TEMPO OF SPECTRUM CHANGE AS A CUE IN SPEECH-SOUND DISCRIMINATION BY INFANTS

JAMES HILLENBRAND, FRED D. MINIFIE, and THOMAS J. EDWARDS

University of Washington, Seattle

Six- to seven-month-old infants were tested on their ability to discriminate among three speech sounds which differed on the basis of formant-transition duration, a major cue to distinctions among stop, semivowel and diphthong classes. The three speech sounds, [be], [we], and [ue] were produced in two different ways. The stimuli for one experiment were two-formant synthetic tokens which differed in formant-transition duration. The stimuli for a second experiment were produced with a computer-modification technique which artificially shortened or lengthened the formant-transition portion of a naturally produced [we], resulting in tokens of [be] and [ue]. The discrimination procedure involved visual reinforcement of a head-turn response following a change from a repeating background stimulus to a contrasting stimulus. Infants in both experiments discriminated [be] from both [we] and [ue]; evidence for [we]-[ue] discrimination was obtained for the "computer modified" tokens only. These findings are discussed in terms of possible mechanisms underlying speech perception in infancy.

Accounts of developmental speech perception have included descriptions of specific acoustic features which are discriminable by young infants (Eimas, 1975a; Kuhl, 1978; Morse, 1978). Studies of speech perception by infants have identified a wide variety of such features which babies can discriminate at six months of age or younger, including: voice-onset time (Eimas, Siqueland, Jusczyk and Vigorito, 1971; Trehub and Rabinovich, 1972; Lasky, Syrdal-Lasky, and Klein, 1975; Streeter, 1976; Eilers, Wilson, and Moore, in press; Eilers, Gavin, and Wilson, 1978), direction and extent of formant transitions (Morse, 1972; Eimas, 1974; 1975; Till, 1976; Jusczyk and Thompson, 1978; Jusczyk, 1977; Williams, 1977a; Williams and Bush, 1978), a variety of cues to distinctions among naturally produced fricatives (Eilers and Minifie, 1975; Eilers, 1977a; 1977b; Eilers, Wilson, and Moore, 1977; Holmberg, Morgan, and Kuhl, 1977), formant-frequency cues to vowel distinctions (Trehub, 1973; Kuhl and Miller, 1975; Swoboda, Morse, and Leavitt, 1976; Kuhl, 1976; 1977), and cues to nonsegmental distinctions, such as stress (Spring and Dale, 1977; Jusczyk and Thompson, 1977) and intonation (Morse, 1972; Kuhl and Miller, 1975).

Although this inventory of the young listener’s phonetic abilities is impres-
sive, virtually all of these studies have examined the infant’s facility to make discriminations within a given manner class. Further, aside from the many studies of voice-onset time discrimination, relatively little attention has been directed toward the infant’s ability to resolve information within the time domain. Our interest in the present study was to examine the ability of six- to seven-month-old infants to discriminate among speech sounds differing along a stop-semivowel-diphthong continuum cued by the tempo of spectrum change at stimulus onset. This study was guided, to a large extent, by data collected on adult listeners by Liberman, Delattre, Gerstman, and Cooper (1956). These investigators showed that the tempo of formant-frequency change is sufficient to cue distinctions among stops, semivowels, and diphthongs. Liberman et al presented adult listeners with two-formant synthetic stimuli like those shown in Figure 1. The starting and ending frequencies of the transitions were identical for all stimuli; formant-transition durations were varied from 10 to 300
msec. Subjects were asked to identify the stimuli as [be], [we], or [ue]. The authors reported that if the transition portion was short, less than about 60 msec, listeners tended to identify the stimulus as a stop. For transition durations between about 60 and 150 msec, listeners tended to hear a semivowel. Beyond about 150-200 msec, subjects tended to hear a diphthong. Similar results were obtained for a series of stimuli that contained falling instead of rising second formants: the cross-over points in this [ge]-[je]-[ie] series occurred at transition durations comparable to those found in the bilabial series. (For reasons that are not entirely clear, the stop-semivowel change occurred at a somewhat greater duration in the velar series than in the bilabial series. For a discussion of this, see Cooper, Ebert, and Cole, 1976.)

As Liberman et al (1956) pointed out, these results can be attributed to one of two components of what they called “tempo”: the rate of frequency change (that is the slope of the transition) or the duration of the transition. To determine which of these features was responsible for the phonetic distinction, the [b]-[w] contrast was tested in a variety of vowel contexts. Varying the vowel context made it possible to test a given transition slope at several transition durations, and vice versa. Their results showed that the [b]-[w] change occurred at a relatively stable transition duration, independent of the rate of frequency change. Liberman et al (1956) concluded that “duration of transition is the essential component of what we have been calling ‘tempo’—that is to say that duration rather than rate is the controlling cue.” (p. 135) (It is perhaps wise to view this conclusion cautiously since it is based exclusively on the results of this vowel-variation condition. Although not very parsimonious, the possibility of a vowel-consonant rate effect can not be ruled out.)

One further point worth noting about the Liberman et al study is that amplitude and bandwidth change during the transition co-varied with frequency change; that is, increases in amplitude and bandwidth occurred during the formant transition that were comparable in tempo to the change in frequency. It seems possible, if not likely, that one or both of these parameters interact with tempo of frequency change to cue these manner-class distinctions (Diehl, 1976). For this reason, tempo of spectrum change is perhaps a better term, with spectrum being used to denote overall spectrum shape, including frequency, amplitude, and bandwidth characteristics.

The present investigation was designed to test whether six- to seven-month-old infants could discriminate among three speech sounds, [be], [we], and [ue], which differed in the tempo of spectrum change at stimulus onset. Two experiments were run. The first tested infants with synthetic stimuli and the second used natural speech sounds which were modified by computer to effect the desired temporal variations. The synthetic stimuli were used so that comparisons to the adult data reported by Liberman et al (1956) would be less hazardous. The purpose of the second experiment was to test the same contrasts with stimuli more closely approximating natural speech.
EXPERIMENT I

Method

Stimuli. The stimuli for Experiment I were two-formant synthetic tokens of [be], [we], and [ue] generated on the Haskins Laboratory parallel-resonance synthesizer according to the specifications of the stimuli used in the Liberman et al. (1956) study. The transition durations were 30 msec, 120 msec and 250 msec for [be], [we], and [ue], respectively. Starting frequencies for all three stimuli were 120 Hz for F₁ and 1080 Hz for F₂. The formants rose linearly to steady-state values appropriate for the vowel /e/ (480 Hz for F₁ and 1800 Hz for F₂). All stimuli had a monotone fundamental frequency of 120 Hz. It should be noted that, although the acoustic parameters were nominally the same as those used by Liberman et al. (1956), the original Haskins stimuli were produced on the Pattern Playback (Cooper, Liberman, and Borst, 1951) rather than the parallel-resonance synthesizer.

Audio recordings of the synthesized stimuli obtained from Haskins Laboratories were subsequently digitized and stored in the disk memory of a PDP 11/10 computer at the University of Washington. A sampling rate of 20 kHz was used, with eight-bit amplitude resolution. The gain on an input amplifier was adjusted to make maximal use of the ±4-volt dynamic range of the computer. In order to prevent aliasing, and to reduce any noise introduced by the digitizing process, all signals going into and coming out of the computer were low-pass filtered at 8 kHz and passed through an auto-correlator noise-reduction unit (Phase Linear 1000).

A four-channel tape deck with simul-sync capability (TEAC 3340-S) was used to record the three speech samples with the stimulus arrangement shown in Figure 2; that is with the stimuli synchronous on three separate channels of the audio tape and a one-second inter-stimulus interval. This was accomplished by converting one of the digitized stimuli into analog form at one-second in-

![Figure 2. Arrangement of stimuli on audio tape.](image-url)
tervals and recording it on channel one of the audio tape. The tape was then rewound and a second stimulus was recorded synchronously on channel two. The onsets of the stimuli on the two channels were synchronized by using each of the samples recorded on channel one to trigger the D/A output of the second stimulus. (The "simul-sync" option on the TEAC 3340-S allows the record and playback heads to be synchronized.) The same procedure was used to record the third stimulus on channel three, again triggering the D/A output of the computer by the onset of each audio signal from channel one of the tape.

To determine the overall quality of the stimuli, an identification test was run on adult listeners. Subjects were twenty-three University of Washington students taking an introductory course in phonetics. Five tokens of each stimulus were dubbed off the tape prepared for the infant discrimination task and presented to subjects in random order at three-second intervals. Listeners were asked to identify the stimuli as [be], [we], or [ue]. Results indicated that both [be] and [we] were readily identifiable (100% and 98%, respectively), but that [ue] was misidentified as [we] 21% of the time.

To determine if the [we] and [ue] tokens were discriminable by adult listeners, three phonetically trained and three phonetically untrained listeners were tested on a discrimination task. Subjects were run in the same room, using the same equipment and stimulus tapes, and with a repeating-standard procedure nearly identical to that used to test infants (described in detail in the next section). Subjects were given six trials, each four seconds in duration. On three of these trials a change occurred from [we] to [ue] (or vice versa) and on the other three no change occurred. The trials were presented in random order; subjects were asked to press a button when they detected a change in the sound being presented.

Five of the subjects completed the task without an error. The remaining subject failed on two consecutive blocks of six trials, producing only a single correct response to a stimulus change and no responses on trials in which no stimulus change occurred. This subject was eventually successful on a third test after being instructed to listen carefully for a subtle change. All of the subjects reported that this contrast was difficult, requiring a good deal of concentration.

*Procedures, Apparatus, and Subjects.* The test procedure was a visually reinforced head-turn paradigm originally developed by Wilson, Moore and Thompson (1976) for obtaining auditory thresholds from infants, and adapted recently to study speech discrimination by Eilers et al (1977). The basic idea of the discrimination procedure is to train the infant to make a head turn in the direction of the sound source when a change occurs from a repeating background stimulus to a comparison stimulus. A visual reward is used to reinforce the head turn.

A schematic of the experimental site for the present study is shown in Figure 3. The infant was seated on the parent's lap at a small table with an assistant seated to the right. A visual reinforcer was housed in a smoked plexiglass box directly in front of a loudspeaker (Electro-Voice SP-12), one meter
to the infant's left at a 45° angle. The visual reinforcer was an electrically operated stuffed toy bear. When activated, the plexiglass box was illuminated and the bear tapped on a drum. The loudspeaker and the visual reinforcer were mounted on stands, and both were approximately at eye level to the infant. An experimenter was housed in an adjoining control room, separated from the experimental room by a one-way observation window. The control room contained a four-channel tape deck (TEAC 3340-S) and a logic device. The intensity of the recorded stimuli was controlled by a clinical audiometer (Maico MA-24). Signals were calibrated by a sound-level meter (Bruel and Kjaer Model 2203) with the microphone at the position of the infant's head. Calibration was checked before each discrimination test.

To reduce the possibility that the parent or assistant might respond to a stimulus change, and potentially cue the infant's response, music was presented
over headphones to both adults in the test room at an intensity sufficient to mask a change from one stimulus to another. The music could be interrupted by an audio intercom which allowed communication between the experimenter, assistant, and parent.

The technique involved training the infant to turn toward the loudspeaker when a change occurred from one stimulus to another. The assistant entertained the infant at midline by manipulating silent toys. With the baby facing the center of the room, the experimenter changed tape recorder channels from a repeating background stimulus to a comparison stimulus, for example, from [we] to [be]. If a head turn toward the loudspeaker was observed within four seconds of the stimulus change (three repetitions of the comparison stimulus), the infant was reinforced by the activation of the toy. After three seconds of reinforcement, tape-recorder channels were changed from the comparison stimulus back to the repeating standard.

The head-turn response was conditioned by pairing the presentation of the visual reinforcer with a stimulus change. Because the animated toy was a prominent event, both visually and auditorily, its activation nearly always resulted in a head turn away from the assistant and toward the reinforcer/loudspeaker. After a variable number of these conditioning trials, the infants generally began to turn toward the loudspeaker spontaneously when the stimulus change occurred. Training of the head-turn response was facilitated by initially incorporating an intensity difference between background and comparison stimuli—at the start of training the background stimulus was presented at 50 dBA and the comparison stimulus at 70 dBA. After each correct head turn, the intensity of the comparison stimulus was reduced by 5 dB; after each miss, the intensity of the comparison stimulus was increased by 5 dB. When the infant made three consecutive head turns with a background-comparison intensity difference, the intensity of the comparison was reduced to 50 dBA (equal to the intensity of the background stimulus) and the discrimination test was begun. For the infants who completed the experiment, the number of trials required to meet the training criterion ranged from three to 29, with a mean of 9.1 trials.

A discrimination trial was begun when the experimenter judged the infant to be in the "ready state," that is, quiet and attending to the toys being manipulated by the assistant. At this time the experimenter initiated a four-second observation interval. The interval was signaled to the experimenter by an audio tone and to the assistant by a hand-held vibrating pin. On half of the trials a change occurred from background to comparison (stimulus-change trials) and on the other half the infant continued to hear the background stimulus (control trials). If, on stimulus-change trials, the experimenter and assistant observed a head turn during the four-second response interval, they independently pressed "vote" buttons which activated the reinforcer—that is, the toy was activated when both judges indicated the presence of a head turn during the four-second observation interval. The toy was not activated if either judge failed to vote, or if either vote was outside of the interval.
On control trials, the experimenter recorded the votes of both judges, but reinforcement was never provided, regardless of the infant's response. Control trials on which either judge signaled the presence of a head turn were considered errors.

A probability generator was used to select, with a 0.5 probability, whether a stimulus-change or control trial would be presented. The experimenter bypassed the probability generator and manually selected trial type to insure that (1) a given discrimination test would contain an equal number of stimulus-change and control trials and (2) no more than three in a row of either trial type would be presented. (This last constraint was motivated by tendencies on the part of some of our infants to either lose interest in the task or increase their proportion of false-positive head turns after several change trials.)

Judges were instructed to vote only when a full 45° head turn was observed. Eye and body movements in the appropriate direction were ignored. The head-turn response was found to be relatively unambiguous, with interjudge agreement, sampled over 30 sessions (372 trials), averaging 96%.

Subjects. Nine normal six- to seven-month-old infants (five males and four females) served as subjects. Subjects were obtained by mail solicitation to the parents of newborns in the Seattle area. A parent questionnaire was used to screen out infants who had been treated for ear infections, had a family history of hearing loss, or whose parents had any concerns about the health and development of their child. Four of these subjects (two males and two females) were eliminated for failing on two consecutive visits to meet the training criterion of three consecutive head turns with an intensity difference between the background and comparison stimuli.

Order of Presentation. The order of presentation of the contrasts was randomized with one constraint: pilot data, and the results of the adult listening tests, suggested that [we]-[ue] might be the most difficult contrast. To make training more efficient, it was never presented first. The token serving as the repeating background stimulus was appropriately counterbalanced across subjects.

Criterion for Discrimination. The original design of the study called for a discrimination criterion similar to that employed by Eilers et al. (1977) in which a comparison is made, for each subject's responses, of the proportion of head turns during stimulus-change intervals to the proportion of head turns during control intervals. An infant was said to discriminate a contrast if he or she tended to turn during change intervals but not during an equal number of equal-length control intervals. According to the type of analysis employed by Eilers et al. (after Bruning and Kintz, 1968), either five appropriate responses in six trials or six appropriate responses in eight trials reach statistical significance. A discrimination test, therefore, was ended when the infant reached this criterion, or at the end of eight trials, whichever came first. If an infant failed on a contrast, it was retested once. However, following completion of the experiment, the statistical analysis used to determine criterion was
judged to be inappropriate. The data analysis eventually used, described in the next section, made use of group rather than individual data, and examined the responses on the first six trials only.

Results

Table I shows the number of head turns observed during stimulus-change and control intervals for each subject tested on the three synthetic contrasts. Data are shown for the first six trials only. The last two responses from subjects given eight trials, as well as all retests, have been excluded from analysis. (Data from all the trials that were run are presented in Appendix A.) Inspection of Table 1 shows that, for the [be]-[we] and [be]-[ue] contrasts, more head turns were observed during stimulus-change intervals than control intervals for all of the infants. For the [we]-[ue] contrast, the number of responses observed during change intervals was greatly reduced and there is essentially no evidence that the infants were turning more often during the change intervals. These trends are shown graphically in Figure 4 which displays, for each contrast, the total number of responses from all subjects during the two types of intervals.

The results of correlated t-tests comparing the number of responses in the two types of intervals are shown in Table 2. This comparison reached significance beyond the 0.05 level of confidence for [be]-[we] and [be]-[ue], but not for [we]-[ue].

Subjects in the first experiment demonstrated an ability to detect differences between stimuli with a rapid change in the spectrum from those with either

<table>
<thead>
<tr>
<th>Contrast</th>
<th>t value</th>
<th>p less than</th>
</tr>
</thead>
<tbody>
<tr>
<td>b-w</td>
<td>3.09</td>
<td>0.05</td>
</tr>
<tr>
<td>b-u</td>
<td>3.14</td>
<td>0.05</td>
</tr>
<tr>
<td>w-u</td>
<td>0.54</td>
<td>NS</td>
</tr>
</tbody>
</table>

Table 2. Results of correlated t-tests on head-turn data from Experiment I (df = 4).
a moderate or a slow spectrum change. They appeared to have much more difficulty resolving the moderate versus slow spectrum change that was used to cue the synthetic [we]-[ue] contrast. However, the results of our adult identification tests, as well as informal, subjective listening, suggested that these simplified two-formant synthetic patterns were only a rough approximation to natural speech sounds. For this reason, a second experiment was run to test infants on the same contrasts with stimuli that represented a closer (though not exact) approximation to natural speech.

Method

Stimuli. The stimuli for this experiment were produced by artificially modifying a natural speech sample, using the ALSAS programs developed by Edwards (1978) for the Speech Analysis Laboratory at the University of Washington. A natural [we], spoken by a male talker, was tape recorded, digitized, and stored in computer memory, using the same procedures that were described for the synthetic samples. The digitized signal was then displayed on a computer graphics terminal and segmented into pitch periods. A spectrogram was used to determine the starting and ending points of the transition. Since our interest was in modifying the portion of the signal showing spectrum change, an initial steady-state portion of 38 msec was removed.
Then, by deleting two out of every three successive pitch periods during the transition, it was possible to decrease the transition duration from 87 msec to 28 msec. The result was a computer-modified version of [be] which matched the original [we] in frequency and amplitude excursion during the transition, but differed in the rate and duration of the change. A computer-modified version of [ue] was produced by making two copies of each pitch period of the [we] transition, resulting in a stimulus with a transition duration three times that of the original [we], or about 260 msec. Using the same method, vowel durations were modified to yield constant syllable durations of 400 msec for the three stimuli. Spectrograms of the computer-modified stimuli are shown in Figure 5.

Stimulus tapes for the three contrasts were produced by recording each pair of stimuli synchronously on two separate channels of audio tape on the same four-channel tape deck used in Experiment I (TEAC 3340-S). Three two-channel tapes (rather than the single three-channel tape used in the first experiment) were produced simply because the four-channel tape deck was not available for use in playing the stimuli to subjects in the second experiment. All other recording procedures were identical to those used for the synthetic stimuli.

Twenty-six introductory-phonetics students participated in an identification task with the computer-modified stimuli employing the same techniques used to test adults on the synthetic stimuli. None of these subjects had participated in the previous study. Results showed that all three stimuli were reliably identified: correct identifications for [be] and [we] were 100 and 99%, respectively; [ue] was labeled correctly 89% of the time, with 11% [we] identifications. Because of the relatively high accuracy with which these stimuli were identified, it was not considered necessary to run any discrimination tests on adult subjects.

Procedures and Apparatus. The procedures and equipment for infant discrimination testing were the same as those described for Experiment I except that, (1) the stimuli were played on a two-channel tape deck (TEAC A-2300SD) instead of the four-channel recorder used in the first experiment, and (2) the intensity of the recorded stimuli was controlled by an amplifier (McIntosh MC-30) and attenuator (Hewlett Packard 350D) instead of the clinical audiometer used for the synthetic stimuli.

Subjects. Ten normal six- to seven-month-old infants (four males and six females) served as subjects. Infants were obtained from the same subject pool and with the same selection criteria used for Experiment I. Four of these subjects (two males and two females) were eliminated for failing to meet the training criterion on two consecutive visits.

Results

Table 3 shows the number of head turns observed during change and control intervals for each subject tested on the computer-modified contrasts. Again,
Figure 5. Spectrograms of the computer-modified stimuli.
Table 3. Number of head turns during stimulus-change and control intervals for infants tested with the three computer-modified contrasts.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Change</th>
<th>Control</th>
<th>Change</th>
<th>Control</th>
<th>Change</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td>15</td>
<td>5</td>
<td>15</td>
<td>5</td>
<td>14</td>
<td>3</td>
</tr>
</tbody>
</table>

Data are shown for the first six trials only. (Appendix B shows the data from all the trials that were run, including retests and blocks of eight trials.) It can be seen from the table that, with the single exception of subject 4 on [be]-[we], more head turns were observed during change intervals than control intervals for all three contrasts. Figure 6 shows the total number of responses during the two types of intervals. It can be seen that, for all three contrasts, many more responses were observed during the change intervals.

Table 4 presents the results of correlated t-tests comparing the number of

---

Figure 6. Total number of head turns during stimulus-change and control intervals for infants tested with the computer-modified stimuli.

Hillenbrand et al: Speech-Sound Discrimination 159
TABLE 4. Results of correlated t-tests on head-turn data from Experiment II (df = 5).

<table>
<thead>
<tr>
<th>Contrast</th>
<th>t value</th>
<th>p less than</th>
</tr>
</thead>
<tbody>
<tr>
<td>b-w</td>
<td>2.71</td>
<td>0.05</td>
</tr>
<tr>
<td>b-u</td>
<td>5.00</td>
<td>0.005</td>
</tr>
<tr>
<td>w-u</td>
<td>5.97</td>
<td>0.005</td>
</tr>
</tbody>
</table>

responses observed during change and control intervals for each contrast. This comparison was statistically reliable beyond the 0.05 level for all three contrasts. The only major difference between the results of this experiment and those of Experiment I was strong evidence for discrimination of the computer-modified [we]-[ue] contrast.

Discussion

The results of these experiments have a bearing on both theoretical and methodological issues in developmental speech perception. First, and most obvious, these results provide evidence for the infant's ability to discriminate speech sounds on the basis of an important temporal feature. They also represent a first demonstration of manner-class discrimination, previous research on consonant discrimination having been directed exclusively at place and voicing contrasts.

Stimulus Considerations. A relatively prominent feature of the results was that infants were successful on the computer-modified [we]-[ue] contrast but showed no evidence of discriminating the same contrast when synthetic tokens were used. There is good reason to believe, from the adult listening tests, that this effect is probably not peculiar to infants, and that there was simply a more pronounced quality difference between the more natural tokens.

It is not entirely clear what aspects of the signals allowed the infants to discriminate the computer-modified stimuli but not the synthetic samples. Analysis of the computer-modified tokens suggests that a plausible explanation might relate to the fact that these stimuli differed in the tempo of spectrum change across a broader frequency range than the synthetic samples. Spectral analysis showed a rise in the third formant as well as a 12.6-Hz fundamental frequency increase during the transition; the synthetic tokens consisted of two formants only and had flat fundamental-frequency contours. The two types of stimuli also differed somewhat in amplitude rise-time and formant-bandwidth characteristics.

Whatever the precise acoustical reasons for the infant's poorer performance on the synthetic samples, however, these results recommend caution in drawing inferences from synthetic to natural speech. We do not, of course, mean to question the validity of all studies in this area that have used synthetic speech sounds, and it was not the purpose of this study to make such a demonstration. Most studies examining infant speech discrimination have used more complex
synthetic patterns than those employed in the present study. These findings do, however, serve to highlight the inevitable trade-off between naturalness and precise control over acoustic parameters. The computer-modification technique used in the present study offers a fair compromise to this problem since a good deal of control can be exercised over specific acoustic characteristics without a great loss in natural quality. The present findings also serve as a reminder that generalizations from synthetic to natural speech should be made with some care, especially where negative results are concerned. (For similar comments on synthetic voice-onset time continua, see Williams, 1977b; Eilers et al, in press.)

Comparison of Infant and Adult Findings. Adults tested on the same synthetic [we]-[ue] contrast, using the same equipment and essentially the same discrimination-testing technique, succeeded where the infants had failed. It is tempting to conclude from this that the adults’ linguistic experience gave them a perceptual advantage over the infant, and thereby argue for the role of listening or speaking experience in the acquisition of speech-discrimination skills. It should be remembered, however, that any discrimination task involves a good deal of the listener’s memory, attention and other cognitive skills, and it is unreasonable to suppose that infants and adults are well matched on these abilities. The remarks of the adult subjects on the amount of attention necessary to succeed on this contrast are suggestive that it was an ability, or perhaps motivation, to attend that separated infants and adults on this task. The fact that the single adult subject who initially failed on this contrast subsequently passed after being asked to attend very carefully also supports this idea. It should also be noted that the assistant’s toys competed with the stimuli and visual reinforcer for the infant’s attention; our adults obviously did not have to contend with this distraction.

It is possible, of course, that the adults’ superior performance does, in fact, reflect a perceptual advantage, and that this advantage is attributable to linguistic experience. However, until a method is found which reliably sorts out perceptual and cognitive factors, comparisons between infants and adults should be made cautiously.

Considerations such as these combine with the earlier comments on stimulus problems to amplify a familiar theme in infant research: negative results are generally open to several competing explanations and great care should be taken in drawing conclusions from them.

Procedural Considerations

The discrimination task used in these experiments represents an important step in the development of the visually reinforced head-turn technique by eliminating several potential sources of bias. First, in the original Eilers et al procedure, both the assistant and the parent were allowed to hear the stimulus change, creating the possibility that some modification in their behavior could cue the infant’s head turn. Also, the assistant was aware of when a con-
control trial was presented. Since the assistant was entertaining the infant with toys, it was possible for him to decrease the likelihood of head turns on control trials simply by being more interesting during these intervals. The presentation of music over headphones to the parent and assistant masked the stimulus change and excluded the possibility of any type of bias resulting from their knowledge of trial type.

Another potential source of bias in the original version of the procedure resulted from the experimenter's knowledge, in advance of an observation interval, of whether a stimulus-change or control trial would be presented. This created the possibility that the experimenter's judgment of "ready state" would vary with trial type. Specifically, the experimenter might tend to present control trials when the infant was quite absorbed in the assistant's activities—decreasing the likelihood of a false-positive head turn—and to present stimulus-change trials when the infant was only moderately interested in the toys—increasing the likelihood of a correct head turn. The use of the probability generator to select trial type circumvented this problem. The possibility for this type of bias still existed, however, on trials in which the experimenter bypassed the probability generator to insure equal numbers of each trial type and to prevent four or more consecutive trials of the same type. We have not yet come up with a satisfactory solution to this problem.

Another problem that remains in the version of the procedure described here involves the discrimination criterion. The use of relatively short trial blocks of fixed length was originally appealing because of its potential usefulness as an assessment technique with clinical populations, such as hearing-impaired or mentally retarded infants (Wilson, 1978). As a research tool, however, this method is problematic because, (1) the small number of trials makes statistical analysis at the level of individual subjects hazardous, and (2) the analysis yields only a binary decision: infants either show evidence of discriminating a contrast, or they do not. Information on the relative difficulty of various contrasts is not readily available. A method currently in use in our laboratory adopts a more-or-less arbitrary criterion (for example, nine correct responses in ten trials—Kuhl, 1976; 1977). The infant is then allowed a relatively large number of trials to meet the criterion. Besides offering the infant a fair opportunity to show evidence of discriminating the contrast, the record of the number of trials required to meet the criterion provides a metric of the relative difficulty of the contrast. (For a more complete discussion of this issue, see Holmberg, 19781 and Morgan, 19782.)

Thoughts on Future Research

It is often the case that studies in infant speech perception are inspired by

experimental findings on the perception of speech by mature listeners. A recent study on adult listeners by Minifie, Kuhl, and Stecher (1977) raises some interesting questions about how the use of temporal features, such as the one examined in the present study, might develop. Minifie et al (1977) tested a [b]-[w] contrast in sentence-length utterances at three overall speaking rates. They found that the [b]-[w] boundary changed systematically with variations in speaking rate. Specifically, the change from [b] to [w] occurred at shorter transition durations in the fast-speaking-rate condition, and vice versa for the slow speaking rate. These results suggested that adult listeners were "taking into account" overall speaking rate in making their decisions on phonetic contrasts. The findings of the present study indicate that the ability to discriminate this temporal contrast is developed before the child has gained control over meaningful speech. It is not at all clear whether the "remote context" effects reported for adults by Minifie et al (1977) (see also Port, 1976, 1978, and Miller and Liberman, 1978, for similar effects on other speech continua) would also be demonstrated by pre-linguistic infants. It is possible that this type of "taking into account" operation might require rather sophisticated learning about the acoustic effects of variations in speech rate. Plans are currently underway to put this question to a test in our infant laboratory.

ACKNOWLEDGMENT

This work was supported by a Research Contract from the National Institute of Child Health and Human Development (HD 32793). The authors are indebted to Patricia Kuhl for her valuable advice throughout the project and for her comments on earlier versions of this paper. We are also grateful to Tristan Holmberg, Timothy Cullinan, and John Moore for their help in running subjects, to Gary Kuhn for his assistance in the synthesis of stimuli at Haskins Laboratories (NICHD-HD 01994), and to Dan Stivers, Kyum-Ha Lee, and Wes Wilson who designed much of the equipment used in these experiments. In addition, we are indebted to two anonymous reviewers for suggesting analysis of group rather than individual data. Requests for reprints may be sent to James Hillenbrand, Child Language Study, CDMRC WJ-10, University of Washington, Seattle, Washington 98195.

REFERENCES


Hillenbrand ET AL: Speech-Sound Discrimination 163


Received February 6, 1978.
Accepted July 31, 1978.

APPENDIX A

Number of head turns during stimulus-change and control intervals for all the trials run in Experiment I.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Contrast</th>
<th>b-w</th>
<th>Control</th>
<th>b-u</th>
<th>Control</th>
<th>w-u</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Change</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Control</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Change</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Control</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Change</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td>17</td>
<td>3</td>
<td>15</td>
<td>4</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

APPENDIX B

Number of head turns during stimulus-change and control intervals for all the trials run in Experiment II.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Contrast</th>
<th>b-w</th>
<th>Control</th>
<th>b-u</th>
<th>Control</th>
<th>w-u</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Change</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Control</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Change</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Control</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>Change</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>Control</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td>20</td>
<td>5</td>
<td>20</td>
<td>6</td>
<td>16</td>
<td>3</td>
</tr>
</tbody>
</table>