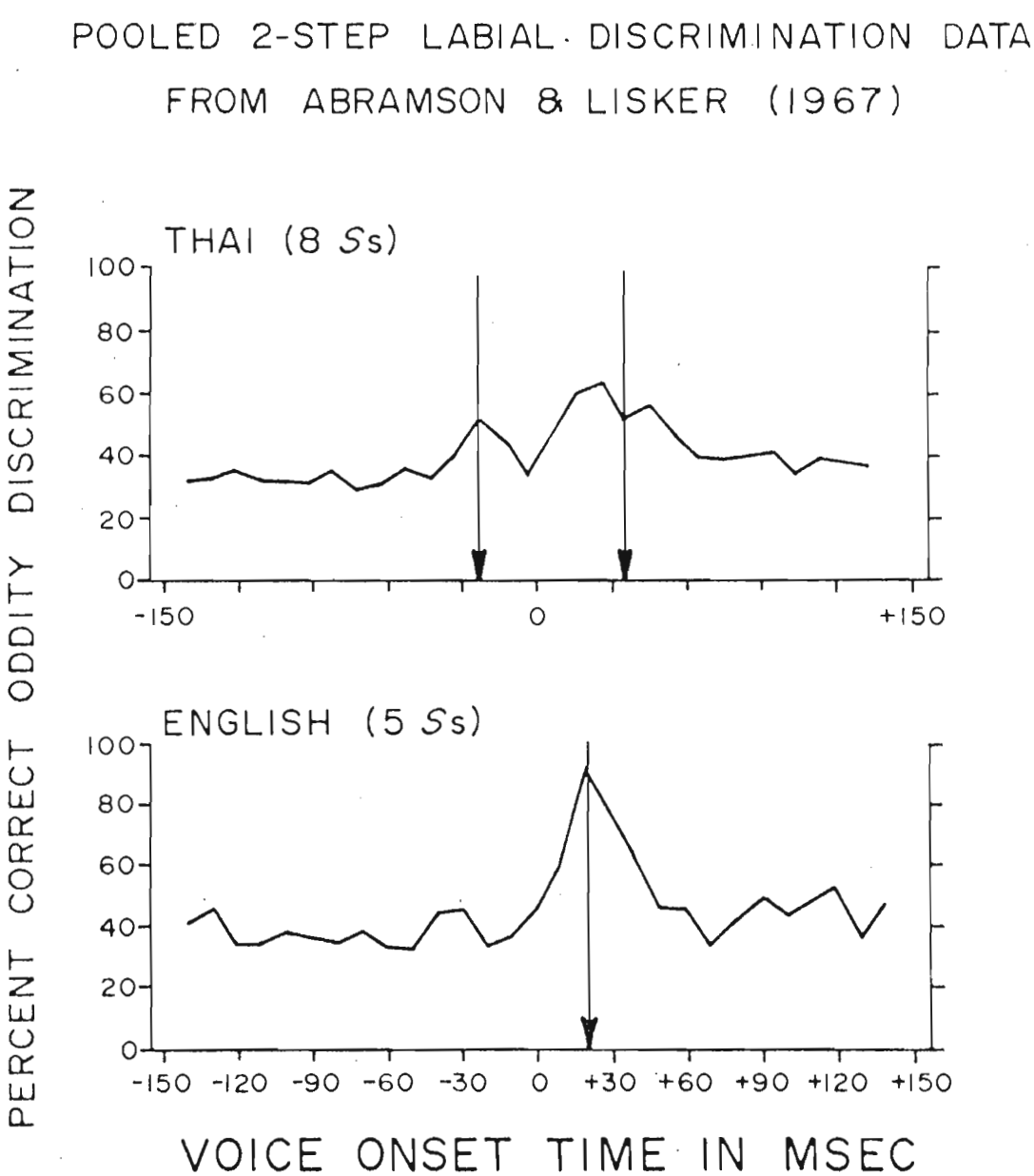

POOLED 2-STEP LABIAL DISCRIMINATION DATA
FROM ABRAMSON & LISKER (1967)
PERCENT CORRECT ODDITY DISCRIMINATION
THAI (8 Ss)
100
80
60
40
20
0
-150
0
+150
ENGLISH (5 Ss)
100
80
60
40
20
0
-150 -120 -90 -60 -30 0 +30 +60 +90 +120 +150
VOICE ONSET TIME IN MSEC

FIGURE 7.

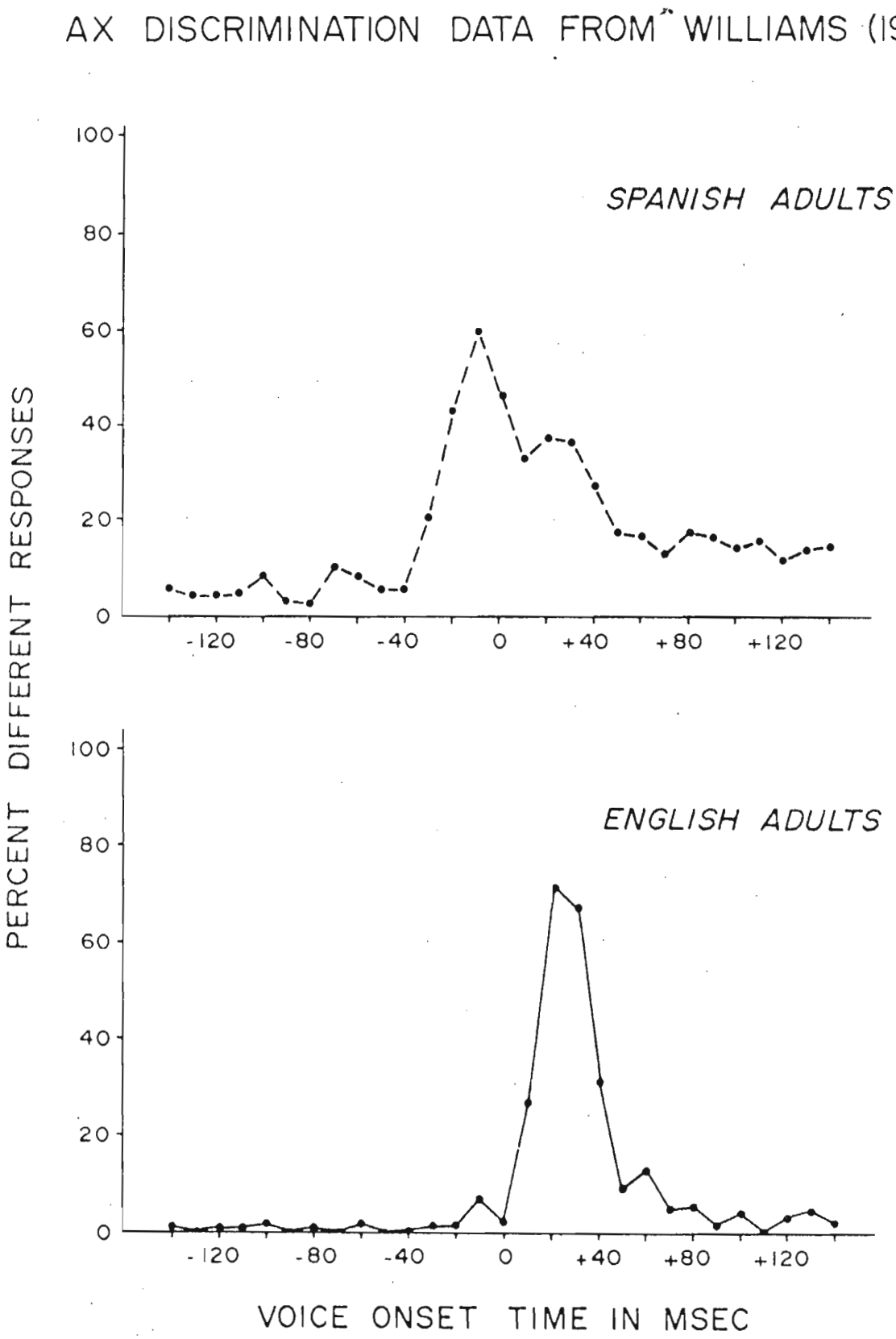

AX DISCRIMINATION DATA FROM WILLIAMS (1974)
SPANISH ADULTS
ENGLISH ADULTS
PERCENT DIFFERENT RESPONSES
100
80
60
40
20
0
-120
-80
-40
0
+40
+80
+120
VOICE ONSET TIME IN MSEC

FIGURE 8.

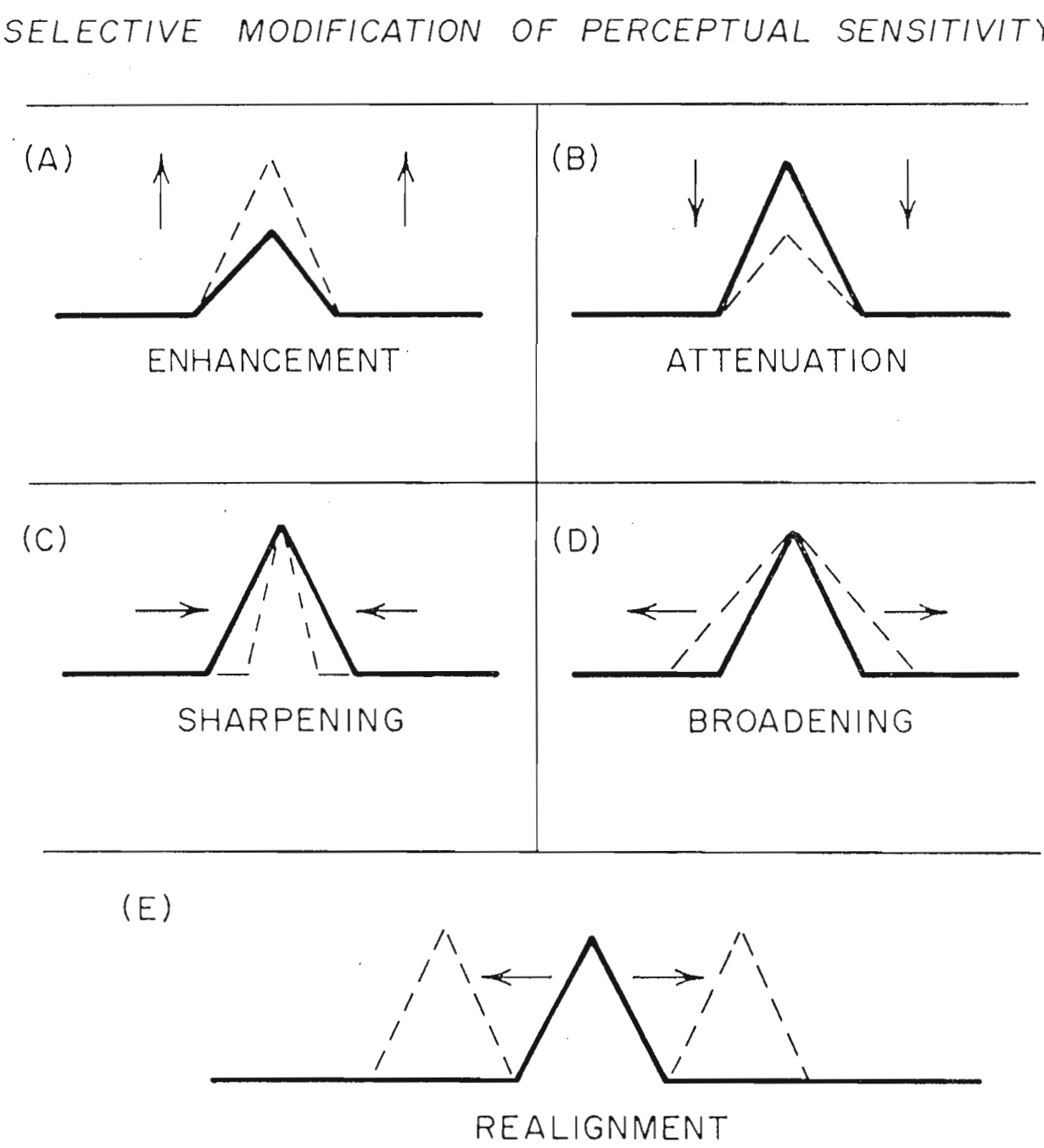

SELECTIVE MODIFICATION OF PERCEPTUAL SENSITIVITY
(A)
ENHANCEMENT
(B)
ATTENUATION
(C)
SHARPENING
(D)
BROADENING
(E)
REALIGNMENT

FIGURE 9.

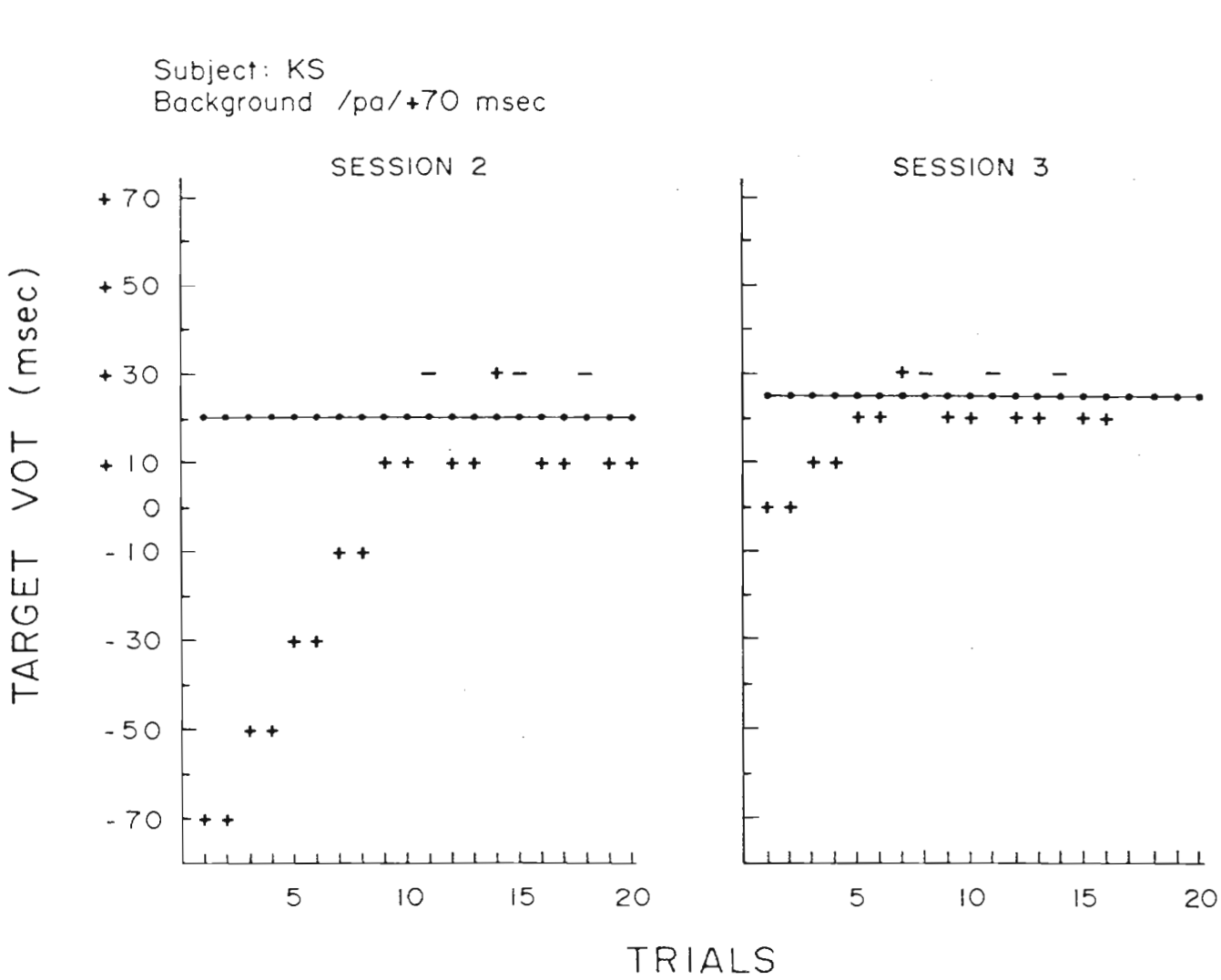

STAIRCASE DISCRIMINATION PROCEDURE
Subject: KS
Background /pa/+70 msec
SESSION 2
SESSION 3
TARGET VOT (msec)
+70
+50
+30
+10
0
-10
-30
-50
-70
5
10
15
20
TRIALS

FIGURE 10.

[RESEARCH ON SPEECH PERCEPTION Progress Report No. 4 (1977-78) Indiana University]

SUSCEPTIBILITY OF A STOP CONSONANT TO ADAPTATION ON A SPEECH-NONSPEECH CONTINUUM: FURTHER EVIDENCE AGAINST FEATURE DETECTORS IN SPEECH PERCEPTION

Robert E. Remez

Department of Psychology

Indiana University

SHORT TITLE: STOP CONSONANT SPEECH-NONSPEECH ADAPTATION

Abstract

The present experiment uses the perceptual adaptation paradigm to establish the validity of a previous test of the phonetic feature detector model of speech perception. In the present study a synthetic stimulus series varied from a CV, [ba], to a nonspeech buzz. When the endpoint tokens were employed alternatively as adaptors, the category boundary was shifted relative to unadapted identification in each adaptor condition. This result suggests that a prior test which had used a vowel as the speech endpoint was legitimate because a stop consonant, an exemplary speech sound, was also susceptible to perceptual adaptation in a speech-nonspeech context. The phonetic feature detector model predicts, incorrectly, that this outcome is impossible, and therefore this finding may be taken to undermine considerably the feature detector model of speech perception.

One current approach to speech perception, the phonetic feature detector proposal, derives its design from the hierarchical cascade model of neurophysiology (Eimas & Corbit, 1973). This conceptualization typically applies a series of increasingly abstract analyzers to the speech signal, first to extract a sequence of instantaneous spectra of the stimulus; then to derive a set of time-varying auditory patterns from the preceding simpler level of analysis; and finally to detect in the auditory patterns the phonetic features which, in combination, constitute the segmental sequence of the talker`s message. This characterization of speech perception postulates that the highest level of analysis consists of a small set of phonetic analyzers, the inventory of which is described by the universal distinctive feature set from linguistic theory. As such, the proposal suggests that the specializations of the neurological substrate for speech perception are matched to the elements of the sound pattern of language (see Cooper, 1975, and Eimas & Miller, 1978, for reviews). With the addition of the single assumption that the specialized detectors develop innately, this hypothesis has also been used to explain the precocious speech perception abilities of human newborns (Eimas, Siqueland, Jusczyk & Vigorito, 1971; Eimas, 1975). Moreover, the binary nature of many phonemic contrasts (Chomsky & Halle, 1968) is easily incorporated into the model through an opponent-process design in which pairs of feature detectors are yoked in oppositions corresponding to the linguistic analysis.

The notion of a genetically endowed set of hypercomplex cells tuned one-to-one to the distinctive features of speech parsimoniously addresses these theoretical and empirical considerations in perception, phonetics, development and neurophysiology.

The behavioral test of this model has used the technique of perceptual adaptation, which was understood as the induction of fatigue in selected parts of the detector ensemble. The ensuing perceptual changes----category boundary shifts[1] along acoustic continua of speech sounds----were well predicted by distinctive feature theory in the original experiment. This gave credence to the claim that the particular assortment of neural perceptual analyzers was identical point for point to the analytic units described by linguists. Indeed, several subsequent studies appeared to require an abstract phonetic level of analysis to explain the adaptation (Diehl, 1975; Ganong, 1975; Rudnicky & Cole, 1977; Sawusch, 1977). The inclusion of a phonetic feature level of analysis in the model therefore appeared to be justified empirically. However, a variety of adaptation effects fail to find adequate explanation in a phonetic detector account, and the set of distinctive features cannot be used to predict the conditions under which adaptation does (or does not) take place (e.g., Ades, 1974; Bailey, 1975; Tartter & Eimas, 1975; Cutting, Rosner & Foard, 1976; Ganong, 1978; Hall & Blumstein, 1978). The occurrence of adaptation would in these latter cases be explained as the result of auditory properties common to the fatiguing stimulus and the continuum used to test the parameters

of fatigue. The difficulty in determining the level of perceptual analysis at which adaptation occurs and the frequent failure of the phonetic feature account to predict the conditions under which adaptation takes place have detracted from the otherwise appealing simplicity of the original conception.

In addition to the inability of the behavioral test to unequivocally establish the role of phonetic feature detectors, a problem of long standing also makes the detector hypothesis less desirable. It has traditionally been difficult to specify the correspondence between acoustic pattern and phonetic segment (Pisoni, 1978, reviews this matter). Although there have been hypotheses advanced about the nature of the acoustic-phonetic correspondence----both those emphasizing contextual aspects (e.g., Dorman, Studdert-Kennedy & Raphael, 1977) and those emphasizing context independence (e.g., Stevens, 1975)----no collection of descriptions or cues suffices yet for more than a small set of features in more than a restricted set of contexts. Under the circumstances, the invoking of feature detectors as the performers of operations which can hardly be described seems entirely beside the point. And, on the adaptation test itself, the argument has recently been made that a change in identification performance occasioned by adaptation may have less to do with normal perception than with the phenomenal contrast between adaptor and test continuum inspired by the test situation (Diehl, Elman & McCusker, 1978; Simon & Studdert-Kennedy, 1978). As evidence accumulates against a simple distinctive feature rationale borrowed from linguistics, we may well wish to question

the assumptions which have permitted adaptation to be interpreted as the change in excitability of individual neural units set to detect the primitives of linguistic analysis.

In the spirit of the reservations about the feature detector approach, a test was recently performed (Remez, 1979) which demonstrated phonetic-like adaptation of a category boundary between a speech sound (the vowel [æ]) and a nonspeech sound (a buzz). In this experiment the [æ] and buzz endpoints were used alternatively as adaptors in the adaptation paradigm. The boundary shifts which were noted for each adaptor on this speech/nonspeech continuum were comparable to those observed on continua of speech sounds only, but could not be explained by resort to a feature detector explanation of any kind. This is because fatigue within the speech detector set should not affect the perception of nonspeech sounds, which would be mediated independently, it is presumed, by a different, nonspeech set of detectors. The finding could only be assimilated by the phonetic detector model (1) if its feature detectors included a [+speech], [-speech] distinction, or (2) if the speech detectors were individually connected opponent-style to a corresponding set of nonspeech detectors. Neither of the alternatives affords a plausible extension of the detector inventory. The first proposal, which essentially claims that a new feature was discovered, could be eliminated on linguistic grounds; no language makes phonemic use of a distinction between abstract segments that are speech and those that are not. The phonemic use of a vocal sound would necessarily confer a [+speech] status

to that sound, despite any similarity the sound might have to nonspeech whistles, clicks, hisses or buzzes. The second proposal, that the speech detector set is merely a component of a larger detector ensemble responsive to speech and nonspeech sounds alike, could be eliminated on neuropsychological grounds (reviewed by Hecaen & Albert, 1978); the higher cortical mediation of speech and language appears to be anatomically segregated from the nonspeech auditory counterpart. In rejecting these alternative explanations for the finding of speech-nonspeech adaptation, it was concluded that the adaptation technique could not have been tapping a fundamental, fixed analytic structure reserved for speech perception. To conclude otherwise would unacceptably require the rejection of the relevant linguistic analyses or the contradiction of the appropriate neuropsychology.

The principal argument of the present report is that the speech-nonspeech test of the earlier study, though it may have been suggestive, may not have posed a challenge to the detector model. This is because the speech sound used in the test continuum was a vowel, and vowels appear to be the least speech-like of speech sounds (Liberman, 1970). Unlike the consonants, vowels are not perceived categorically (Fry, Abramson, Eimas & Liberman, 1962); vowels do not exhibit a right-ear superiority when presented dichotically, though consonants do (Studdert-Kennedy & Shankweiler, 1970); memory processes for consonants and vowels differ (Crowder, 1971; Pisoni, 1975). Vowels and consonants also appear to be

distinctly different species of motor acts (Ohman, 1966, 1967; Perkell, 1969; Fowler, Rubin, Remez & Turvey, in press). In short, the precondition of the test of Remez (1979), that one of the continuum endpoints be speech and the other not, may not have been met. On the evidence just cited, we might presume vowel feature detectors to be closely related to nonspeech processes. If so, then adaptation along the [æ]-buzz continuum may actually have been mediated by a generalized nonspeech processor which happened to include vowel feature detectors. The speech-nonspeech adaptation test employing a vowel as the speech sound would lose its force if that argument is credible, and the phonetic feature detector model of speech (i.e., consonant) perception would remain secure, at least with respect to the arguments and evidence presented earlier by Remez (1979).

To insure that the test series ranging from speech to nonspeech be constructed properly, the present experiment uses a continuum in which the speech sound is a CV, [ba], a synthetic token with formants of narrow bandwidth. The nonspeech sound here is a buzz, a synthetic sound which has formants of broad bandwidth. The prediction of the phonetic feature detector model is that fatigue in the analytic "cascade" leading up to and including the speech feature detectors should not affect the identification of the [ba], because there is no mechanism by which the fatigued speech detectors are integrated with or compared to the nonspeech detectors. In other words, adaptation should not occur. This prediction would be the same even if the actual site of adaptation were sensory, as many authors have

claimed (see above). Because the range of alternatives for the speech processor is presumed to be fixed by the early linguistic experience of the listener (Eimas, 1975), it is not free to undergo demand-specific alteration occasioned by particular tasks. The perceptual analysis of speech by feature detectors, whether it is made on the basis of fatigued or unfatigued sensory data, is defined to be fixed, and independent of processes sensitive to nonspeech sounds.

In summary, the present experiment extends the prior finding of Remez (1979) by using a speech-nonspeech test series in which the speech endpoint is a CV. Because stop consonants are exemplary speech sounds, we can be reasonably certain that if detectors mediate speech perception, this stimulus series will receive detector mediation, at least in the [ba] portion. Unlike our earlier test which employed a vowel, this consonantal version is an appropriate test of the hypothesis that the speech signal is processed by an ensemble of phonetic detectors tuned one-to-one to the distinctive features of linguistic analysis.

METHOD

Subjects. Eight Indiana University undergraduate students served as listeners in the experiment. They had been recruited by handbill advertising, and were paid for their participation. All were native speakers of English, and none had a history of impaired speech or hearing.

Stimuli. A software version of the Klatt digital synthesizer was used in cascade mode to create the stimuli. This program runs on a PDP-11 computer in the Speech Perception Laboratory of the Psychology Department, Indiana University. The acoustic continuum of ten tokens varied from the CV [ba] to a nonspeech buzz. All synthesis parameters with the exception of formant bandwidth were identical in each of the tokens. The rising formant transition pattern characteristic of [o] in syllable-initial position was made by starting F1 at 400Hz, F2 at 1000Hz and F3 at 2400Hz, and linearly increasing the value of each over 30msec to 700Hz, 1200Hz and 2600Hz respectively. The values of F4, 3300Hz, and F5, 3700Hz, were constant for the duration of each token, 290msec. These synthesis values for formant frequency were based on a natural utterance of [ba]. F0 was flat at 110Hz throughout each token. Initially, the formant bandwidths were set at 200Hz, which produced the [ba] endpoint. Iterative 50Hz increments in each of the five bandwidth parameters, to 650Hz, were made to create the ten-item continuum. Each token was synthesized at a sample rate of 10KHz, stored on a magnetic disk, and retrieved for presentation during the experimental sessions. Stimuli were output at that time by digital-to-analog conversion with a 5KHz low-pass filter in effect. Smoothed spectra of the endpoints at onset (initial 20 msec) are displayed in Fig. 1.

Insert Figure 1 about here

The listeners sat at carrels in a sound attenuating room. Stimuli were presented over TDH-39 earphones at 80dB SPL, and responses were scored on two-button response modules which were monitored by computer. The buttons were labelled "BA" and "BUZZ."

Procedure. Each listener participated in two sessions, which corresponded to the use of each endpoint as an adaptor. Testing was done in two groups of four subjects; the first group received the [ba] adaptor in the first session and the buzz adaptor in the second, while the second group received the adaptor conditions in the reverse order. Each testing session included three segments: a practice sequence, an unadapted identification test, and an adapted identification test.

An experimental session began with an introductory practice sequence of ten trials, five each of the two endpoint tokens in random order. This segment of the procedure acquainted the listeners with the best example of each category. Identifications were scored as "BA" or "BUZZ" on each trial, and listeners were informed of the "correct" response by the onset of a light alongside the appropriate response button. At the conclusion of this segment, every listener admitted that the distinction between [ba] and buzz was clearly perceptible. Then, the unadapted identification test began.

In the unadapted identification segment of this test, listeners judged the continuum members presented in a random order, fifteen repetitions of each of the ten, with three seconds

of silence between trials. A different random order was used in each of the four occurrences of the identification test (two groups of subjects, two sessions per group). At the conclusion of this segment, subjects were instructed in the routine of the adaptation test. When subjects said that they understood the instructions, the adaptation test began.

The adaptation test contained, first, a block of fifty repetitions of the adaptor token, either the [ba] or the buzz endpoint as the case may have been, with one second of silence between successive presentations. This repetitive stimulation was designed to induce adaptation, hence no response was called for during adaptor blocks. Immediately after the fiftieth adaptor presentation, a "ready" light on the response module went on, and two seconds later the identification series started. A sequence of ten randomly ordered continuum tokens was then presented with three seconds of silence between trials. Subjects responded by depressing the appropriate button on the response module. The "ready" light was offset after the tenth item, at which point another block of fifty adaptor repetitions was started. The alternation of fifty adaptors and ten identifications continued for the duration of the test. Ten-trial sequences within identification blocks were made by randomly ordering a series of 150 trials (ten continuum items by 15 repetitions of each) and dividing it into successive groups of ten trials. A different random order was used in each of the four occurrences of the adaptation test.

RESULTS AND DISCUSSION

A least-squares normal ogive (Woodworth, 1938) was fitted to the data of each of the four tests contributed by each subject. The mean of each ogive could then be used as an estimate of the category boundary. A one-way Analysis of Variance of the ogive means was used to calculate the error term for planned orthogonal comparisons using the t-statistic (Hayes, 1973). Both adaptors produced a significant change in the labelling function in the direction of change predicted by the adaptation rationale. Speech adaptation caused the category boundary to move from a mean baseline value of 372Hz to a mean adapted value of 282Hz, buzz adaptation caused the boundary to move from 398Hz to 470Hz ($p<.01$ in both cases). The group identification functions, drawn for each adaptor condition, appear in Fig. 2. Each point on the curve is the average of 120 trials (fifteen repetitions of each token by eight subjects).

Insert Figure 2 about here

The adaptation of the boundary between [ba] and buzz is neither predicted by nor is compatible with the phonetic feature detector model of speech perception. In that conception, speech sounds are perceived independently of other ongoing activity by a discrete special purpose device tuned to phonetic features. To explain the present result the model would have to incorporate an additional specialization allowing the fatigue of [b] detectors

to affect nonspeech sounds and speech sounds alike. Not only would this require the device to sacrifice the linguistic rationale for the speech processor, but it would also contradict evidence from neuropsychology on the anatomical separation of these two kinds of auditory processes. To take a single neuropsychological example, Hecaen and Albert (1978) review the relative importance of left hemisphere structures in supporting language functions; but, for the recognition and identification of objects by ear, a complementary nonspeech ability, they write that particular areas of the right hemisphere appear to be critical. The inference to be drawn here is that linguistic and nonlinguistic processes may each be supported by distinct cortical areas which are widely separated from one another. Though it would not be absolutely inconceivable that a fine-grained opponents-process network spans speech and nonspeech cortical areas, present neuropsychological understanding would not be encouraging for the conceptualization of a general perceptual structure serving speech as well as nonspeech sounds. We might therefore take the speech/nonspeech adaptation reported here as evidence of a functional bridge between the two domains, rather than as a point of support for a general auditory perceptual structure in which feature detectors for speech sounds interact with feature detectors for nonspeech sounds.

Even supposing that a large part of the adaptation effect may be due to low level auditory fatigue will not solve the problem and allow the retention of the detector model. It would still be necessary to explain why the perception of buzzes should

ever be contingent on the state of the auditory detectors for [ba], and vice versa, if the perceptual processes for categorizing speech sounds and nonspeech sounds are segregated and independent to begin with. The current model accounts for phonetic contingencies by the opponents-process organization of detectors. Perceptual analysis is accomplished through comparing the activity of paired, rival detectors, whether auditory or phonetic. In this manner a sequence of binary "choices" is established in a fixed analytic structure. But, although fatigue could be an auditory phenomenon, in order for it to affect the [ba]-buzz judgment, a level of integration beyond the auditory detectors would be required nonetheless for the analysis. There would be no other way, within the detector framework, for a sound to seem more [ba]-like by virtue of becoming less buzz-like. This result would indicate that the speech processor is potentially linkable to general auditory processes in a nontrivial way, and such a modification is certainly beyond the scope of the detector proposal. It seems, then, that these findings considerably undermine the claim that adaptation reveals the existence of underlying detectors, and perhaps, as well, the notion that feature detectors mediate speech perception.

One reason that the detector model proves critically testable by so simple an experiment is that the sources of direct behavioral evidence in support of it are limited to those which have employed the adaptation test. The rationale for the adaptation test is that fatigue in portions of an analytic sequence cause the eventual output to be altered; the specific

alterations are thought to result from the operation of the detector ensemble without the full participation of some of its members. Thus, the details of perceptual alterations occasioned by fatigue reveal the individual sensitivities of formerly robust analyzers. A test of this kind of "pandemonium" scheme in speech perception looks to see whether the perceptual system will break down along the dimensions established by linguistic analysis; we have seen that the evidence is equivocal and does not verify the conception. Nevertheless, many other studies also seem to warrant the psychological use of distinctive features appropriated from linguistics (e.g., Miller & Nicely, 1955; Singh, 1966; Wickelgren, 1966; Studdert-Kennedy & Shankweiler, 1970) but it should be emphasized that the use of techniques other than adaptation prevents those results from supporting the hypothesis of distinctive feature detectors. The serviceability of distinctive feature descriptions of perceptual or memory experiments does not in itself count in favor of a specific proposal of neural mechanism; any physiological model of speech perception, including those which may avoid a literal interpretation of psycho-neural identity, must be able to derive these phenomena.

In addition to challenging the detector model, the present study also makes use of a property of speech perception which has begun to receive scrutiny only recently (Aslin, Pisoni, Hennessy, & Perey, 1979), namely, perceptual plasiticity. Although it would be very obviously false to allege that every perceptual constancy is in a perpetual state of change, it would seem

equally false to represent categories as fixed, resistant to differentiation. The present experiment shows that no more than ten trials are required to enable listeners to differentiate novel waveforms as speech or nonspeech. The adaptation test at the very least suggests that, in the naming of the continuum items as speech or nonspeech, the categories are treated as contingent. Could this have resulted from the increased salience of those particular physical attributes of the waveforms which in this case distinguish the categories (for instance, multi-peaked spectrum versus flat spectrum)? Such a proposal would be compatible with a channels-of-analysis view (Kay & Matthews, 1972; Simon & Studdert-Kennedy, 1978), taken here to mean that the listener implicitly determines which functional channels in combination are competent to perform the contingency task. However, the assortment of channels which supported categorization here were unlikely to have been developed beforehand for detecting the difference between speech sounds and nonspeech sounds. The spectral differences between the categories of the continuum do not distinguish them as speech/nonspeech, but merely distinguish them as acoustic patterns. It might be argued that listeners know what kind of spectral envelope speech typically has, for example, what the typical peak-to-trough value (in dB) is, given a scalloped spectrum. However, the definitive characteristic of speech as an acoustic stream may be the principle of moment-to-moment change observable across successive spectra: from high amplitude energy to low; from periodic excitation to aperiodic; from diffuse

spectrum to compact. Thus, the distinction between speech and nonspeech may not depend on a characteristic kind of static spectrum. Rather, the description must ultimately confront the apparent continuity of the speech stream in face of continual change in the acoustic pattern, which itself originates in the continually changing configuration of the articulators as speech is uttered. The subjects in the present experiment may have used a spectral measure as the basis for their perceptions, but it is not to be supposed that this criterion suffices generally for distinguishing speech sounds from nonspeech sounds, and therefore may have been adopted specially for use in the task imposed by the experiment.

In conclusion, the finding of susceptibility to adaptation of a stop consonant on a speech-nonspeech continuum offers proper counterevidence to the feature detector model of speech perception. Although the status of vowels as speech may be in dispute, thus mitigating the force of some of our earlier arguments and evidence, stop consonants are acclaimed to be unequivocal speech sounds. Were an ensemble of speech feature detectors operating, the present test would have tapped them. Because the detector model incorrectly predicts that adaptation along the speech-nonspeech continuum is impossible, due to the presumed independence of the speech detectors from other perceptual processes, the present finding may be taken to undermine the conceptualization of phonetic adaptation as neural fatigue; we may no longer want to consider the process of speech perception a kind of analysis performed by a genetically endowed

set of hypercomplex cells tuned one-to-one to detect the distinctive features of linguistic analysis.

References

Ades, A.E. How phonetic is selective adaptation? Experiments on syllable position and vowel environment. Perception & Psychophysics, 1974, 16, 61-66.

Aslin, R.N., Pisoni, D.B., Hennessy, B.L., and Perey, A.J. Identification of a new linguistic contrast. Journal of the Acoustical Society of America, 1979, 65, S113.

Bailey, P.J. Perceptual adaptation in speech. Unpublished dissertation, Cambridge University, 1975.

Chomsky, N., and Halle, M. Sound Pattern of English. New York: Harper & Row, 1968.

Cooper, W.E. Selective adaptation to speech. In F. Restle, R.M. Shiffrin, N.J. Castellan, H.R. Lindman, and D.B. Pisoni (Eds.), Cognitive Theory, Vol. I. Hillsdale: Lawrence Erlbaum Associates, 1975. Pp. 23-54.

Crowder, R.G. The sound of vowels and consonants in immediate memory. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 587-596.

Cutting, J.E., Rosner, B.S., and Foard, C.F. Perceptual categories for musiclike sounds: implications for theories of speech perception. Quarterly Journal of Experimental Psychology, 1976, 28, 361-378.

Diehl, R. The effect of selective adaptation on the identification of speech sounds. Perception & Psychophysics, 1975, 17, 48-52.

Diehl, R.L., Elman, J.L., and McCusker, S.B. Contrast effects on stop consonant identification. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 599-609.

Dorman, M.F., Studdert-Kennedy, M., and Raphael, L.J. Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent context-dependent cues. Perception & Psychophysics, 1977, 22, 109-122.

Eimas, P.D. Speech perception in early infancy. In L.B. Cohen and P. Salapatek (Eds.), Infant Perception. New York: Academic Press, 1975. Pp. 193-231.

Eimas, P.D., and Corbit, J.D. Selective adaptation of linguistic feature detectors. Cognitive Psychology, 1973, 4, 99-109.

Eimas, P.D., and Miller, J.L. Effects of selective adaptation on the perception of speech and visual patterns: evidence for feature detectors. In R.D. Walk and H.L. Pick, jr. (Eds.), Perception and Experience. New York: Plenum, 1978.

Eimas, P.D., Siqueland, E.P., Jusczyk, P., and Vigorito, J. Speech perception in infancy. Science, 1971, 171, 303-306.

Fowler, C.A., Rubin, P.E., Remez, R.E., and Turvey, M.T. Implications for speech production of a general theory of action. In B. Butterworth (Ed.), Language Production. New York: Academic Press. In press.

Fry, D.B., Abramson, A.S., Eimas, P.D., and Liberman, A.M. The identification and discrimination of synthetic vowels. Language and Speech, 1962, 5, 171-189.

Ganong, W.F. An experiment on "phonetic adaptation." Progress Report, Research Laboratory of Electronics, Massachusetts Institute of Technology, 1975, 116, 206-210.

Ganong, W.F., III. The selective adaptation effects of burst-cued stops. Perception & Psychophysics, 1978, 24, 71-83.

Hall, L.L., and Blumstein, S.E. The effect of syllabic stress and syllabic organization on the identification of speech sounds. Perception & Psychophysics, 1978, 24, 137-144.

Hayes, W.L. Statistics for the Social Sciences. New York: Holt, Rinehart and Winston, 1973.

Hecaen, H., and Albert, M.L. Human Neuropsychology. New York: Wiley, 1978.

Kay, R.H., and Matthews, D.R. On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones. Journal of Physiology, 1972, 225, 657-677.

Liberman, A.M. Some characteristics of perception in the speech mode. In Perception and its Disorders, Research Publications of the A.R.N.M.D., vol. 48. Baltimore: Williams & Wilkins, 1970. Pp. 238-254.

Miller, G.A., and Nicely, P.E. An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 1955, 27, 338-352.

Ohman, S.E.G. Coarticulation in VCV utterances: Spectrographic measurements. Journal of the Acoustical Society of America, 1966, 39, 151-168.

Ohman, S.E.G. Numerical model of coarticulation. Journal of the Acoustical Society of America, 1967, 41, 310-320.

Perkell, J.S. Physiology of Speech Production. Cambridge: MIT Press, 1969.

Pisoni, D.B. Auditory short-term memory and vowel perception. Memory & Cognition, 1975, 3, 7-18.

Pisoni, D.B. Speech perception. In W.K. Estes (Ed.), Handbook of Learning and Cognitive Processes, Vol. 6: Linguistic Functions in Cognitive Theory. Hillsdale: Lawrence Erlbaum Associates, 1978. Pp. 167-233.

Remez, R.E. Adaptation of the category boundary between speech and nonspeech: A case against feature detectors. Cognitive Psychology, 1979, 11, 38-57.

Rudnicky, A.I., and Cole, R.A. Adaptation produced by connected speech. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 51-61.

Sawusch, J.R. Selective adaptation effects on end-point stimuli in a speech series. Perception & Psychophysics, 1976, 20, 61-65.

Sawusch, J.R. Peripheral and central processes in selective adaptation of place of articulation in stop consonants. Journal of the Acoustical Society of America, 1977, 62, 738-750.

Simon, H.J., and Studdert-Kennedy, M. Selective anchoring and adaptation of phonetic and nonphonetic continua. Journal of the Acoustical Society of America, 1978, 64, 1338-1357.

Singh, S. Cross-language study of perceptual confusions of plosive phonemes in two conditions of distortion. Journal of the Acoustical Society of America, 1966, 40, 635-656.

Stevens, K.N. The potential role of property detectors in the perception of consonants. In G. Fant and M.A.A. Tatham (Eds.), Auditory Analysis and Perception of Speech. New York: Academic Press, 1975. Pp. 303-329.

Studdert-Kennedy, M., and Shankweiler, D.P. Hemispheric specialization for speech perception. Journal of the Acoustical

Society of America, 1970, 48, 570-594.

Tartter, V.C., and Eimas, P.D. The role of auditory and phonetic feature detectors in the perception of speech. Perception & Psychophysics, 1975, 18, 293-298.

Wickelgren, W. Distinctive features and error in short-term memory for consonants. Journal of the Acoustical Society of America, 1966, 39, 388-398.

Woodworth, R.S. Experimental Psychology. New York: Holt, 1938.

Footnotes

Address correspondence to Robert E. Remez, Department of Psychology, Indiana University, Bloomington, Indiana 47405.

This research was supported by NIMH Grant MH 32848-01 to the author and by NIMH Grant MH 24027-05 and NIH Grant NS 12179-03 to David B. Pisoni. I would like to thank Professor Pisoni and Steve Simnick for their assistance during the course of this study.

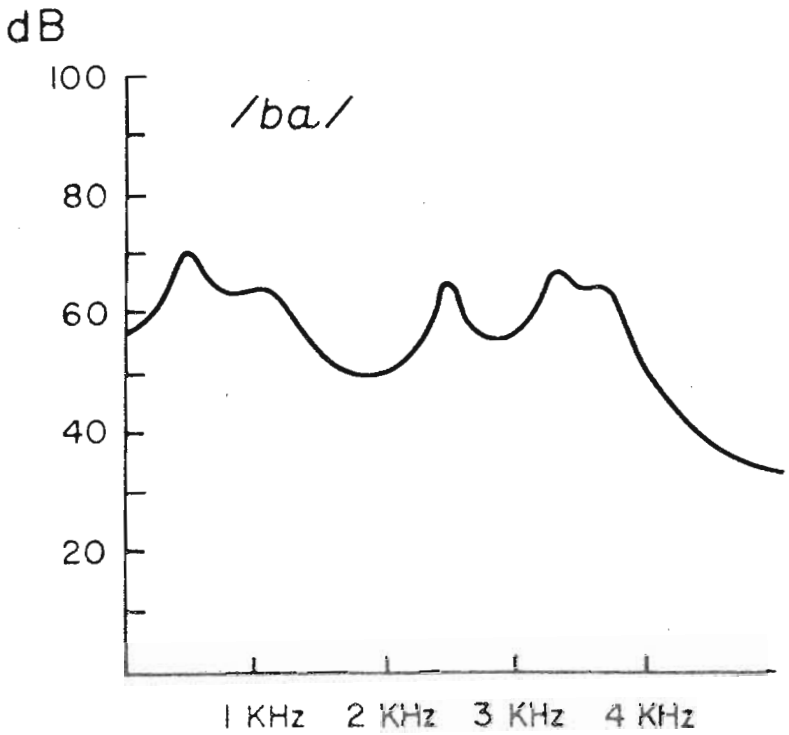
[1]The use of a rating procedure has shown that the effect of adaptation can be observed throughout an entire category, not only at the boundary stimuli (Sawusch, 1976).

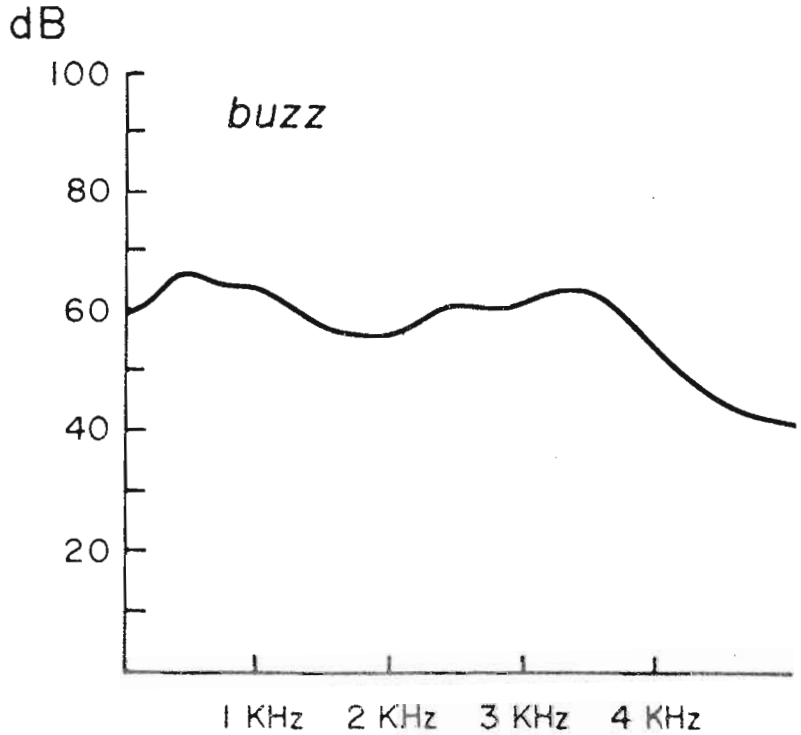
FIGURE CAPTIONS

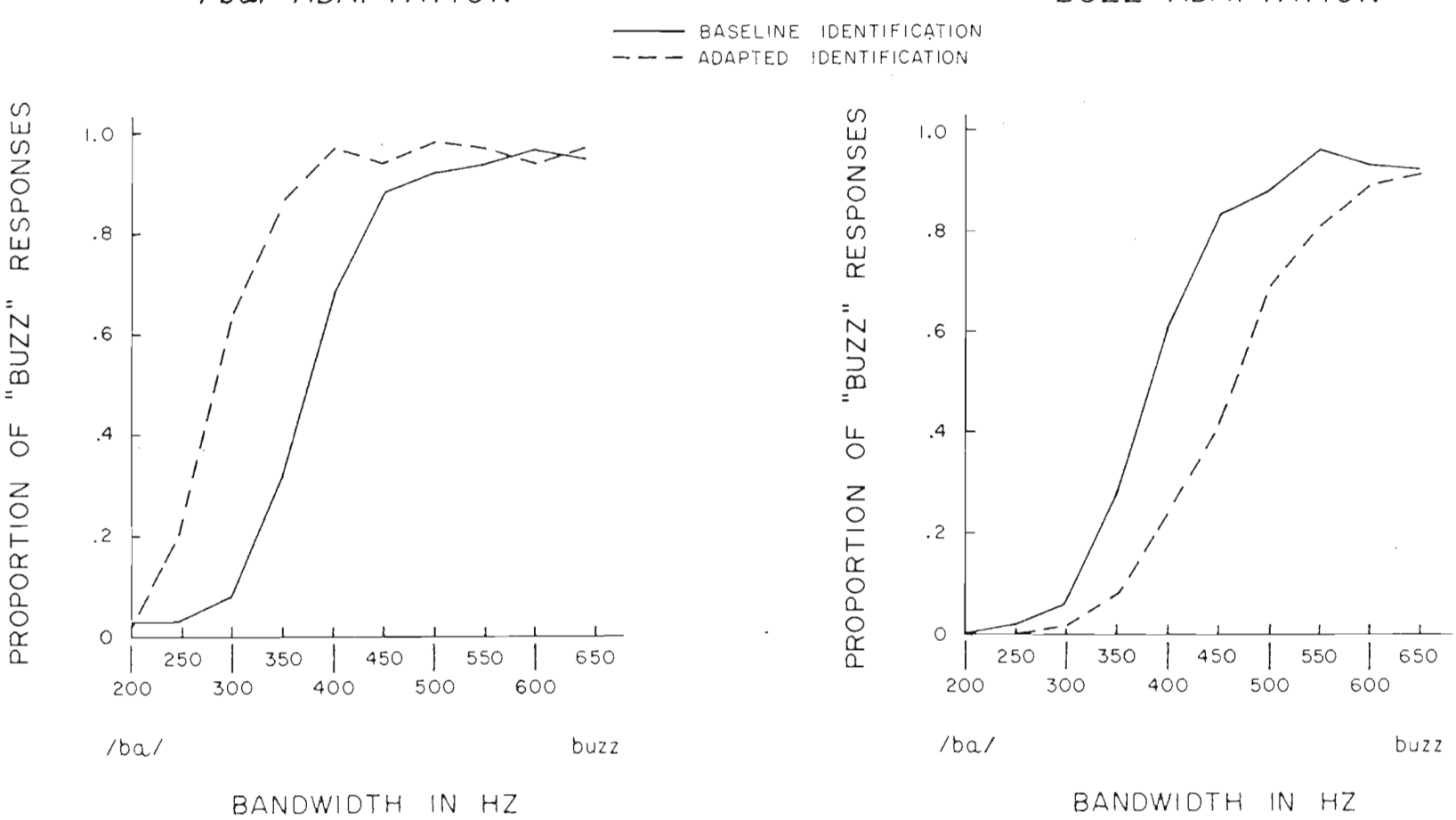
Figure 1. Spectra of the continuum endpoints at onset: [ba] (left panel) and buzz (right panel).

Figure 2. Baseline and adapted identifications in two adaptor conditions: [ba] adaptor (left panel) and buzz adaptor (right panel).

SPECTRUM AT ONSET (INITIAL 20 MSEC)






/ba/ ADAPTATION
BUZZ ADAPTATION
BASELINE IDENTIFICATION
ADAPTED IDENTIFICATION
PROPORTION OF "BUZZ" RESPONSES
1.0
.8
.6
.4
.2
0
200
250
300
350
400
450
500
550
600
650
/ba/
buzz
BANDWIDTH IN HZ

SHORT REPORTS AND WORK-IN-PROGRESS

Dual Processing vs. Response-Limitation Accounts of Categorical Perception: A Reply to Macmillan, Kaplan and Creelman*

A. J. Perey and D. B. Pisoni

Department of Psychology

Indiana University

Bloomington, Indiana 47401

Abstract

At the last meeting of the Acoustical Society, Macmillan, Kaplan and Creelman argued that measures of discrimination typically obtained in speech perception experiments employing single-interval absolute identification tasks do not reveal the "true sensitivity" of the observer because only a very small number of response categories are permitted and identification into these putative phonological categories is often without error. In the present paper we present the results of two experiments, one employing synthetic stop consonants and the other employing synthetic steady-state vowels, in which observers were permitted to use a larger number of response categories in an absolute identification (AI) procedure similar to that used by Braida and Durlach. Discrimination performance in the identification task was then compared to discrimination obtained in the traditional ABX format in order to assess the correspondence between the two measures of discrimination. The results revealed systematic differences in discrimination between consonants and vowels and between AI and ABX procedures. While improved discrimination performance was observed in the AI task for vowels by permitting a larger number of response categories in identification, the same effects were not observed for stop consonants. The results suggest that the account of categorical perception offered by Macmillan et al. in terms of an equivalence of the spacing of signals in identification and discrimination is not sufficient to capture the observed differences between stimuli showing tendencies toward either categorical or continuous modes of speech perception. Our results and analyses also indicate that dual coding models of speech perception which assume separate and distinct levels of coding phonetic information in speech perception cannot be explained away simply on the basis of the results obtained in identification and discrimination experiments as supposed earlier by Macmillan et al.

*This paper was presented at the 94th meeting of the Acoustical Society of America, Miami Beach, Florida, December 15, 1977. This research was supported by NIH grant NS-12979-03 and NIMH grant MH-24027-03 to Indiana University.

Dual Processing vs. Response-Limitation Accounts of Categorical Perception: A Reply to Macmillan, Kaplan and Creelman

A. J. Perey and D. B. Pisoni

Department of Psychology

Indiana University

Bloomington, Indiana 47401

In the last few years there has been a renewed and quite vigorous interest in categorical perception of speech and other complex signals with particular attention focused on the psychophysical, perceptual and cognitive factors that may underlie these results. At the last meeting of the Society at State College, Macmillan, Kaplan and Creelman presented a theoretical analysis of several previous accounts of categorical perception as they have been applied to both speech and nonspeech continua and offered a number of interesting suggestions concerning the ways in which these different views may be distinguished experimentally. Their ideas have been amplified and extended in a number of directions in a paper published recently in Psychological Review where the claim is made that categorical perception should be redefined as simply "the equivalence of discrimination and identification measures." More specifically, Macmillan et al. state that "a stimulus dimension is categorically perceived if the perceptual spacing of signals along that dimension is the same in discrimination as in identification."

The purpose of the present paper is to examine, in greater detail, several of the factors that led Macmillan et al. to offer this redefinition of categorical perception and to report some relevant new data that we think bear on these particular issues.

The results of our experiments suggest that the "modified" definition of categorical perception--"that the spacing of signals along a dimension in identification is equal to the spacing of these signals in discrimination" is inappropriate and potentially misleading because it fails to capture differences in discrimination between continua that have traditionally been thought to show tendencies toward either categorical or continuous modes of perceptual analysis. Moreover, our results show that the analyses of identification and discrimination experiments proposed by Macmillan et al. cannot distinguish between dual coding models and single dimensional models of categorical perception.

The issue we consider first in this paper concerns the relationship between identification and discrimination tasks used in speech perception experiments. Macmillan et al. suggested that discrimination performance measured in a single AI task often fails to correspond to discrimination performance measured in a multiple-interval procedure such as ABX. That is, it is generally agreed that observers do not resolve differences between stimuli as well in identification as they do in discrimination. Macmillan et al. suggest that this result is often observed in speech perception experiments because the subject is permitted to use only a very small number of response alternatives in identification. These response alternatives represent, of course, the "distinctive" phonological categories of the subjects' language such as /b/ or /p/ -- /i/, /I/, /ɛ/ -- or /b/, /d/, /g/.

The specific empirical question that we were interested in is quite straight-forward--can subjects display better resolution in identification as compared to ABX discrimination if a larger number of response alternatives

are available? To answer this question we followed up the recommendation suggested by Macmillan et al. and carried out an AI experiment of the kind used previously by Braida and Durlach (1972) in their intensity experiments which would provide us with estimates of d' for adjacent stimuli along the continuum. In our experiments, subjects were presented with seven stimuli and were permitted to use seven response alternatives in a rating scale format. Stimuli were presented one at a time for identification.

Insert Figure 1 about here

Figure 1 shows the idealized predictions for two possible outcomes in an AI task such as this. Mean rating response is displayed on the ordinate whereas stimulus value is displayed on the abcissa. Panel (a) on the left represents what we would anticipate from a continuous model. In this case, we would expect that the rating responses in an AI task would be a monotonically increasing function of the stimulus value. In contrast, panel (b) on the right shows the predictions anticipated from a "categorical" model of identification where the rating response would be non-monotonic with the physical scale. Such an outcome would indicate that the subjects cannot use the additional response alternatives in the AI task to show improved discrimination either because of a tendency toward perceptual categorization or because of a true limitation on resolution in the AI task.

Insert Figure 2 about here

Figure 2 shows the mean rating responses obtained for a set of seven steady-state vowels varying from /i/ to /I/ and for a set of seven stop-vowel syllables varying in VOT (voice onset time) from /ba/ to /pa/. Both are standard sets of synthetic stimuli that have been used in previous speech perception experiments. Notice that the function displayed in the left panel for the steady-state vowels shows a fairly close correspondence to the function expected from the continuous model shown in the previous slide. Subjects can, in fact, order their responses to steady-state vowels more-or-less monotonically with changes in the physical scale. On the other hand, the shape of the rating function for the consonants on the right shows a fairly close correspondence to the predictions of the categorical model. Subjects in this condition do not appear to be able to use all of the response alternatives available to them because they fail to map their responses consistently across the stimulus continuum.

Insert Figure 3 about here

Figure 3 shows the individual data for the 10 subjects in the vowel condition which match the group data reasonably well.

Insert Figure 4 about here

Figure 4 shows the individual data for the nine subjects in the consonant condition which also correspond closely to the group data shown earlier.

From the AI data presented so far for vowels and consonants, it seems reasonable to conclude that subjects can, under some conditions, take advantage of a larger number of response alternatives in AI and therefore display substantially better discrimination between signals than when a smaller number of responses is permitted. However, as we have seen from these data, such an outcome will depend on the particular stimulus continuum under consideration (i.e., vowels vs. consonants) and the extent to which any "perceptual categorization" might occur.

To quantify the observed differences in identification between vowels and consonants and to obtain a measure of discrimination for the stimuli in each stimulus series, we computed from the stimulus-response matrix a set of scale values for the stimuli and a set of criteria or "cutoffs" for the response categories. The decision model used here has been outlined in earlier papers by Braida and Durlach (1972) and is essentially the same as Thurstone's "Law of Categorical Judgment" as summarized by Torgerson (1958).

Insert Figure 5 about here

Figure 5 shows the results of this analysis for vowels in the upper panel and consonants in the lower panel. This analysis assumes that each stimulus in the series gives rise to a unit-variance normal distribution spaced along some continuous internal dimension. Notice that the distributions for the vowel stimuli are spaced more-or-less equally along the perceptual continuum, whereas the distributions for the consonants are displaced toward both ends of the range. A measure of discrimination

between any two stimuli on the continuum can be computed from the scale values displayed in this figure by simply finding the distance between the means of the two distributions. This "discrimination distance" is assumed to be equivalent to the d' that would have been obtained in a Yes-No discrimination experiment. Following Braida and Durlach, let us call this measure of discrimination the "identification distance" of the stimuli in each series.

In order to compare resolution in a single interval AI experiment with resolution in a multiple-interval experiment, we also collected 1- and 2-step ABX discrimination functions from the subjects in the vowel and consonant conditions. From the response frequencies of the various ABX triads, we then computed hit and false alarm rates for adjacent pairs of stimuli along the continuum. These probabilities were then converted into Yes-No d's assuming equal variance normal distributions and "no bias" in the ABX discrimination task. We will call this measure of discrimination the "discrimination distance" to distinguish it from the "identification distance" obtained in the AI experiment.

Now, according to Macmillan et al., if the two distances--the "identification distance" and the "discrimination distance" coincide, the two experiments would therefore yield the same information about the subjects' "perception" of these stimuli. Such an outcome could then be taken, in their view, as evidence for categorical perception within the context of a continuous TSD model rather than a discrete low-threshold model as assumed by the earlier Haskins account of categorical perception.

Insert Figure 6 about here

Figure 6 shows the cumulative d's for the 1-step stimulus comparisons obtained in ABX discrimination (open triangles) and in the AI experiment (filled circles). For the vowel condition as shown in the left panel, discrimination is much better in ABX than AI. The total discrimination distance is 5.0 whereas the total identification distance is 4.2, a difference of about 20%. Thus, for the vowel continua, even with the additional response alternatives in identification, discrimination distance still exceeds identification distance. In this case, according to Macmillan et al. one would therefore be justified in claiming noncategorical or "continuous" perception for these signals, a result that is not entirely unexpected with steady-state vowels.

Turning to the cumulative d's for the consonants on the right, we can see immediately that the correspondence between the two curves is substantially better than the fit for the vowels. Total discrimination distance is now 4.4 as compared to a total identification distance of 4.3. Based on these results, we would be justified, according to Macmillan et al., in concluding that the consonants are perceived "categorically" because the spacing of signals in identification is equivalent, or very nearly so, to the spacing of signals in discrimination. Again, this is a result that is not entirely unexpected based on previous experiments with stop consonants.

A similar set of curves for the vowels and consonants is obtained when we consider the 2-step stimulus comparisons as shown in Figure 7.

Insert Figure 7 about here

Again, discrimination distance exceeds identification distance for the vowels, shown on the left, although the two curves show a very close fit for the consonants as shown in the right panel.

It may be appropriate at this point to step back for a moment and consider what the continuous model of categorical perception as proposed by Macmillan et al. provides over the previous accounts which have assumed a discrete number of perceptual states. At first glance, the two accounts seem quite similar, at least with regard to capturing differences in identification and discrimination between steady-state vowels and stop consonants, stimuli that have been studied extensively in speech perception experiments. Indeed, if we compare the ability of the Haskins model to predict ABX discrimination from identification with the continuous signal detection model, we find that the continuous model provides a much better fit to the observed ABX data. These results can be seen quite clearly in the following two figures.

Insert Figure 8 about here

This figure shows the observed ABX discrimination functions (open triangles) and the predicted ABX discrimination (filled circles) from the Haskins model under the assumption that a listener can discriminate between two stimuli only to the extent that he can identify these stimuli as different. Notice that the discrepancy between the predicted and obtained ABX functions is present for both vowels and consonants although it is somewhat larger for the vowels.

Insert Figure 9 about here

In this figure the same observed ABX data are plotted for vowels and consonants. The predicted ABX functions shown here are plotted in terms of $P(C)_{MAX}$ which was derived from the d's obtained in the AI experiment. It can be seen that the fit between the observed ABX and the predicted ABX from the AI experiment is now substantially better than the predictions from the Haskins model for both vowels and consonants although the close match for the consonants is particularly noteworthy. Thus, within the context of these experiments, the predictions derived from the continuous model show a much closer fit to the observed ABX discrimination data and suggest that the extreme definition of categorical perception in terms of a discrete number of perceptual states may indeed be incorrect.

However, while the assumptions underlying a continuous TSD model of categorical perception may provide an improvement over previous conceptualizations, the new revised definition of categorical perception offered by Macmillan et al. in terms of an equivalence of the spacing of signals in identification and discrimination is not only misleading but also incorrect in our view. Such a definition fails to capture differences between continua that clearly show quite different tendencies toward categorical and continuous modes of perception.

It is quite easy to imagine a set of signals whose spacing is the same in discrimination and identification yet there is no evidence of perceptual categorization in identification nor the presence of a peak or discontinuity in discrimination. The failure to distinguish between continua of this kind

kind and those which show both attributes of categorical perception, at least as they have been considered in the past, must surely call into quesiton the usefulness of the redefinition offered by Macmillan et al.

It should also be apparent from the results of our experiments and the outcome of the analyses suggested by Macmillan et al., that differences between dual coding and other alternative accounts of speech perception cannot be resolved simply on the basis of identification and discrimination experiments alone.

Thus, we conclude that the discrepancy observed between discrimination and identification for some stimulus continua cannot now be explained away by recourse to constraints or limitations on the number of response alternatives available in identification as supposed by Macmillan et al. The extent to which dual-coding models can account for this discrepancy is a matter for future research to decide.

In summary, we have shown that a continuous model of categorical perception can capture the important differences in perception between vowels and consonants and that it is better able to predict observed discrimination performance from identification. However, we feel that the redefinition of categorical perception offered by Macmillan et al. as simply an equivalence of the spacing of signals in identification and discrimination will fail to capture important differences between stimulus continua that show tendencies toward either categorical or continuous modes of perception. Finally, while dual coding models have often been invoked in speech perception to account for perceptual categorization and other effects, the results from identification and discrimination experiments alone cannot be used to decide their fate relative to other accounts which assume that only a single underlying dimension is relevant.

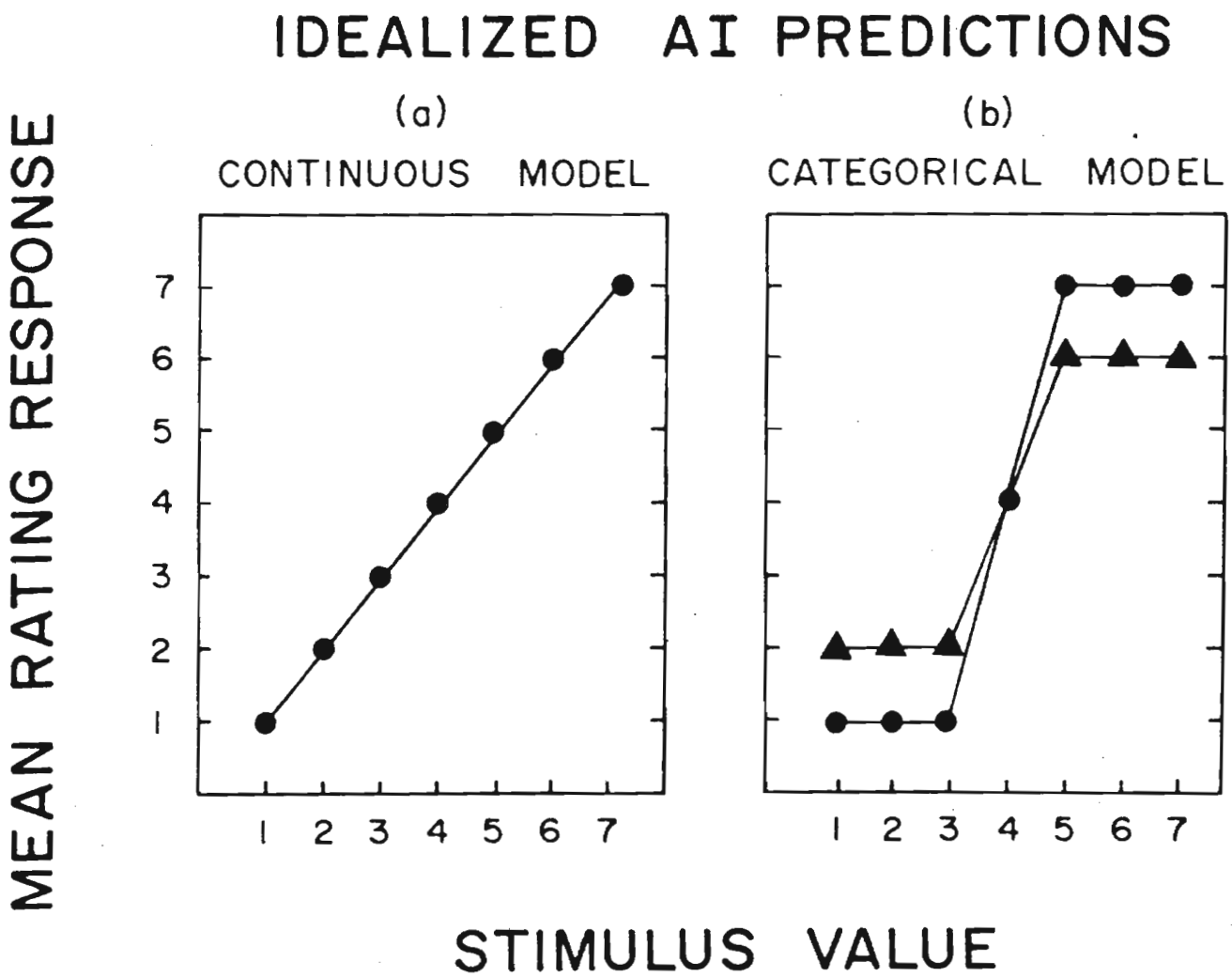

IDEALIZED AI PREDICTIONS
(a)
CONTINUOUS MODEL
(b)
CATEGORICAL MODEL
MEAN RATING RESPONSE
7
6
5
4
3
2
1
1 2 3 4 5 6 7
1 2 3 4 5 6 7
STIMULUS VALUE

Figure 1.

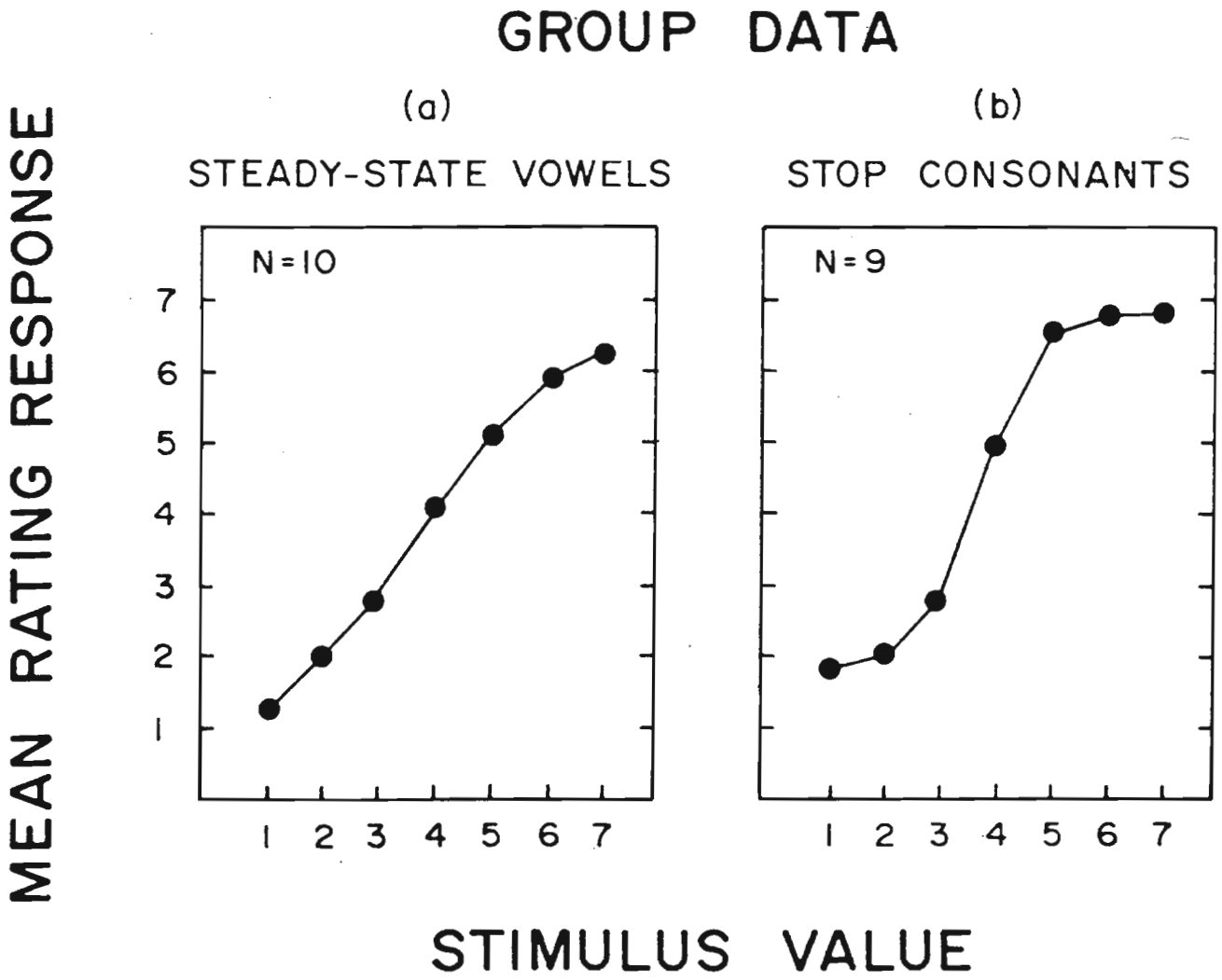

GROUP DATA
(a)
STEADY-STATE VOWELS
N=10
(b)
STOP CONSONANTS
N=9
MEAN RATING RESPONSE
1
2
3
4
5
6
7
1 2 3 4 5 6 7
STIMULUS VALUE

Figure 2.

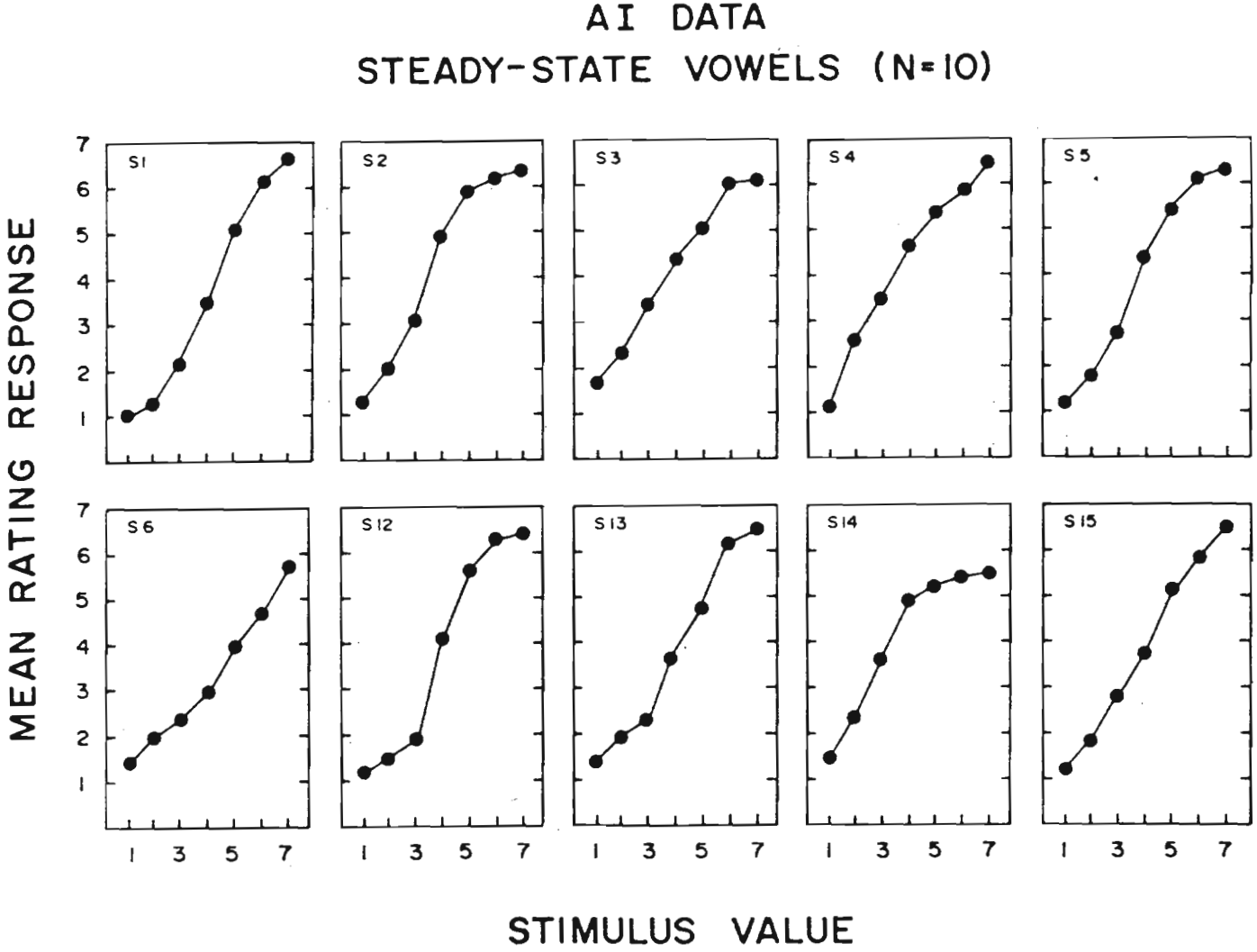

AI DATA
STEADY-STATE VOWELS (N=10)
MEAN RATING RESPONSE
S1
S2
S3
S4
S5
S6
S12
S13
S14
S15
1
2
3
4
5
6
7
STIMULUS VALUE

Figure 3.

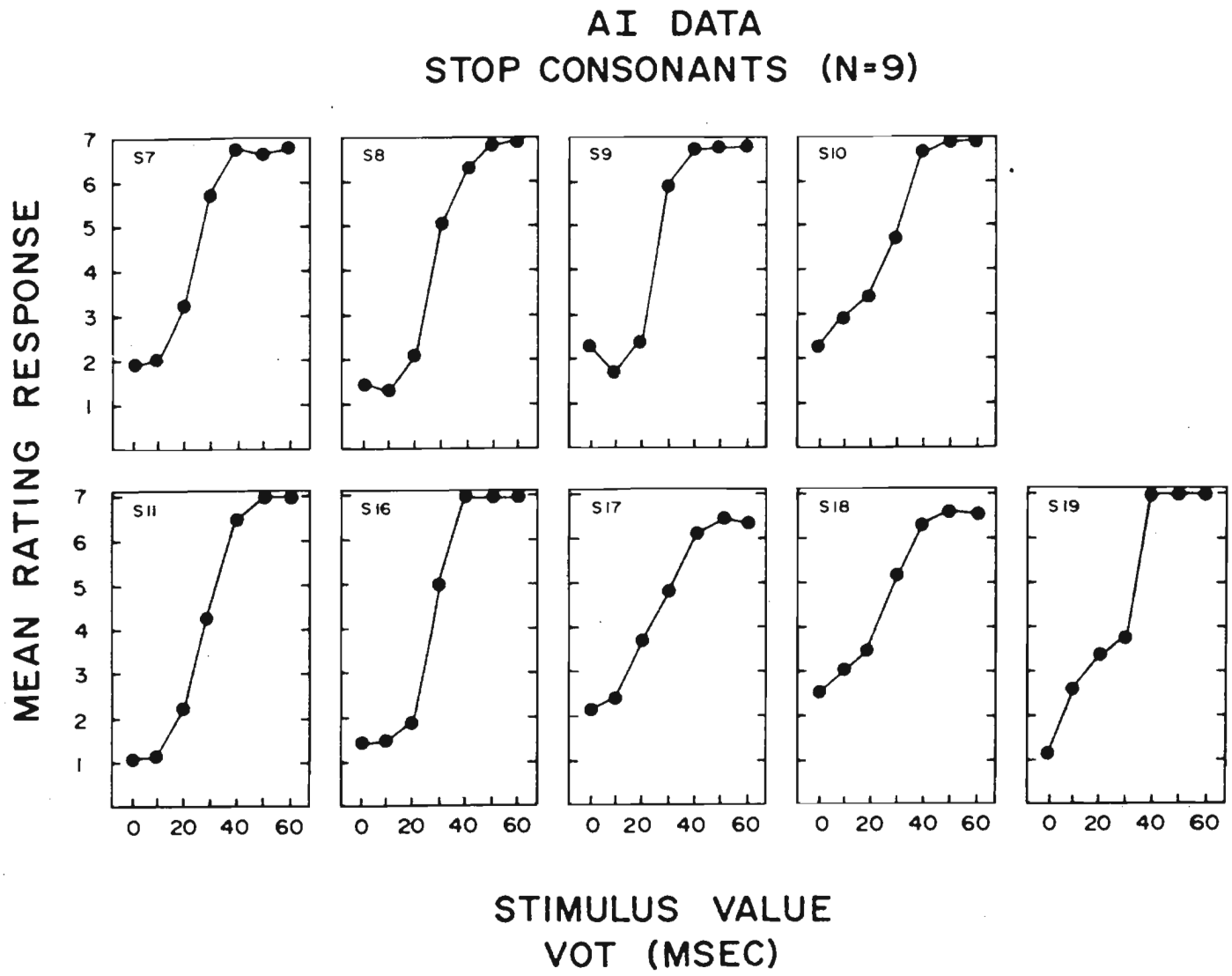

AI DATA
STOP CONSONANTS (N=9)
MEAN RATING RESPONSE
S7
S8
S9
S10
S11
S16
S17
S18
S19
1
2
3
4
5
6
7
0
20
40
60
STIMULUS VALUE
VOT (MSEC)

Figure 4.

SCALE VALUES & RESPONSE CRITERIA FROM IDENTIFICATION

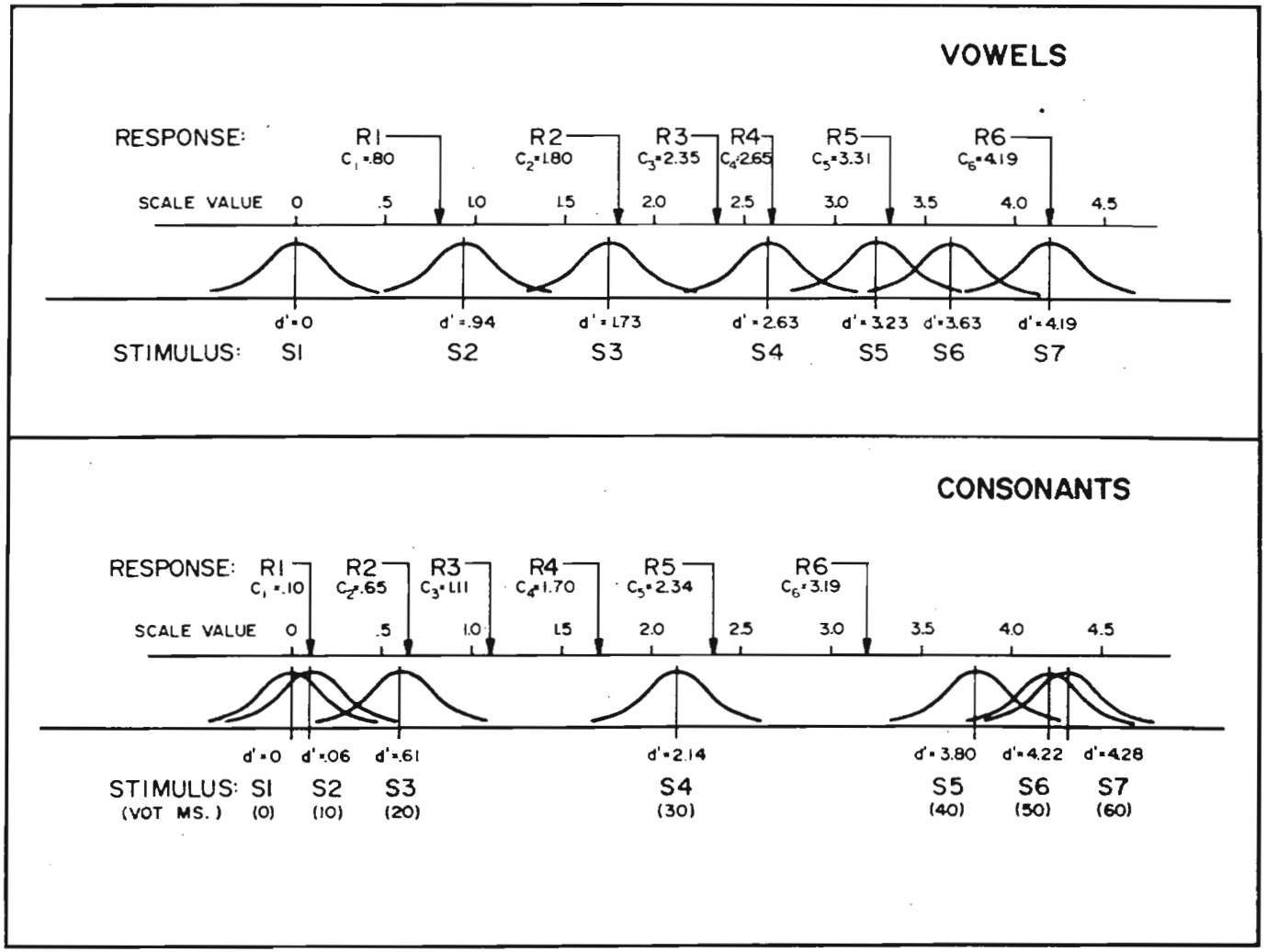


Figure 5.

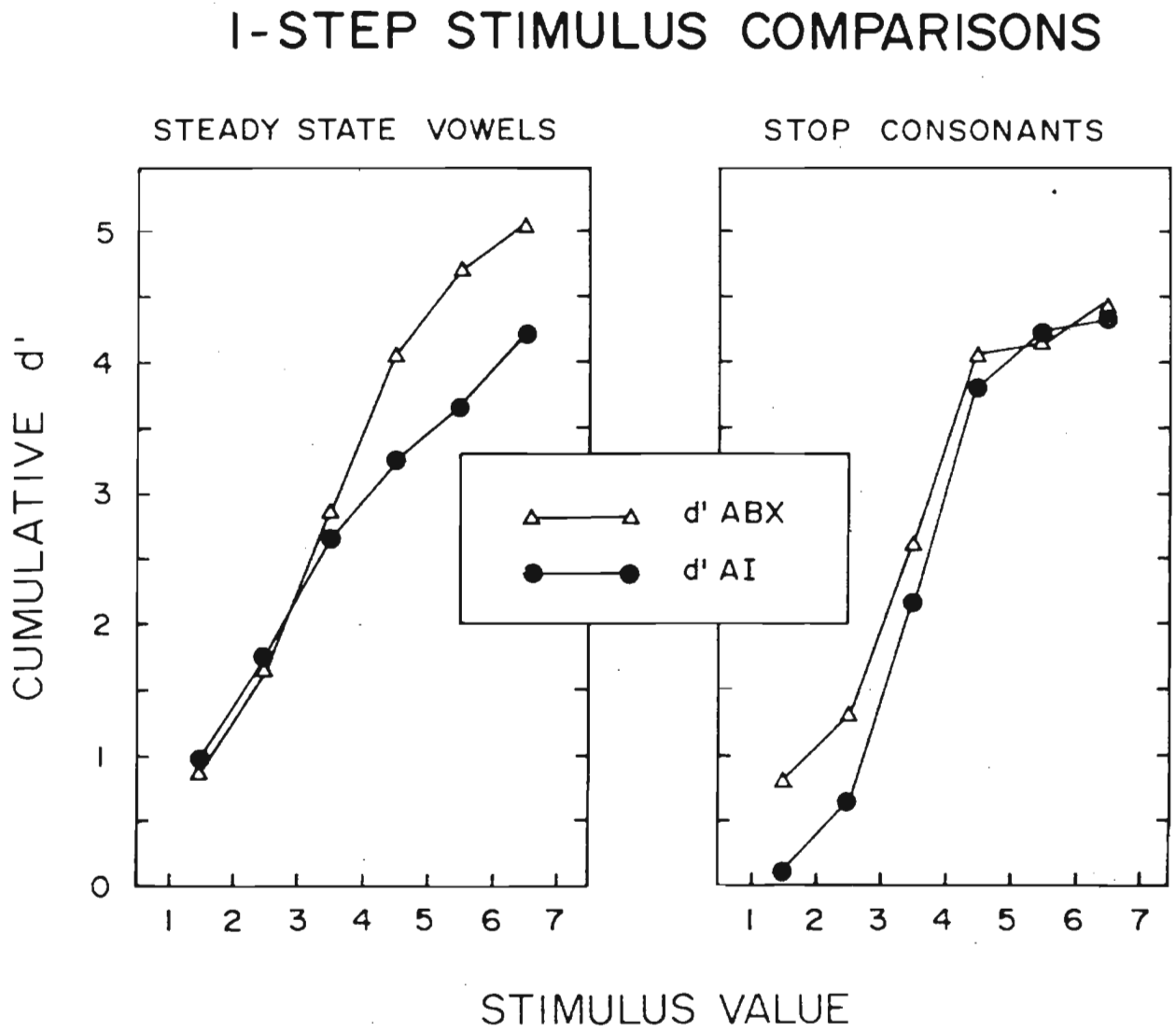

1-STEP STIMULUS COMPARISONS
STEADY STATE VOWELS
STOP CONSONANTS
CUMULATIVE d'
0
1
2
3
4
5
1 2 3 4 5 6 7
d' ABX
d' AI
STIMULUS VALUE

Figure 6.

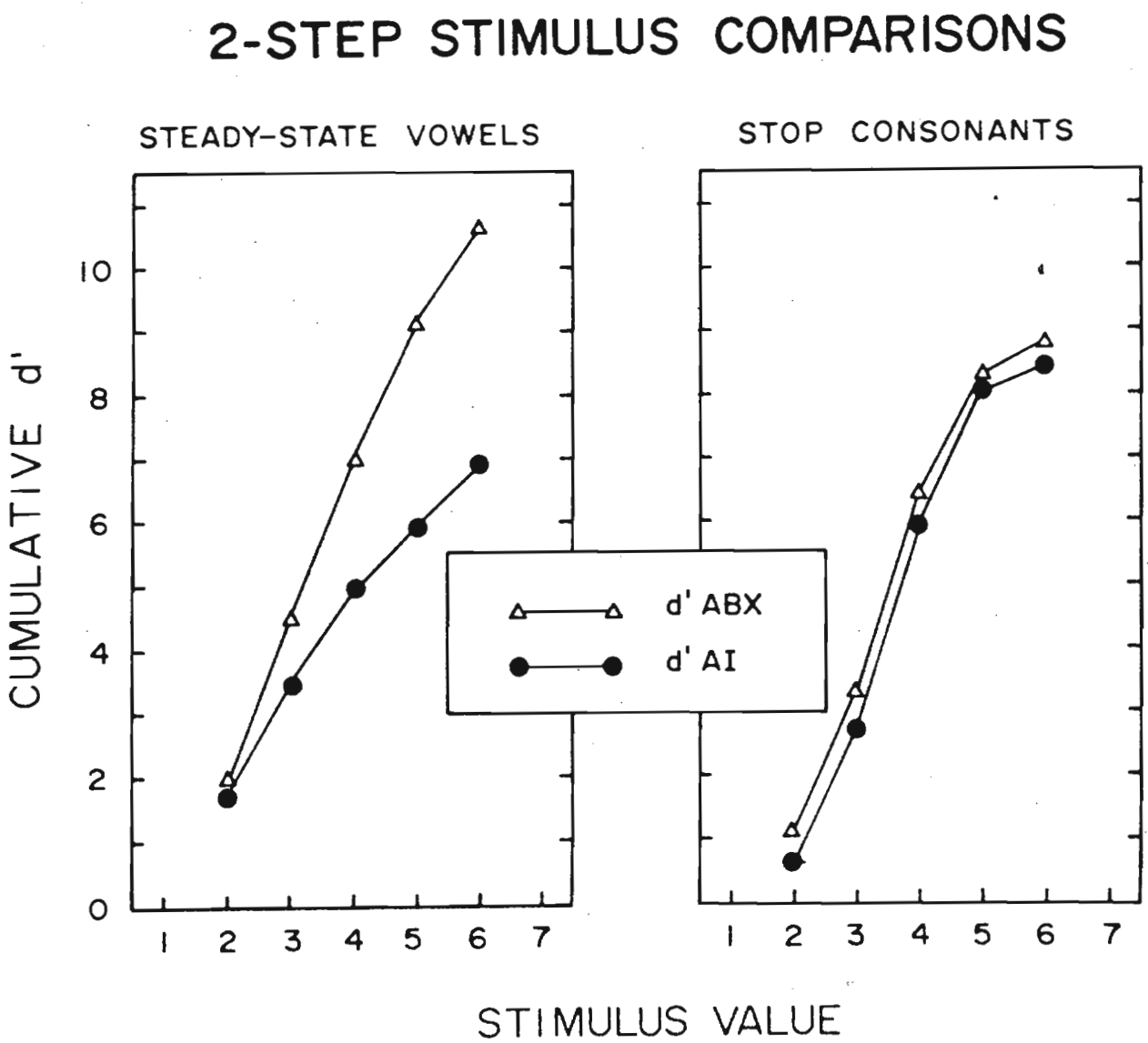

2-STEP STIMULUS COMPARISONS
STEADY-STATE VOWELS
STOP CONSONANTS
CUMULATIVE d'
0
2
4
6
8
10
1 2 3 4 5 6 7
d' ABX
d' AI
STIMULUS VALUE

Figure 7.

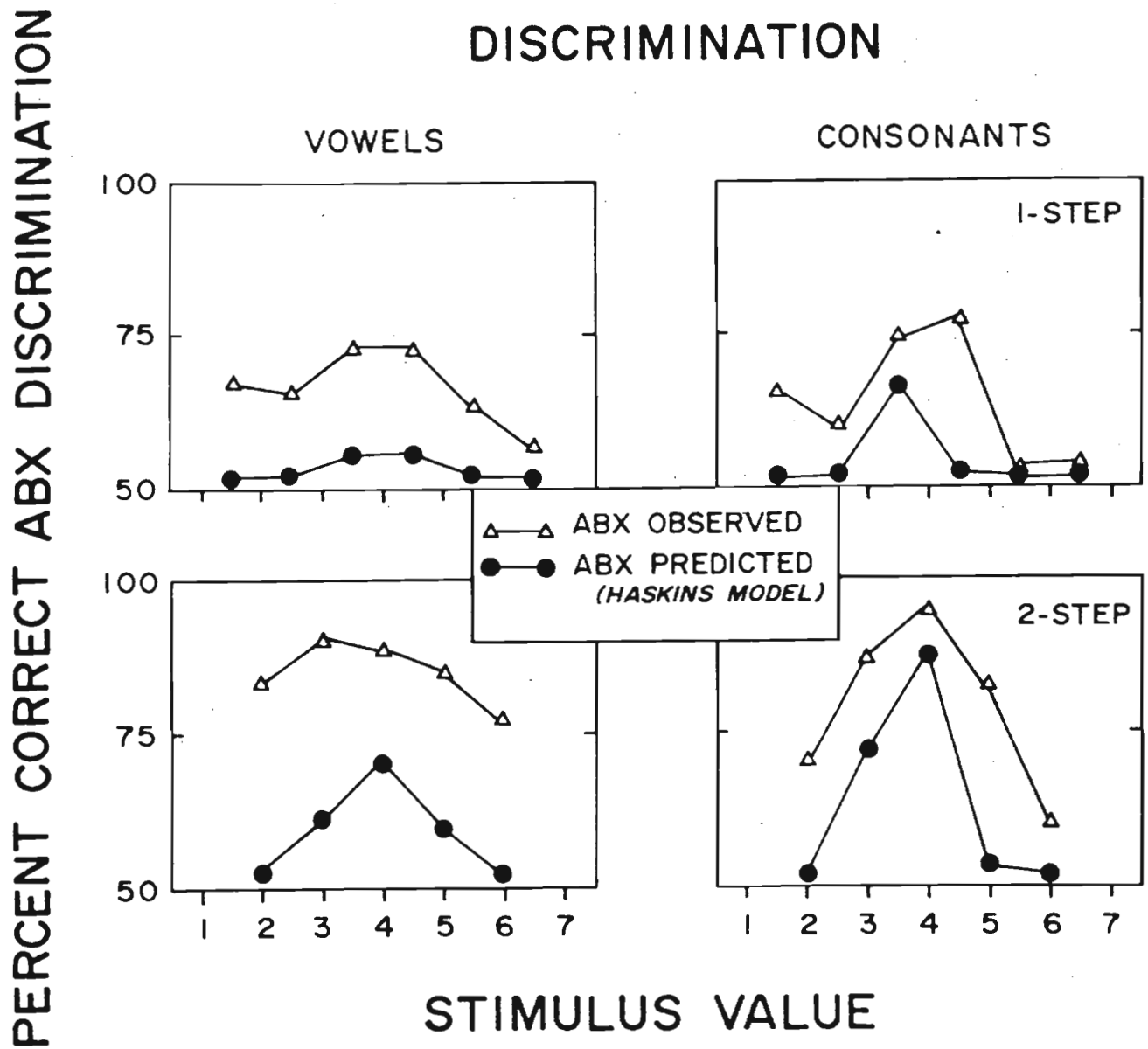

DISCRIMINATION
VOWELS
CONSONANTS
PERCENT CORRECT ABX DISCRIMINATION
100
75
50
1-STEP
2-STEP
ABX OBSERVED
ABX PREDICTED
(HASKINS MODEL)
1 2 3 4 5 6 7
STIMULUS VALUE

Figure 8.

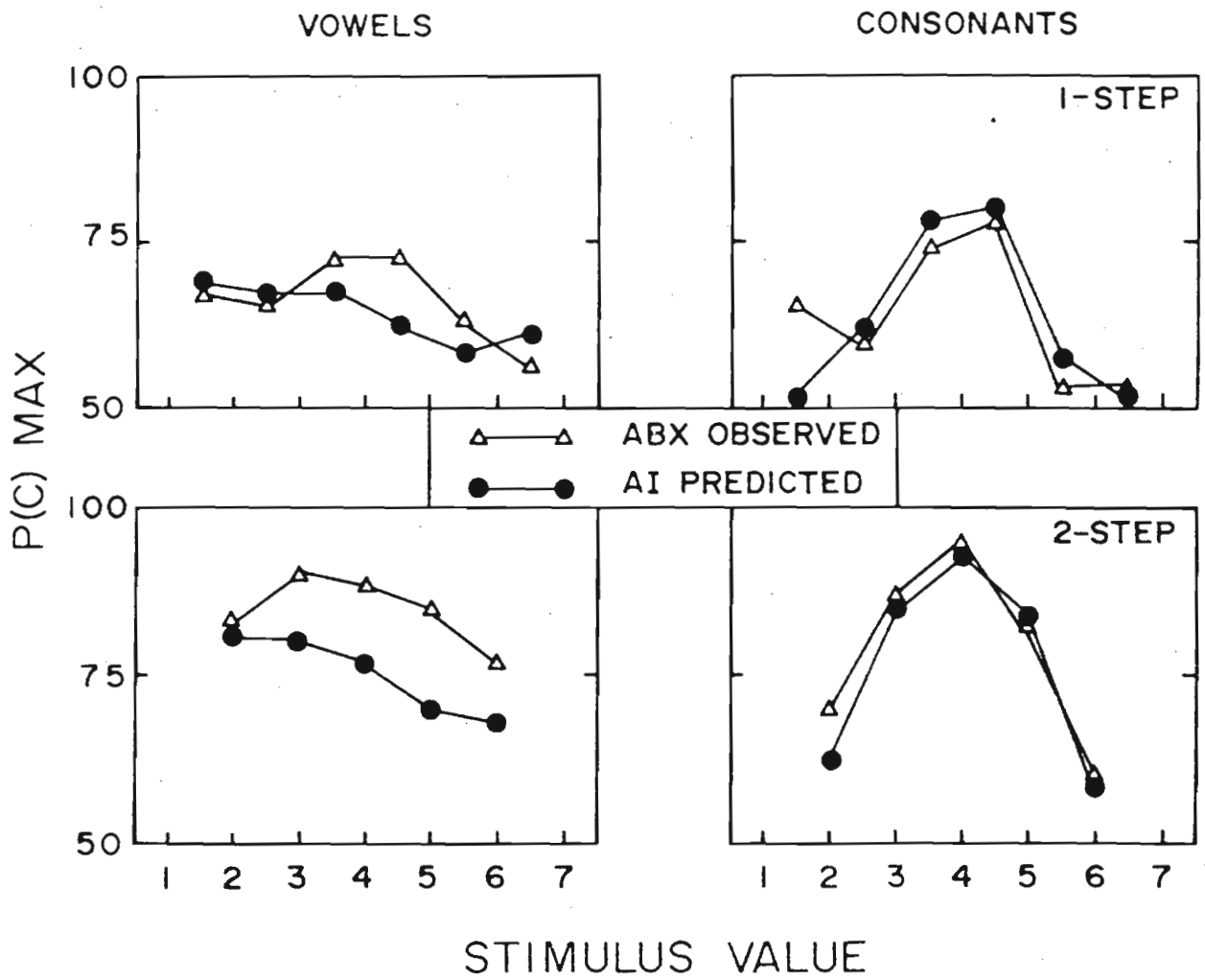

DISCRIMINATION
VOWELS
CONSONANTS
1-STEP
2-STEP
ABX OBSERVED
AI PREDICTED
P(C) MAX
100
75
50
100
75
50
1 2 3 4 5 6 7
1 2 3 4 5 6 7
STIMULUS VALUE

Figure 9.

Perceptual Analysis of Speech Sounds by Prelinguistic Infants: A First Report*

R. N. Aslin, A. J. Perey, B. Hennessy, and D. B. Pisoni

Department of Psychology

Indiana University

Bloomington, Indiana 47401

Abstract

In this paper we summarize the results of our efforts using a two-alternative go/no go head-turning paradigm to study speech perception abilities of 5-6 month old infants. Two synthetically produced speech sounds were presented as training stimuli during initial shaping and conditioning phases and the infant was reinforced with the presentation of a visual stimulus (i.e., an animated toy monkey) for an appropriate differential head-turn response toward the left for one stimulus (S_1). A second response involved inhibiting a headturn response by orienting toward the center when another stimulus (S_2) was presented. Implications of our findings and methodology for questions surrounding perceptual constancy, feature analysis and the role of early environmental experience in development of speech perception abilities in young infants are discussed.

*This paper was presented at the 94th meeting of the Acoustical Society of America, Miami Beach, Florida, December 16, 1977. This work was supported by NIH grant NS-12179-03 and NIMH grant MH-30424-01 to Indiana University.

Perceptual Analysis of Speech Sounds by Prelinguistic Infants: A First Report

R. N. Aslin, A. J. Perey, B. Hennessy, and D. B. Pisoni

Department of Psychology

Indiana University

Bloomington, Indiana 47401

For several years, researchers interested in the speech perception abilities of human infants have been limited to a single measure: discrimination. Being able to show discrimination of some speech sounds is, however, only one aspect of the speech perception process. To gain a better understanding of how infants perceive speech signals, it is necessary to secure identification or labeling functions as well as discrimination data. For example, all of the pairs of speech stimuli used in past infant studies contrasted on some particular sets of acoustic cues. Demonstrating that these acoustic differences can be discriminated by infants may tell us that discrimination was based on nonlinguistic variables such as the specific psychophysical properties of these sounds, but it does not tell us anything about perceptual categorization or whether infants perceive speech like mature adults. For the past year our group at Indiana has been attempting to devise a technique for use with infants which will provide an analog to the adult labeling procedure. Today, we wish to report on the progress of our efforts in this regard.

The three major responses that have been used in the past to measure infant's speech discrimination are heart rate (HR), high amplitude sucking (HAS), and operant head-turning (OHT). As illustrated in Figure 1, previous

---

Insert Figure 1 about here

---

uses of these measures have consisted of presenting a standard speech stimulus repeatedly until some criterion of habituation has been reached and then measuring a dishabituation response to the presentation of a novel stimulus. As reported in San Diego by Pat Kuhl, a variation on this procedure has been to present multiple-standards in the form of a category formation paradigm. If the infants show discrimination of only the criterial differences between these stimuli, it can be inferred that the stimuli in the multiple-background standard have been generalized to a single perceptual category. Thus, one could argue that some evidence for perceptual constancy has been obtained. However, a measure of identification or labeling requires that at least two independent responses be present--one response to stimulus S_1 and the other to stimulus S_2. This is shown schematically in the middle panel of Figure 1.

After the successful demonstration of a two-alternative identification, a generalization procedure can then be used to determine the relative similarity of novel stimuli to the original two discriminative stimuli as shown in the lower panel. Thus, for example, one could use /ba/ and /pa/ as training stimuli and /bo/, /bu/, /bi/, /bɛ/, and /po/, /pu/, /pi/, /pɛ/ as generalization stimuli. The outcome of this kind of experiment can determine if infants perceive all acoustically different versions of /b/ and /p/ to be members of two general consonant categories.

We chose to use an operant head-turning response as our dependent measure for two reasons. First, the head-turning procedure as used earlier by Eilers and Kuhl is subject to a relatively low attrition rate. Second, the identification-generalization procedure requires within-subject data, and in our view the head-turning procedure could provide us with reliable within-subject results.

Our first attempt at modifying head-turning procedures to obtain identification data involved presenting several repetitions of one of two possible speech stimuli, /ba/ or /ta/, and then reinforcing a directional head turn for a correct response, e.g., a headturn to the right for /ba/ or one to the left for /ta/. The complexity of this modification of the earlier procedure, from the infant's standpoint, is probably best exemplified by considering the earlier headturning procedure in detail. This procedure, which we have called "Go'No-Go," is illustrated in Figure 2.

Insert Figure 2 about here

A background habituating stimulus (S_B) is presented repeatedly while the infant is fixating straight ahead toward an experimenter who is attempting to capture and maintain the infant's attention. At variable intervals, two types of trials are presented: experimental trials and control trials. On experimental trials the background stimulus (S_B) is changed to a novel target stimulus (S_1). On control trials the background stimulus continues with no change. Appropriate head-turns on experimental trials are reinforced by the brief presentation of an animated toy. Some measure of performance is then used such as per cent correct or number of trials to a criterion to establish whether the infant has demonstrated discrimination of the target stimulus from the background.

Our initial procedure, which we have called two-alternative forced choice (2AFC) is also illustrated in Figure 2b in the bottom panel. One stimulus, S_1, is paired with a head-turn toward the left; the other stimulus, S_2, is paired with a head-turn toward the right. Directionally-appropriate

head-turns are reinforced by the brief presentation of one of the two animated toy monkeys. The apparatus is shown in Figure 3 in the bottom panel. During inter-trial intervals a series of blinking lights was

---

Insert Figure 3 about here

---

presented to return the infant's gaze to the center. Although we had great difficulty obtaining reliable identification performance from infants in a short period of time, we did have some success with this procedure. One set of results is shown in Figure 4. This infant, after 13 shaping

---

Insert Figure 4 about here

---

trials and 30 training trials maintained a level of responding above chance for over 30 additional trials. However, it was clear from working on this procedure for several months that the technique was too difficult for the average infant that we see in our laboratory and despite some successes, even these infants did not reach a level of performance that was high enough to allow generalization data to be gathered.

There are many possible reasons for the difficulty infants showed in this procedure involving discrete left and right headturns and it is useful to summarize these here for anyone contemplating research in this area. First, infants were required to display a differential response--right versus left headturns in the absence of any localization cues. Response differentiation appears to be poor for bilateral symmetrical motor movements in infants, a finding common in the animal research literature. In short,

associating /ba/ with the right and /ta/ with the left may simply be too difficult for young infants in the absence of additional cues and may be thought of as a possible "constraint on learning." Second, the arbitrary association or pairing of stimuli and responses, regardless of poor response differentiation, may also be responsible for our difficulties. This task demands that infants not only encode two stimulus-response pairings, but also that they remember the pairings for at least 15 seconds (the duration of two trials). A third possibility with this procedure may be that infants are distracted by the repeated shifts from the visual stimuli at center to the auditory stimuli. Finally, the Go/No-Go discrimination procedure used earlier takes advantage of an apparently "natural" tendency to orient toward a stimulus change in the infant's environment. That is, the sound source presenting both the background and the target is located to the right or left of the infant adjacent to the visual reinforcer. In our left-right procedure we initially shaped a left or right head-turn by presenting the stimuli from speakers located either to the left or right of center. However, to demonstrate identification, or a response to "stimulus quality," it was necessary to fade-out the localization cue and present both stimuli from a centrally located speaker where directional cues were absent. When we attempted this, on subjects who had been trained with localization cues, these subjects often failed to continue responding to the stimuli in the absence of localization cues.

To remedy several of these problems, we redesigned our apparatus to conform more closely to the successful Go/No Go procedure of Eilers and Kuhl. Figure 5 illustrates an outline of the current testing situation.

Insert Figure 5 about here

As shown, an experimenter attracts the gaze of the infant toward center while a stimulus such as a tone or neutral vowel is presented repeatedly in the background. The infant is seated on the mother's lap and she wears headphones during the entire session. A PDP-11/10 computer is used to present all stimuli, record responses and deliver visual reinforcement to the infant. Two raters who are blind to the stimuli and reinforcement contingencies record headturns directly into the computer which has been programmed to provide the visual reinforcement when both raters agree on the direction of a headturn. In contrast to the earlier procedure of Eilers and Kuhl, the experimenter who is located in the room with the infant does not have control over the stimuli or reinforcer.

The new procedure, which we call two-alternative go/no go (2AGNG) is outlined schematically in Figure 6. "Go" trials consist of presenting stimulus S_1 and reinforcing an appropriate headturn. "No-Go" trials

---

Insert Figure 6 about here

---

consist of presenting stimulus S_2 and not reinforcing the same headturn which was reinforced during S_1 trials. In essence, we are giving the infant feedback that only S_1 in particular is the discriminative stimulus for a headturn response whereas S_2 is the stimulus for inhibiting a response.

Figure 7 summarizes the performance of four infants tested recently on the 2AG/NG procedure. In all four cases the background stimulus was

---

Insert Figure 7 about here

---

the synthetic vowel /u/ which was 350 msec in duration and repeated once every second. Stimuli S_1 and S_2 were the synthetic vowels /a/ and /i/. It can be seen from the data displayed here that all infants rapidly learned to turn toward a change from the background stimulus to the reinforced target stimulus, S_1. Summed over all four infants, the level of appropriate headturns towards S_1 was 88%. Difficulty arises initially, however, when the infants are required to inhibit their headturning to a change from the background stimulus to the "No-Go" stimulus, S_2. Of course, the fact that they do turn their head indicates that they have discriminated the "No-Go" stimulus (S_2) from the background--a fact which is clearly necessary for this procedure to be a meaningful two-alternative situation. Despite the infant's difficulties in initially inhibiting their headturn on S_2 trials, the data from the four infants show that they can learn to respond correctly on No-Go trials and they can do this fairly rapidly. The four infants shown in Figure 7 performed at a 66% level averaged across S_1 and S_2 trials. Although this level of correct responding is not overwhelming, it should be noted that this value includes all trials on which the infant was initially acquiring the differential responses. Moreover, these results were obtained in only a single session, and for all of the four infants shown here this single session was their very first using the 2AG/NG procedure. Indeed, this was also their first visit to our lab.

We are confident from these results that by using several repeated testing sessions we can increase the overall probability of correct responding to at least an 80% level, a reasonable criterion in our view for subsequently assessing the infant's abilities to show generalization

to stimuli other than S_1 and S_2--the stimuli that they were originally trained on. It is here that we can begin to assess issues surrounding categorization and perceptual constancy.

Because all of our experiments are run "on-line" in real-time under computer control, we can present a very wide range of different signals for generalization testing which should enable us to answer a whole host of theoretically motivated questions concerning perceptual constancy, feature extraction, and the role of early experience in perceptual development. Finally, our research using this procedure, is ultimately aimed at claims surrounding the early predisposition or specialization for linguistic categorization--an issue that has received a great deal of attention in recent years because it may be a species-specific phenomena.

In final summary, we feel that discrimination paradigms such as the earlier Go/No-Go head-turning procedure do provide useful information about the ability of young infants to discriminate or detect small and perhaps very subtle differences between various kinds of speech signals, but additional information is needed about an infant's ability to classify or categorize acoustically different sounds that adults perceive as the same. To this end, our work on the two-alternative Go/No-Go paradigm should provide a way to collect identification or labeling data and will therefore permit us to make some gross comparisons of the perceptual capabilities of infants, adults and even chinchillas, comparisons which should shed light on the problems surrounding the existence of species-specific mechanisms for processing biologically significant acoustic signals like speech.

Aslin, Perey, Hennessy & Pisoni

Indiana University

Figure 1

I. Discrimination Paradigm

Standard Stimulus (Background) → Novel Stimulus → Response

Standard Stimulus (Background) → Standard Stimulus (Control) → No Response

II. Identification Paradigm

Stimulus S_1 → Response R_1

Stimulus S_2 → Response R_2

III. Generalization (Transfer) Phase

Stimulus S_1 → Response R_1

Stimulus S_2 → Response R_2

Stimulus S_3 → Response R_1 or R_2?

Stimulus S_4 → Response R_1 or R_2?

Stimulus S_5 → Response R_1 or R_2?

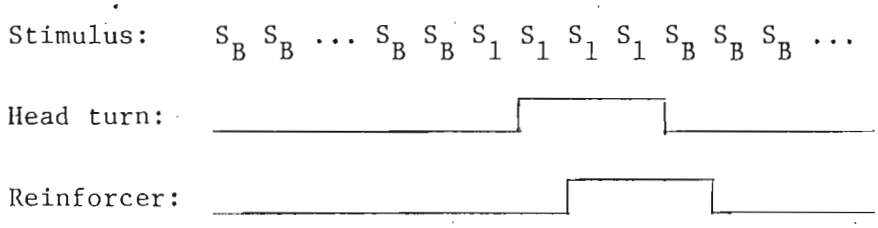
.
.
.

Stimulus S_n → Response R_1 or R_2?

Figure 2(a)

GO/NO GO Procedure

I. Experimental trial



II. Control trial

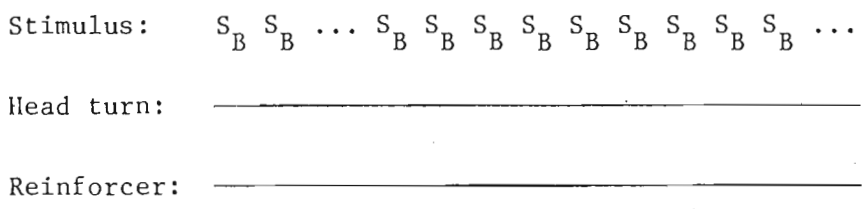
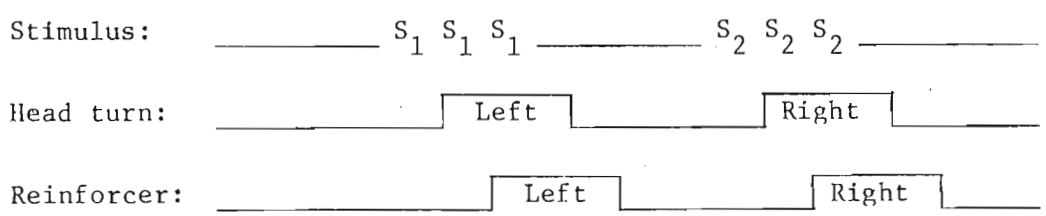


Figure 2(b)

Two-alternative Forced-choice Procedure

I. Shaping and Training



II. Transfer

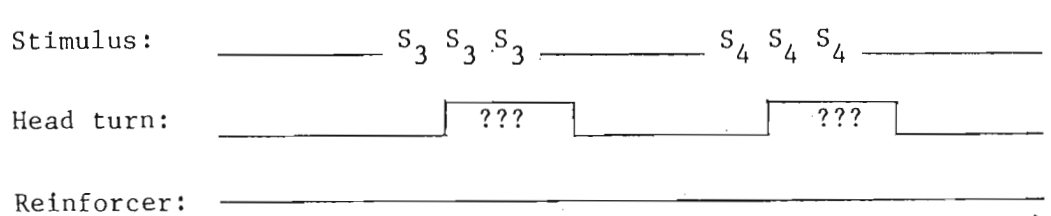
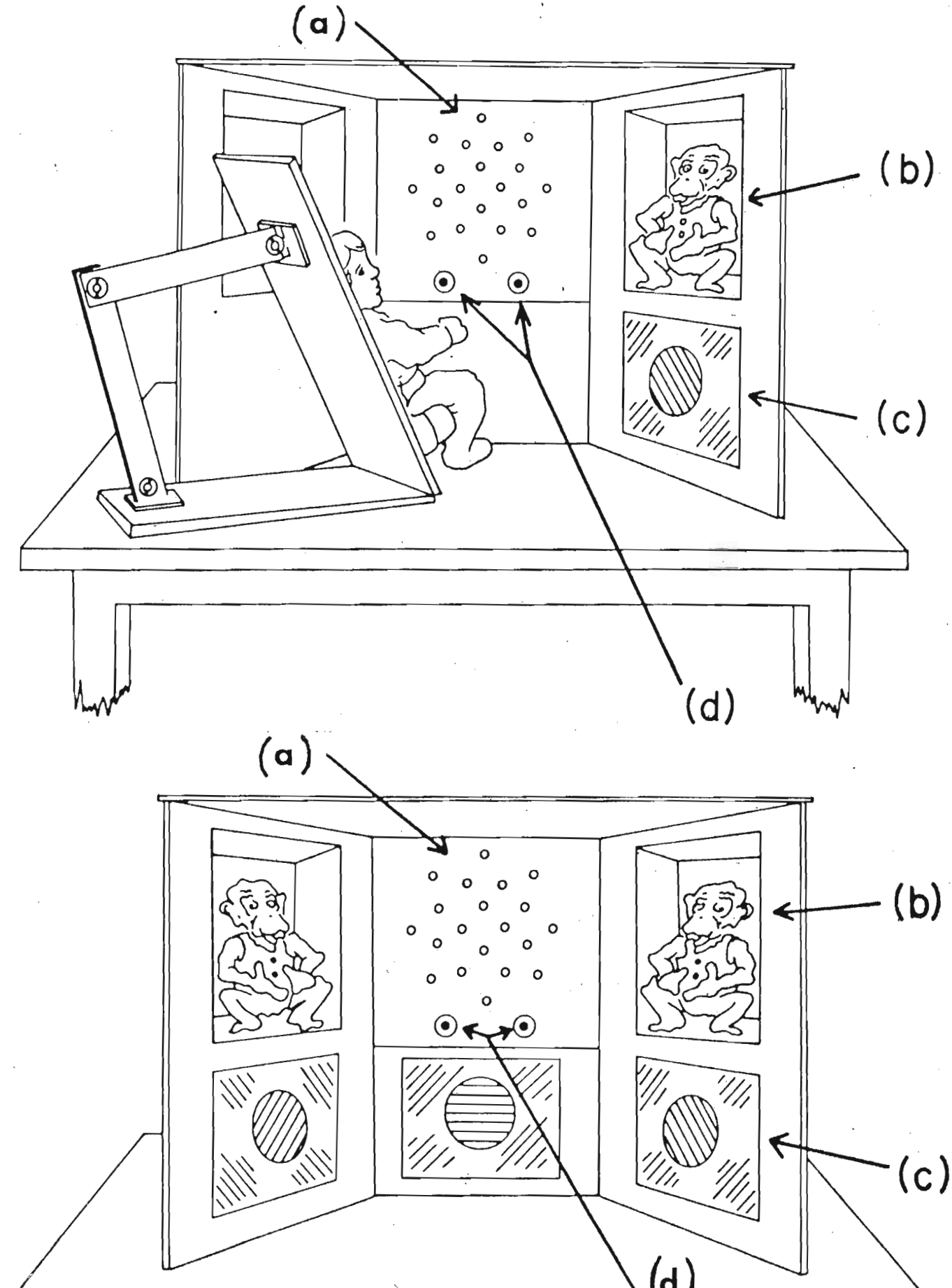


Figure 3



Aslin, Perey, Hennessy & Pisoni

Indiana University

Figure 4

Subject: RP

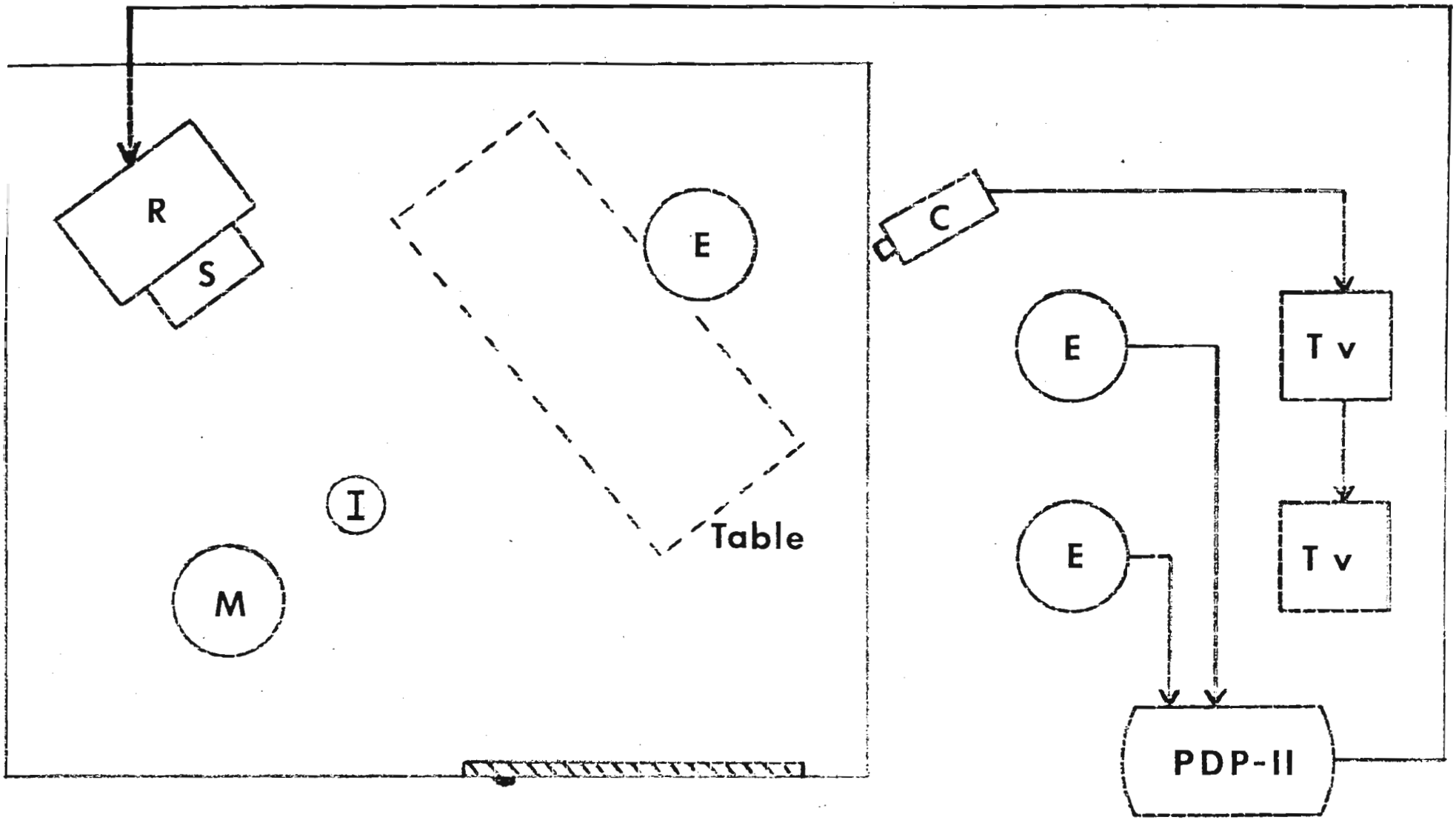
Age: 5½ months

S_1 = /ba/ = LEFT

S_2 = /ta/ = RIGHT

Trials	S_1	S_2	Total
1-13	Shaping		
14-23	0/0	2/10	20%
24-33	2/2	2/8	40%
34-43	1/3	2/7	30%
44-53	3/5	4/5	70%
54-63	4/5	3/5	70%
64-73	5/8	2/2	70%
	67%	75%	

Figure 5



Two-alternative Go/No Go apparatus. R = reinforcer, S = speaker, E = experimenter, M = mother, I = infant, C = video camera.

Aslin, Perey, Hennessy & Pisoni

Indiana University

Figure 6

Two-alternative Go/No Go Procedure

I. Identification

A. "GO" Trial

Stimulus: $S_B\ S_B \ldots S_B\ S_B\ S_1\ S_1\ S_1\ S_1\ S_B\ S_B\ S_B \ldots$ P(HIT)

Head turn: 

Reinforcer: 

B. "NO GO" Trial

Stimulus: $S_B\ S_B \ldots S_B\ S_B\ S_2\ S_2\ S_2\ S_2\ S_B\ S_B\ S_B \ldots$ P(False alarm)

Head turn:

Reinforcer:

II. Generalization

A. Same as (A) and (B) above

B. Stimulus: $S_B\ S_B \ldots S_B\ S_B\ S_3\ S_3\ S_3\ S_3\ S_B\ S_B\ S_B \ldots$

Stimulus: $S_B\ S_B \ldots S_B\ S_B\ S_4\ S_4\ S_4\ S_4\ S_B\ S_B\ S_B \ldots$

Stimulus: $S_B\ S_B \ldots S_B\ S_B\ S_5\ S_5\ S_5\ S_5\ S_B\ S_B\ S_B \ldots$

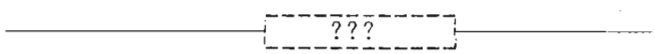
Headturn: 

Figure 7

Subject	Trials	S_1=GO	S_2=NO GO	TOTAL
RF (6½ months)	1-4	shaping		
S_B = /u/	5-9	3/3	0/2	60%
S_1 = /i/	10-19	3/3	3/7	60%
S_2 = /a/	20-29	5/5	3/5	80%
JH (9 months)	1-5	shaping		
S_B = /u/	6-15	6/6	1/4	70%
S_1 = /i/	16-25	5/5	2/5	70%
S_2 = /a/	26-35	4/5	2/5	60%
	36-45	4/5	2/5	60%
	46-52	4/4	2/3	85%
MB (6 months)	1-18	shaping		
S_B = /u/	19-28	6/6	1/4	70%
S_1 = /a/	29-38	4/5	3/5	70%
S_2 = /i/	39-48	4/5	2/5	60%
CM (8 months)	1-10	shaping		
S_B = /u/	11-20	5/5	3/5	80%
S_1 = /i/	21-30	4/7	3/3	70%
S_2 = /a/	31-40	5/7	1/3	60%
	41-50	4/5	3/5	70%
	51-60	4/4	1/6	50%

INSTRUMENTATION AND SOFTWARE DEVELOPMENT

KLTEXC: Executive Program to Implement the KLATT Software Speech Synthesizer

Diane Kewley-Port

KLTEXC is a FORTRAN program designed to interface with the flexible digital speech synthesizer algorithm designed by Dennis Klatt. The goal was to make KLTEXC a human oriented research tool for investigating various properties of speech. This report gives a brief summary of both the synthesizer and KLTEXC and describes in some detail the design of executive program.

Introduction

The KLATT software digital speech synthesizer was designed by Dennis Klatt (1977) to be a very flexible and natural synthesizer modeled after many principles of the acoustic theory of speech production (Fant, 1960). The KLTEXC program was designed to interface with KLATT so that the flexibility of the synthesizer would be retained to the greatest extent while making the manipulation of the synthesis parameters as easy as possible for the user. In creating the program, two goals were kept in mind. The first was that KLTEXC would be used to produce sets of speech utterances which would be as natural as possible, but differ parametrically from one another along several stimulus dimensions. Thus KLTEXC allows the synthesis parameters to be stored and retrieved from disk as files which can then be conveniently altered in step sizes as small as the synthesizer itself allows. For KLATT these limits are essentially in 1 dB of amplitude, 1 Hz in frequency, and .5 msec in time. Secondly, we also wanted users to be able to produce nonspeech signals on KLATT as control stimuli for psychoacoustic experiments comparing speech and speech-like stimuli. To this end, we disabled the software that constrains parameter values to be within "natural" limits of what the human vocal tract can produce, and an extra parameter (COR SW) was added to KLATT to permit more direct control over the synthesis in certain conditions.

KLTEXC is currently implemented on the PDP 11/05 and PDP 11/34 computers in the Speech Perception Laboratory, both under the RT-11 operating system with two RK05 disks and 28 K of memory. KLTEXC and KLATT are both written in FORTRAN IV. KLTEXC was first operational in August, 1977, with an earlier version of the KLATT synthesizer algorithm. This paper will describe the version of KLTEXC implemented as of October, 1978 with the KLATT synthesizer subroutine obtained in 1978. Differences between these programs and the version described in Klatt, 1979 are slight.

In order to facilitate understanding of KLTEXC, an extremely brief description of KLATT will follow. The interested reader is urged to consult Klatt (1979) for the theory and details necessary to understand and use the KLATT synthesis algorithm.

KLATT Synthesizer

Figure 1 provides a block diagram of KLATT and Table 1 lists more specifically the individual control parameters. A total of six formant resonators, R1 to R6, can be employed in either a cascade or parellel configuration. In the cascade branch there is a pair of nasal pole and zero resonators, RNP and RNZ, and in the parallel branch a nasal resonator RNP. All resonators are specified by two parameters, their center frequency, prefix F, and bandwidth, prefix B. Furthermore, the amplitude of each resonator in the parallel branch is controlled in dB with the A_i parameters, A1 to A6. The resonators can be excited by two voicing sources and a noise source which are controlled by the amplitude parameters AV, AVS, AH and AF. AV controls a "normal" voicing source, while AVS controls a "smoothed" voicing source such as might be found in the prevoicing or closure of stop consonants. In the default configuration of the synthesizer as Klatt designed it, the voicing sources excite only the cascade branch. The frication source (AF) can

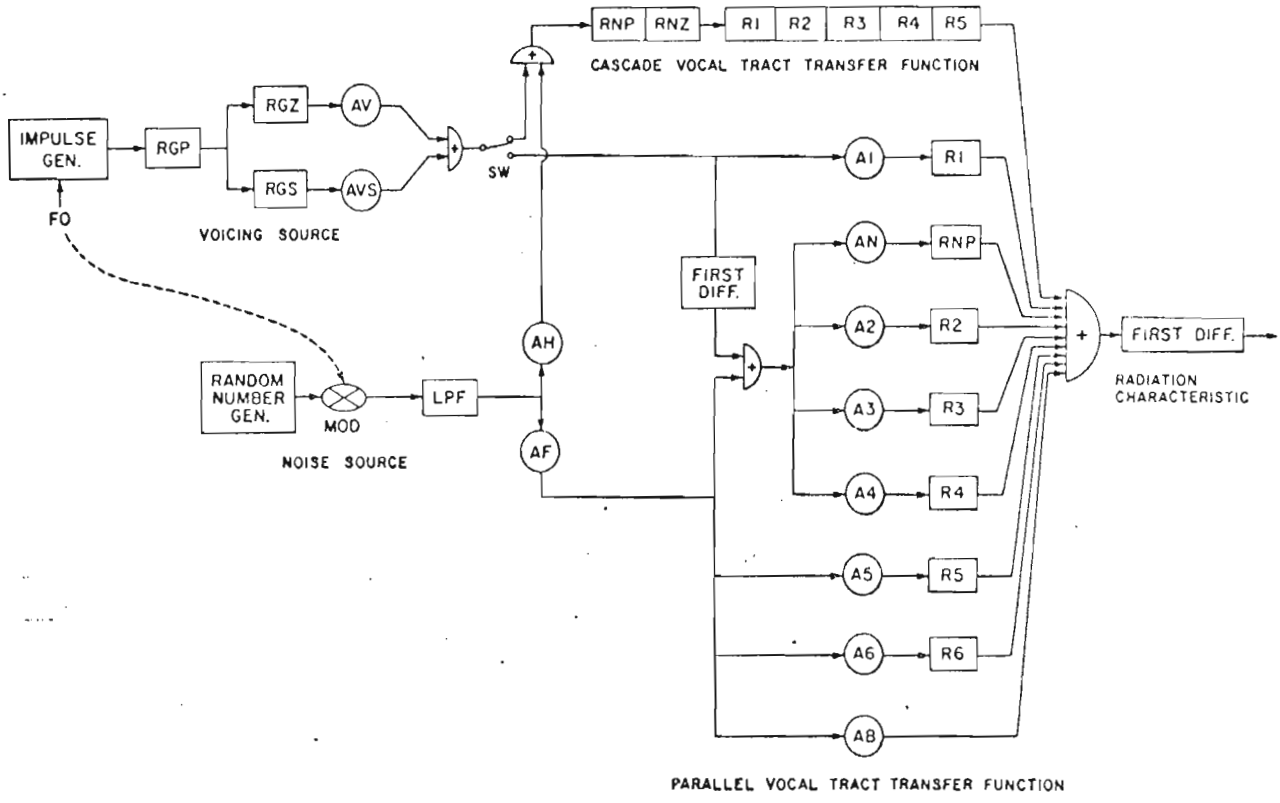


Figure 1. Block diagram of Klatt's cascade/parallel formant synthesizer. Digital resonators are indicated by the prefix R and amplitude controls by the prefix A. Each resonator has an associated resonant frequency control parameter and a resonance bandwidth control parameter. (Taken from Klatt, 1977.)

Table 1: List of control parameters for the software formant synthesizer. The second column indicates whether the parameter is normally constant (C) or variable (V) during the synthesis of English sentences. Also listed are the permitted range of values for each parameter, and a typical constant value. (Taken from Klatt, 1977.)

N	V/C	SYM	NAME	MIN	MAX	TYP
1	V	AV	AMPLITUDE OF VOICING (DB)	0	80	0
2	V	AF	AMPLITUDE OF FRICATION (DB)	0	80	0
3	V	AH	AMPLITUDE OF ASPIRATION (DB)	0	80	0
4	V	AVS	AMPLITUDE OF SINUSIODAL VOICING (DB)	0	80	0
5	V	F0	FUNDAMENTAL FREQ. OF VOICING (HZ)	0	500	0
6	V	F1	FIRST FORMANT FREQUENCY (HZ)	150	900	500
7	V	F2	SECOND FORMANT FREQUENCY (HZ)	500	2500	1500
8	V	F3	THIRD FORMANT FREQUENCY (HZ)	1300	3500	2500
9	V	F4	FOURTH FORMANT FREQUENCY (HZ)	2500	4500	3300
10	V	FNZ	NASAL ZERO FREQUENCY (HZ)	200	700	250
11	V	AN	NASAL FORMANT AMPLITUDE (DB)	0	80	0
12	V	A1	FIRST FORMANT AMPLITUDE (DB)	0	80	0
13	V	A2	SECOND FORMANT AMPLITUDE (DB)	0	80	0
14	V	A3	THIRD FORMANT AMPLITUDE (DB)	0	80	0
15	V	A4	FOURTH FORMANT AMPLITUDE (DB)	0	80	0
16	V	A5	FIFTH FORMANT AMPLITUDE (DB)	0	80	0
17	V	A6	SIXTH FORMANT AMPLITUDE (DB)	0	80	0
18	V	AB	BYPASS PATH AMPLITUDE (DB)	0	80	0
19	V	B1	FIRST FORMANT BANDWIDTH (HZ)	40	500	50
20	V	B2	SECOND FORMANT BANDWIDTH (HZ)	40	500	70
21	V	B3	THIRD FORMANT BANDWIDTH (HZ)	40	500	110
22	C	SW	CASCADE/PARALEL SWITCH	0(CASC)	1(PARA)	0
23	C	FGP	GLOTTAL RESONATOR 1 FREQUENCY (HZ)	0	500	0
24	C	BGP	GLOTTAL RESONATOR 1 BANDWIDTH (HZ)	100	2000	100
25	C	FGZ	GLOTTAL ZERO FREQUENCY (HZ)	0	5000	1500
26	C	BGZ	GLOTTAL ZERO BANDWIDTH (HZ)	100	9000	6000
27	C	B4	FOURTH FORMANT BANDWIDTH (HZ)	100	500	250
28	V	F5	FIFTH FORMANT FREQUENCY (HZ)	3500	4900	3850
29	C	B5	FIFTH FORMANT BANDWIDTH (HZ)	150	700	200
30	C	F6	SIXTH FORMANT FREQUENCY (HZ)	4000	4999	4900
31	C	B6	SIXTH FORMANT BANDWIDTH (HZ)	200	2000	1000
32	C	FNP	NASAL POLE FREQUENCY (HZ)	200	500	250
33	C	BNP	NASAL POLE BANDWIDTH (HZ)	50	500	100
34	C	BNZ	NASAL ZERO BANDWIDTH (HZ)	50	500	100
35	C	BGS	GLOTTAL RESONATOR 2 BANDWIDTH	100	1000	200
36	C	SR	SAMPLING RATE	5000	20000	10000
37	C	NWS	NUMBER OF WAVEFORM SAMPLES PER CHUNK	1	200	50
38	C	GO	OVERALL GAIN CONTROL (DB)	0	80	48
39	C	NFC	NUMBER OF CASCADED FORMANTS	4	6	5

excite either branch. AH controls the amount of frication sent to the cascade branch, in which case it can be considered phonetically as "aspiration." AF controls frication sent through the parallel resonators A2 to A6 and AB which may then be used to synthesize stop bursts and fricatives where the source of turbulence is located above the glottis.

Of the remaining parameters shown in Table 1 only NWS, number of waveform samples per array of input parameters, deserves further comment. KLATT is, of course, a FORTRAN subroutine. It synthesizes a waveform when called from the main program which must supply the 39 parameters in Table 1 to KLATT. KLATT then calculates a waveform piece which will contain NWS output samples and place them in an output array. KLATT has a memory, so that parameters changing between calls to KLATT are linearly interpolated over the output sample interval. Thus to change synthesis parameter values in 5 ms steps, the default value, NWS = (SR/1000 ms)*5 ms = 50, when SR = sample rate = 10000 samples/sec. Since NWS is a variable, very careful control over the synthesis is permitted.

KLTEXC Parameter Files

A decision was made to divide the parameters in Table 1 into a set of 20 variable parameters, and a set of 20 fixed parameters located in the global table. (The extra parameter COR SW, will be discussed below.) An utterance is synthesized from a parameter file consisting of a buffer of the 20 variable parameters, n lines long and a single global table. Each of the n lines is referenced in 5 ms intervals, although during synthesis the value of the NWS parameter will actually control how many waveform samples per line are synthesized. Table 2 contains the names of the variable parameters and global parameters as defined in KLTEXC, associating them with the numbers found in Klatt's original description shown in Table 1. Note that variable names have been shortened to two characters when necessary. The

TABLE 2

KLTEXC VARIABLE TABLE

20 VARIABLE PARAMETERS

TABLE 1 No.	1	2	3	4	5	6	7	8	10	11	12	13	14	15	16	17	18	19	20	21
2 CHAR. NAME	AV	AF	AH	AS	FØ	F1	F2	F3	NZ	AN	A1	A2	A3	A4	A5	A6	AB	B1	B2	B3
RANGE	Ø 8Ø	Ø 8Ø	Ø 8Ø	Ø 8Ø	Ø 4ØØ	2ØØ 9ØØ	6ØØ 24ØØ	13ØØ 31ØØ	25Ø 7ØØ	Ø 8Ø	Ø 8Ø	Ø 8Ø	Ø 8Ø	Ø 8Ø	Ø 8Ø	Ø 8Ø	Ø 8Ø	4Ø 5ØØ	4Ø 5ØØ	4Ø 5ØØ
DEFAULT VALUE	6Ø	Ø	Ø	Ø	12Ø	68Ø	189Ø	265Ø	25Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø	Ø	9Ø	11Ø	17Ø

20 GLOBAL PARAMETERS

TABLE 1 NO.	9	22	23	24	25	26	27	28	29	30	31	32
NAME	F4	C/P SW	F GLT RES	B GLT RES	F GLT ZRO	B GLT ZRO	B4	F5	B5	F6	B6	F NSL POL
DEFAULT VALUE	33ØØ	Ø	Ø	1ØØ	15ØØ	6ØØØ	25Ø	385Ø	2ØØ	49ØØ	1ØØØ	25Ø

TABLE 1 NO.	33	34		36	37	38	35		39	40		
NAME	B NSL POL	B NSL ZRO	NOT USED	SAMP RATE	SAMP/ LINE	GO	B GLT RES 2	NOT USED	NO. CAS FOR	CORSW	NOT USED	
DEFAULT VALUE	1ØØ	1ØØ		1ØØØØ	5Ø	48	2ØØ		5	1		

task of KLTEXC, then, is to allow a user to efficiently create and alter a parameter file to be used by KLATT to synthesize a speech waveform. An example of a parameter file can be seen in Table 3, a copy of the printer listing of the utterance /dIg/ from KLTEXC.

As noted earlier, KLTEXC contains one extra parameter, COR SW, not found in the original KLATT subroutines. The purpose of this parameter was to suppress certain rules implemented automatically during synthesis in the parallel configuration to calculate formant amplitudes. Klatt included these rules to make voiced sounds produced in the parallel configuration more natural, like those produced in the normal cascade synthesis. Since one purpose of KLTEXC was to allow for synthesis of non-speech sounds, COR SW can be set to suppress these rules. The rules are described by KLATT (1979) in greater detail.

One of the major design features of KLTEXC was to place no effective limits on the length of the parameter buffer. To do this, it was decided that the parameter buffer would be split into 1 second pages (200 lines), so that 1 second would always reside in core, and the rest of the buffer would be stored in a scratch file on the systems disk. The maximum length of a parameter file is currently set to 163 seconds or 2.7 minutes. The file structure of the buffer is opaque to the user and places no restriction on use of the KLTEXC commands.

KLTEXC Commands

The commands for creating parameter and waveform files, and related editing operations are specified by single letter mnemonics as they appear in the MENU shown in Table 4 below. Every effort was made to orient the program to naive users, to detect errors and to display messages for appropriate corrective action A brief description of the commands follows, starting with those which manipulate parameter files.

Table 3

DIH4.KPR, PARAMETER FILE FOR /DIG/ DATE: 12-JUL-79 SIGNAL MAX =

GLOBAL PARAMETERS

F4	C/P SW	F GLT RES	B GLT RES	F GLT ZRO	B GLT ZRO	B4	F5	B5	F6	B6	F NSL POL
3300	0	0	100	1500	6000	250	3850	200	4900	1000	250

B NSL POL	B NSL ZRO	F RAD ZRO	SAMP RATE	SAMP/LINE	SCAL AMP	BGLT RES2	NOT USED	NO. CAS FOR	COR SW	NOT USED
100	100	99	10000	50	48	200	0	5	1	0

MSEC	AV	AF	AH	AS	F0	F1	F2	F3	NZ	AN	A1	A2	A3	A4	A5	A6	AB	B1	B2	B3
0.	0	53	0	0	120	229	1634	2598	250	0	0	0	47	60	62	0	0	60	100	170
5.	0	45	0	0	120	257	1669	2596	250	0	0	0	47	60	62	0	0	60	100	170
10.	0	0	0	0	120	286	1703	2593	250	0	0	0	47	60	62	0	0	60	100	170
15.	60	0	0	0	105	314	1737	2591	250	0	0	0	47	60	62	0	0	60	100	170
20.	60	0	0	0	108	326	1746	2589	250	0	0	0	47	60	62	0	0	59	100	167
25.	60	0	0	0	110	339	1755	2586	250	0	0	0	47	60	62	0	0	57	100	164
30.	60	0	0	0	113	351	1764	2584	250	0	0	0	47	60	62	0	0	56	100	161
35.	60	0	0	0	115	363	1773	2582	250	0	0	0	47	60	62	0	0	54	100	158
40.	60	0	0	0	118	375	1782	2579	250	0	0	0	0	0	0	0	0	53	100	155
45.	60	0	0	0	120	388	1791	2577	250	0	0	0	0	0	0	0	0	51	100	152
50.	60	0	0	0	120	400	1800	2575	250	0	0	0	0	0	0	0	0	50	100	149
55.	60	0	0	0	120	400	1800	2572	250	0	0	0	0	0	0	0	0	50	100	146
60.	60	0	0	0	120	400	1800	2570	250	0	0	0	0	0	0	0	0	50	100	143
65.	60	0	0	0	120	400	1800	2570	250	0	0	0	0	0	0	0	0	50	100	140
70.	60	0	0	0	120	400	1800	2570	250	0	0	0	0	0	0	0	0	50	100	140
75.	60	0	0	0	119	400	1800	2570	250	0	0	0	0	0	0	0	0	50	100	140
80.	60	0	0	0	119	400	1800	2570	250	0	0	0	0	0	0	0	0	50	100	140
85.	60	0	0	0	119	400	1800	2544	250	0	0	0	0	0	0	0	0	50	100	140
90.	60	0	0	0	119	400	1800	2517	250	0	0	0	0	0	0	0	0	50	100	140
95.	60	0	0	0	119	400	1800	2491	250	0	0	0	0	0	0	0	0	50	100	140
100.	60	0	0	0	119	400	1800	2464	250	0	0	0	0	0	0	0	0	50	100	140
105.	60	0	0	0	118	400	1800	2438	250	0	0	0	0	0	0	0	0	50	100	140
110.	60	0	0	0	118	398	1822	2411	250	0	0	0	0	0	0	0	0	50	100	140
115.	60	0	0	0	118	396	1844	2385	250	0	0	0	0	0	0	0	0	50	100	140
120.	60	0	0	0	117	393	1867	2359	250	0	0	0	0	0	0	0	0	50	100	140
125.	60	0	0	0	116	391	1889	2332	250	0	0	0	0	0	0	0	0	50	100	140
130.	60	0	0	0	115	389	1911	2306	250	0	0	0	0	0	0	0	0	50	100	140
135.	60	0	0	0	114	387	1933	2279	250	0	0	0	0	0	0	0	0	50	100	140
140.	60	0	0	0	113	384	1956	2253	250	0	0	0	0	0	0	0	0	50	100	140
145.	45	0	0	0	113	382	1978	2226	250	0	0	0	0	0	0	0	0	50	100	140
150.	0	0	0	0	112	380	2000	2200	250	0	0	0	0	0	0	0	0	50	100	140

Table 4

MENU

(October, 1978 Revision)

A = Append two parameter files together

B = Define new parameter buffer

C = Change audio output parameters

D = Deposit parameter files on disk

E = Erase lines in parameter buffer

F = Fetch parameter file from disk

G = Get waveform file from disk

I = Insert a line in parameter buffer

K = Change entire column of parameter values

L = List parameter file on line printer

M = MENU

O = Output waveform file to D/A

P = Parameter values inserted individually

Q = Query last line in parameter file

R = Repeat a line in parameter buffer n times

S = Synthesize parameter file and store waveform file on disk

T = Interpolate linearly between two parameter values

V = View parameter buffer on CRT terminal

X = Change global table parameter values

Z = Set global table parameters to their default values

A parameter file is created in core using either the commands B or F. B (Buffer) creates a file containing steady-state parameter values for an /ae/-like vowel. F (Fetch) retrieves a parameter file from disk which was previously stored using D (Deposit). The parameter buffer can be altered in several ways. Entire lines can be inserted, I, erased, E or repeated, R. Individual parameter columns, like F1, can be altered using K, P and T as defined in the MENU. Any global table parameter can be altered using X. Two parameter files can be joined together as a new file on disk using A (Append). The contents of the parameter files can be listed on the CRT using V (View) or the line printer using L (List).

When the parameter file has been specified, a waveform file is synthesized by KLATT and stored on disk using action S (Synthesize). The waveform can then be output through the D/A for listening using command O (Output) according to the attenuation, interstimulus interval and number of repetitions specified by C. G (Get) retrieves a waveform file from disk for output by O.

File Names in KLTEXC

Although KLTEXC operates under the RT-11 monitor, standard file names are not used so that a number of special features for file manipulation could be implemented. All files are referenced only by a 6 character file name, and KLTEXC adds the appropriate extension, '.KWV' for waveform files and '.KPR' for parameter files. The storage device is always 'DK:'. One of the principle reasons for using only 6 character file names was to allow lists of names to be constructed and used in two commands, S - synthesize and O - output. One 20 item list can be entered or reused by either command. For synthesis, the list is used to synthesize waveform files for all parameter files named. For output, the list can be used to output sequentially the waveform files specified, allowing

the user to do audio comparisons or recordings of the utterances. Other programs are available to the user for preparing audio tapes or using stimuli generated by KLTEXC in real-time on-line applications.

Summary

KLTEXC has been used in our laboratory for about two years now. The goals of having a user oriented program to produce both "natural" speech and comparable nonspeech stimuli utilizing the flexibility of the KLATT synthesizer have been successfully achieved in a wide range of applications. The PDP-11 version of the KLTEXC program has been distributed to a number of other research laboratories in the United States and abroad and seems to be well accepted by other investigators having interests similar to ours.

Acknowledgement

We gratefully acknowledge the time and effort Dennis Klatt has contributed in making KLATT available to us, providing suggestions for KLTEXC and in consulting with us numerous times on the implementation and use of the synthesizer. The development of KLTEXC was supported by NIH grant NS12179 to Indiana University.

References

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

Klatt, D. H. A cascade-parallel terminal analog speech synthesizer and a strategy for consonant-vowel synthesis, Journal of the Acoustical Society of America, 1977, 61, Suppl. 1, 68(A).

Klatt, D. H. Software for a Cascade/Parallel Speech Synthesizer, Journal of the Acoustical Society of America, 1979 (In press).

Graphic Support for KLTEXC

Thomas Carrell and Diane Kewley-Port

KLTEXC, as described in Kewley-Port (1978) optionally supports two graphics input and output devices; the Summagraphics 2000 tablet and the DEC VT-11 graphics display processor. The tablet is used for parameter input to the program while the display processor is used for both parameter and waveform output from the program. The present paper describes the design and operation of these two optional features and provides a brief description of the hardware itself. Both options are available to users of KLTEXC.

Introduction

The process of generating high-quality synthetic speech can be greatly facilitated by visual feedback of either the input parameter file or the synthesized waveform. Another desirable feature is the ability to transfer information from spectrograms of natural speech directly into the parameter files used in KLTEXC. Using a refresh CRT and a graphic input tablet, several of these functions have now been implemented in a version of KLTEXC running in the Speech Perception Laboratory at Indiana University. The parameter file entry and display functions are available through a new command J (Jot routine) within the KLTEXC Menu. Display of waveform files will be added to KLTEXC as command W, although a separate program called DRAWAV currently supports various waveform displays on the system. This report contains a description of the hardware and the Jot and DRAWAV routines which provide the graphic support for KLTEXC.

Hardware

It is now possible to interact graphically with KLTEXC through the use of two peripherals, the Summagraphics 2000 Tablet/Digitizer and the DEC VT-11 Graphic Display Processor. They have been configured into the

RT-11 Version 3 operating system with software written here at Indiana University. KLTEXC (and any other FORTRAN program for that matter) uses FORTRAN calls to communicate with these two peripherals through the use of this custom software.

The Summagraphics 2000 is a 60 cm by 60 cm graphics input device which digitizes x-y coordinates for further processing. The user inputs data by moving a stylus across its drawing surface. Its resolution of .1 mm (100 points per cm) is more than adequate for tracing spectrograms. The data input rate to KLTEXC is limited by the speed of the CPU. With the current implementation on a PDP 11/05 the tracings must be made at a rather slow speed of about 4 cm per second or less for perfect tracking. However, the software linearly interpolates between the points received if the speed is faster than this and therefore a faster useable speed is realized.

The VT-11 is a graphic display processor. It executes graphic display code stored in the system main memory to draw vectors which are displayed on the screen. The graphic display code is executed repeatedly to refresh the display. A considerable amount of FORTRAN software is devoted to converting the x-y and simple vector data needed from KLTEXC to the graphic display code format. The resulting graphs and pictures are high contrast and have a resolution of 1024 by 1024 points on a 17 inch (40 cm diagonal) CRT screen. The data output from KLTEXC is limited to a maximum of 5000 words of graphic display code. This is adequate for the display of most parameter, waveform, amplitude, and spectral information generated by KLTEXC. A Polaroid CU-5 camera with a custom made CRT hood can be used to take high-contrast white on black

photographs of the CRT displays. All figures in this document were photographed originally using the Polaroid. We expect delivery of a hardcopy device within the next few months.

Graphic Parameter Entry and Display

The synthesizer routine in KLTEXC operates on control parameter data which is updated every 5 msec. The KLTEXC program as described in Kewley-Port (1978), permits two basic methods by which parameter data can be entered. The first is to specify parameter values at two or more times and then to linearly interpolate between them to obtain a value at each 5 msec update (T command). The second and more time consuming method is to type in a separate value for each 5 msec of an utterance for the parameter being entered (P command). Of course both methods are usually used together by an experimenter synthesizing speech.

With the capabilities afforded by the Summagraphics 2000 and the VT-11 a new method of parameter entry becomes possible. Parameter values may be input by simply tracing them on the tablet. Time is represented on the x axis and the parameter values are represented on the y axis. KLTEXC currently takes advantage of this form of input for the control parameters which specify the first three formant frequencies. This capability was designed with the specific intent of tracing formants and formant motions directly from spectrograms of natural speech. However, it is also useful to rapidly input idealized stimuli which have been schematized on paper. Graphic input, then, is a practical method of avoiding the tasks of measuring spectrograms at 5 msec intervals and of inputting data on a point-by-point basis.

KLTEXC uses the VT-11 display processor to output formant parameter information in a simple spectrographic representation (see Figures 1 and 2 below). The formant tracks are displayed simultaneously with the values being traced on the tablet. Furthermore, the formant frequencies already stored in a parameter file may be viewed by a user. Another feature permits changing formants of old parameter files with the tablet. Thus, keyboard and graphic input may be mixed to build a complete parameter file for an utterance. A set of routines collectively called "JOT" are used to interact with KLTEXC in the manner just described.

JOT

The JOT routines are designed to allow users of KLTEXC to input and output formant parameters graphically. These routines are invoked at the command level of KLTEXC with the command J (Jot). When this command is issued the user is immediately queried to determine the highest frequency to be displayed so that the output scale will match a spectrogram with equivalent linear expansion. When this information is supplied, three sub-commands then become available for the user: PL (plot), TR (trace), and EX (exit).

The subcommands are two letter sequences to distinguish them from KLTEXC command letters. PL (plot) displays the formant frequencies currently residing in the parameter buffer. TR (trace) allows the user to trace in the formants of a spectrogram or to input an idealized schematic of an utterance on the tablet. As the user is tracing a formant it is simultaneously displayed on the VT-11 and entered into the parameter buffer. EX (exit) allows the user to exit from the JOT routines back to the KLTEXC command level.

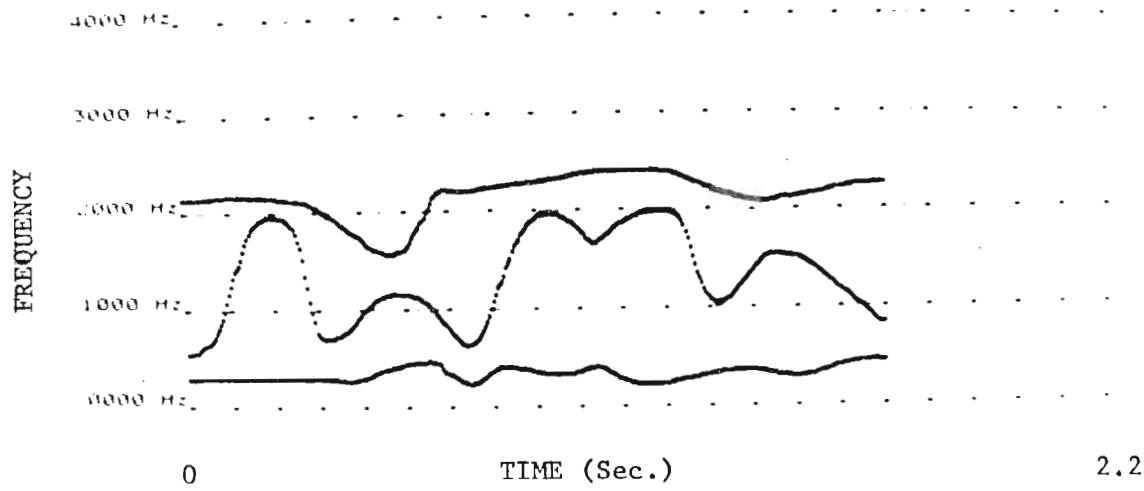


Figure 1. Schematic representation of the first three formants of, "We were away a year ago." as traced from a spectrogram of a natural utterance.

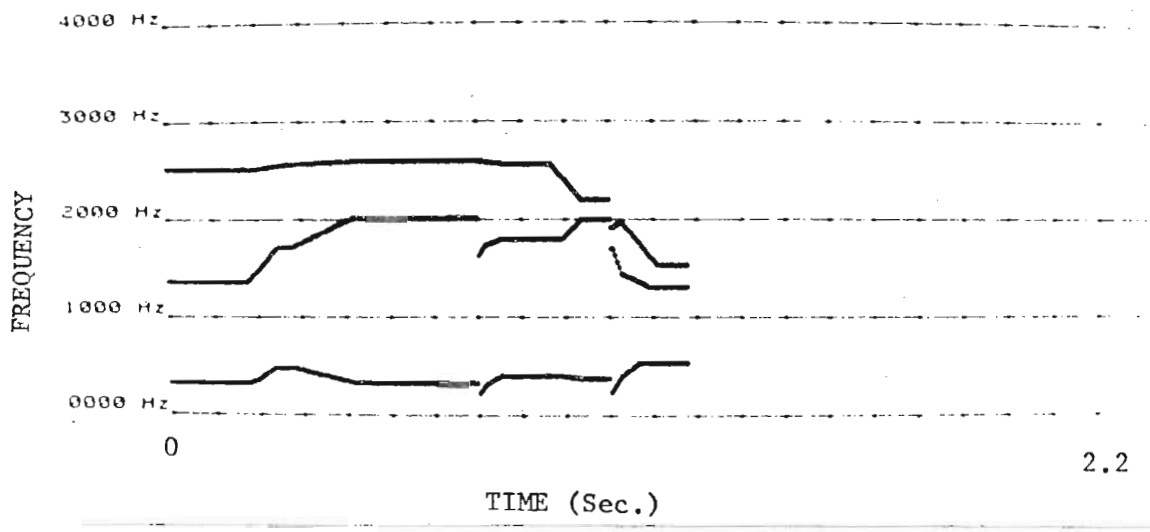


Figure 2. Schematic representation of the first three formants of, "Say digger." as input from the keyboard.

An example should help to clarify the use of these commands. Let us assume that we need to enter the formant frequencies for the utterance, "We were away a year ago." The first step is to obtain a parameter file; either a new one using the B (buffer) command of KLTEXC or an old one using the F (fetch) command. Next, the user enters the Jot mode with the J command and places the spectrogram on the digital tablet. The maximum frequency query is then answered in order to allow the VT-11 representation to model the spectrogram as closely as possible. Next, the user calls the sub-command TR (trace) to enter one formant at a time. When the user is finished entering formants the EX (exit) command is called and the user returns to the KLTEXC command level where the new parameter file may be saved or synthesized. Figure 1 shows the VT-11 display of "We were away a year ago" as traced in using the Summagraphics tablet.

While in Jot it is also possible to use the PL (plot) command. This is useful for viewing a parameter file before changing it or for viewing a parameter file that has been input from a keyboard. Figure 2 shows an example of "Say digger" as input from the keyboard and displayed on the VT-11. In summary, Jot allows the KLTEXC user to interact with the formant parameters graphically with the Trace, Plot, and Exit commands.

Waveform Displays

A visual display of the waveform produced by synthesis can also be a great advantage to the user. In the near future, these displays will will be a part of the KLTEXC synthesis program. However, we have temporarily been using a separate program called DRAWAV. DRAWAV has three primary display functions. One command displays the entire contents of the waveform buffer as seen in Figure 3. The horizontal lines indicate

how the maximum amplitude range has been used. The top and bottom lines correspond to ±11 bits, or the full range of the D/A. The lines at half-scale then indicate ±10 bits used. Peak clipping is obvious when present in the waveform.

Having examined the entire waveform, another command allows the user to specify a subsegment to be expanded, as seen in Figure 4. This expansion allows the user to make a very detailed examination of the temporal structure of the waveform. This is particularly useful if the user finds some aspect of the synthesized signal unacceptable or strange after listening to a synthesized version.

Another feature of DRAWAV is the capability of calculating and displaying the RMS energy under a 20 ms window over the entire utterance. Visual and numeric displays of the signal energy can be very important in generating certain kinds of utterances used in controlled perceptual experiments of the kind carried out routinely in our laboratory. This feature will be made more general with the output in decibels when the waveform displays are incorporated into KLTEXC.

In summary, over the past year or so we have been able to add two very useful graphics displays to the KLTEXC as well as providing a means for entering continuous data via the digital tablet. Other work is continuing on these projects to further improve their usefulness for researchers in the laboratory.

Acknowledgement

The development of these programs was supported by NIH grant NS12179 to Indiana University.

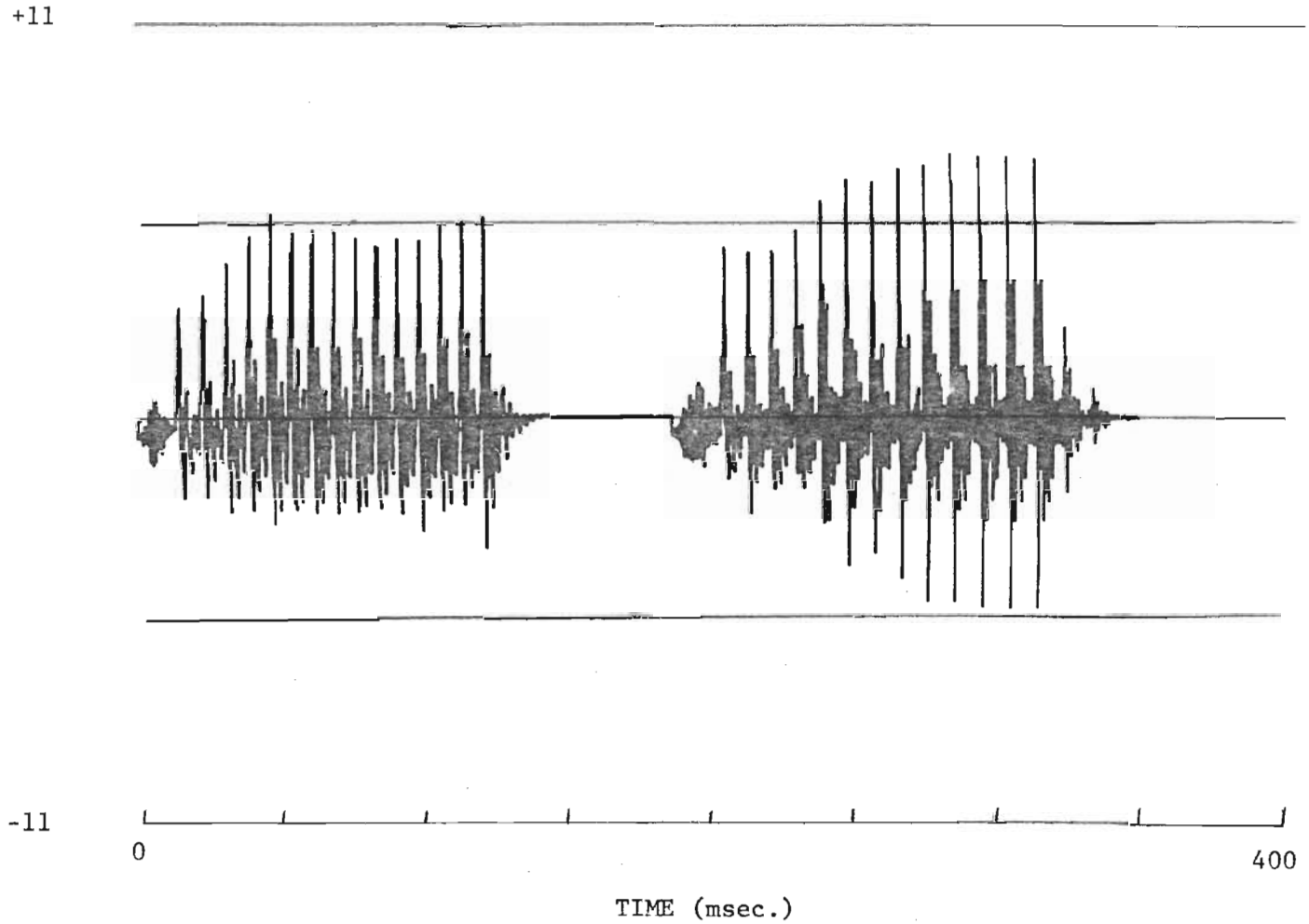


Figure 3. Display of the waveform buffer containing the utterance, "digger."

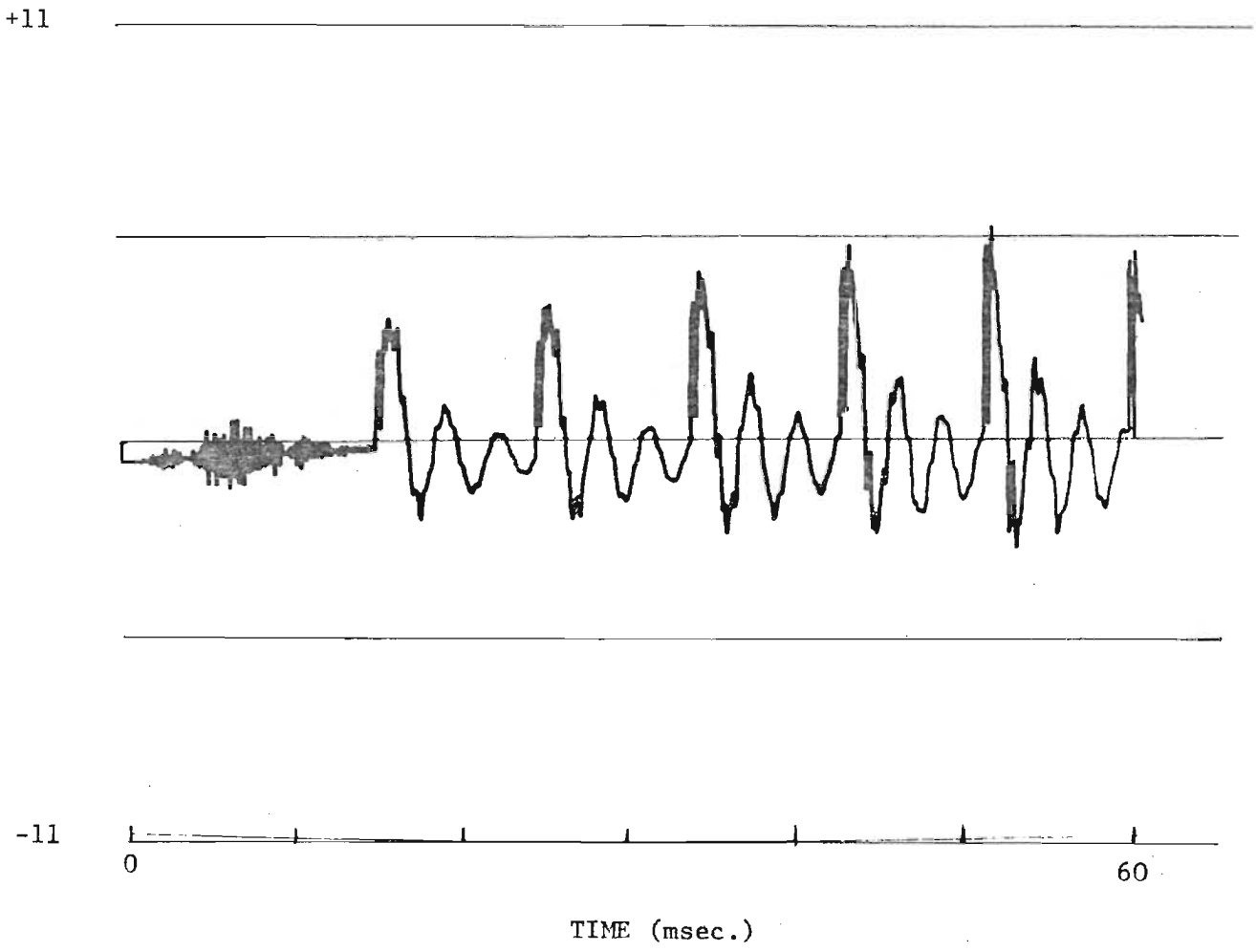


Figure 4. Display of the first 60 msec of the waveform buffer containing the utterance, "digger."

References

Kewley-Port, D. KLTEXC: Executive program to implement the KLATT software speech synthesizer. RESEARCH IN SPEECH PERCEPTION Progress Report No. 4, 1978. Pp.

Publications:

Pisoni, D. B. Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stop consonants. Journal of the Acoustical Society of America, 1977, 61, 5, 1352-1361.

Liberman, A. M. & Pisoni, D. B. Evidence for a special speech perception s bsystem in the human. In T. H. Bullock (Ed.), Recognition of complex acoustic signals. Berlin: Dahlem Konferenzen, 1977, Pp. 59-76.

Sawusch, J. R. Processing place information in stop consonants. Perception & Psychophysics, 1977, 22, 417-426.

Sawusch, J. R. Peripheral and central processes in selective adaptation of place of articulation in stop consonants. Journal of the Acoustical Society of America, 1977, 62, 738-750.

Cutting, J. E. & Pisoni, D. B. An information processing approach to speech perception. In J. F. Kavanagh & W. Strange (Eds.), Implications of basic speech and language research to the school and clinic. Cambridge: The M.I.T. Press, 1978, Pp. 38-72.

Pisoni, D. B. Speech perception. In W. K. Estes (Ed.), Handbook of learning and cognitive processes: Volume 6. Hillsdale, NJ: Erlbaum Associates, 1978, Pp. 167-233.

Sawusch, J. R. & Pisoni, D. B. Simple and contingent adaptation effects for place of articulation in stop consonants. Perception & Psychophysics, 1978, 23, 125-131.

Manuscripts to be Published:

Pisoni, D. B. On the perception of speech sounds as biologically significant signals. To appear in a special issue of Brain, Behavior, and Evolution, 1979. (In Press).

Pisoni, D. B. Review of "Speech Recognition" by D. R. Reddy. Journal of the Acoustical Society of America, 1979. (In Press).

Aslin, R. N. & Pisoni, D. B. Some Developmental Processes in Speech Perception. In G. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (Eds.), Child Phonology: Perception and Production. New York: Academic Press (In Press).

V. Laboratory Staff and Personnel:

David B. Pisoni, Ph.D. --- Professor of Psychology
Richard N. Aslin, Ph.D. -- Assistant Professor of Psychology
Robert E. Remez, Ph.D. --- Visiting Assistant Professor of Psychology

Jerry C. Forshee, M.A. --- Computer Systems Analyst
Diane Kewley-Port, M.S. -- Research Associate

Alan J. Perey, B.A. ------ Research Assistant
Beth L. Hennessy, B.A. --- Research Assistant
Thomas D. Carrell, B.A. -- Research Assistant

David Link ---------- Electronics Engineer
Nancy Layman ---------- Administrative Secretary