The discrimination of vowel duration by infants

Rebecca E. Eilers, Dale H. Bull, D. Kimbrough Oller, and Diana C. Lewis
Mailman Center for Child Development, University of Miami, Miami, Florida 33101

(Received 2 May 1983; accepted for publication 1 November 1983)

Three groups of nine 5-11-month-old infants provided evidence of discrimination of speechlike stimuli differing only in vowel duration. Ease of discrimination was directly related to the magnitude of the ratio of the longer to shorter vowel. Group one infants discriminated three vowel duration contrasts (with ratios of 0.33, 0.67, and 1.0) embedded in a synthetic /mad/ syllable; group two discriminated these same duration contrasts within the bisyllabic /samad/, and group three in the trisyllable /masamad/.

In all cases, the contrasting durations were carried by the last vowel of the synthetic word. These same three infant groups failed to provide evidence of discrimination of a final position released stop consonant contrast (/mat/ versus /mad/) cued by voice excitation during closure of the /d/ and not the /t/. These results suggest that vowel duration may be a primary cue for infants’ perception of the voicing of final position stop consonants.

PACS numbers: 43.70.Dn, 43.70.Ve

INTRODUCTION

Variations in segment duration play a systematic role in the organization of the speech code (Klatt, 1976). Both consonant and vowel duration, as well as overall rhythmic patterns, provide cues not only to the identification of speech segments but also to the phrase structure of utterances and the relative importance of morphemic components within that structure. Most of the research on perception of speech by adults has focused on the duration of syllabic nuclei within three contexts: as a cue to vowel identity (intrinsic vowel duration), as a cue to voicing and manner of final consonants, and as a correlate of linguistic stress.

Intrinsic vowel duration in English has been studied extensively. In general, English tense vowels are longer than lax vowels (Peterson and Lehiste, 1960), low vowels are longer than high vowels (House and Fairbanks, 1953), and diphthongs are longer than monophthongs (Peterson and Lehiste, 1960). Klatt (1975) found that differences in duration as a function of vowel identity explained about half of the variance of stressed vowel durations in connected discourse. These duration differences have been shown to be perceptually relevant (Cohill et al., 1977) and, at least in some cases, duration alone can signal the difference between two English vowels (Stevens, 1959; Mermelstein, 1978). Other languages such as Finnish, Estonian, and Serbocroatian make extensive use of phonemic vowel length distinctions (Lehiste, 1970).

In addition to providing cues to vowel identification, vowel duration differences help to differentiate following consonants. House and Fairbanks (1953), Peterson and Lehiste (1960), and Delattre (1966) have all demonstrated that vowels have greater duration before fricatives than before stops and affricates. Similarly, Kenyon (1940), Thomas (1947), and Jones (1948) asserted that vowels were longer before voiced consonants than before voiceless ones. House and Fairbanks (1953) and Peterson and Lehiste (1960) provided systematic data to support this observation, and other investigations (Denes, 1955; Raphael, 1972; Mermelstein, 1978) have shown that vowel duration alone can serve as the basis upon which final voiced and voiceless consonants are discriminated. In addition, Raphael et al. (1980a) have shown that the length of a formant transition leading into a vowel provides information concerning the duration of that vowel, which in turn influences the perception of the final consonant.

Finally, vowel duration is one of the critical correlates of linguistic stress in many languages (Fry, 1955; Oller, 1973; Lehiste, 1975). Stressed vowels are often twice as long as their unstressed counterparts, given constant wordlength, position in utterance, etc. (Oller, 1971). Using synthesized stimuli, Fry (1955) found that changes in the duration ratio of two vowels in two-syllable test words have much greater impact on stress perception than changes in the intensity ratio of the two vowels. Subsequent work confirms that duration of syllables varies systematically with the degree of stress (Klatt, 1975) and that contrastive stress and semantic novelty also result in increased duration of syllables (Umeda, 1975; Lindblom and Rapp, 1973; Oller, 1971).

It is clear that variations in vowel duration play a very basic role in the perception of speech by adults. Some information is available concerning the development of vowel duration distinctions in speech production by children, and very limited information is available concerning vowel duration perception in infancy. With regard to speech production, the following statements can be made with some confidence: (1) overall vowel duration is greater for children aged 2-10 years than for adults (Subtelny et al., 1966; Naeser, 1970; Fletcher, 1972; Di Simoni, 1974; Smith, 1976); (2) vowels before final voiced consonants are longer than before voiceless consonants, starting as early as 21 months-of-age (Naeser, 1970; Smith, 1976; Raphael et al., 1980b) but do not reach adult norms until years later (Di Simoni, 1974); (3) stressed vowels are longer than unstressed vowels in similar contexts, even in children as young as 29-31 months (Smith, 1976); (4) final position vowels are longer than nonfinal vowels, and the extent of final lengthening by 29-31 months is comparable to that produced by adults (Smith, 1976); (5) 7-13-month-old infants show reduced lengthening of final syllables in reduplicated babbling relative to adult values (Oller and Smith, 1977).
To date, most infant speech perception research has focused on perception of segmental units in monosyllabic contexts rather than on discrimination of suprasegmental features such as stress and its correlates: duration, intensity, or fundamental frequency. Notable exceptions include Kaplan and Kaplan's (1971) investigation of the effect of intonation and stress on the perception of running speech, and the work of Morse (1972) and Kuhl and Miller (1982) concerning the detection of rising and falling pitch. Spring and Dale (1977) found that infants could discriminate the position of stressed syllables in two-syllable stimuli. However, both Jusczyk and Thompson (1978) and Williams (1977) found that stress variations on a syllable within a multisyllabic utterance had no demonstrable effect on infants' ability to discriminate a place of articulation contrast in the syllable. It is important to note that the place contrasts were discriminable in unstressed syllables so that failure to find an interaction between stress and the segmental discrimination may indicate a ceiling effect. Other researchers (Trehub, 1976) have found that increasing the number of syllables in an utterance reduces the discriminability of segmental contrasts within the utterance for infants. Some recent failures to obtain discrimination of selected consonant distinctions has led to renewed interest in the role of stress to highlight difficult-to-perceive contrasts.

To date, most infant speech perception research has focused on perception of segmental units in monosyllabic contexts rather than on discrimination of suprasegmental features such as stress and its correlates: duration, intensity, or fundamental frequency. Notable exceptions include Kaplan and Kaplan's (1971) investigation of the effect of intonation and stress on the perception of running speech, and the work of Morse (1972) and Kuhl and Miller (1982) concerning the detection of rising and falling pitch. Spring and Dale (1977) found that infants could discriminate the position of stressed syllables in two-syllable stimuli. However, both Jusczyk and Thompson (1978) and Williams (1977) found that stress variations on a syllable within a multisyllabic utterance had no demonstrable effect on infants' ability to discriminate a place of articulation contrast in the syllable. It is important to note that the place contrasts were discriminable in unstressed syllables so that failure to find an interaction between stress and the segmental discrimination may indicate a ceiling effect. Other researchers (Trehub, 1976) have found that increasing the number of syllables in an utterance reduces the discriminability of segmental contrasts within the utterance for infants. Some recent failures to obtain discrimination of selected consonant distinctions has led to renewed interest in the role of stress to highlight difficult-to-perceive contrasts. In particular the role of vowel duration as a cue to perception of voicing in final stop and fricative consonants has drawn considerable attention.

Eilers (1977) demonstrated that 3-4-month-old infant's discrimination of /s/ and /z/ was context dependent. Discrimination of final position fricatives was superior to discrimination of initial position fricatives when natural vowel duration cues were present (i.e., a lengthened vowel before /z/). When vowel duration cues were removed, discrimination was not above chance. Further, infants failed to provide evidence of discrimination of a 100-ms vowel duration increment in the pair /at/ vs /at/. Eilers et al. (1977) employed these same speech stimuli with 6- and 12-month-old infants. The following developmental pattern emerged: (1) the youngest infants were able to discriminate final consonants when vowel duration cues were present, (2) fricative voicing contrasts were discriminable without additional vowel duration cues developmentally earlier than stop voicing contrasts, (3) vowel duration increments unaccompanied by segmental contrasts were poorly discriminated even by the 12-month-olds. Despite the fact that this area of investigation has yielded intriguing developmental data, little attention has been focused on the ability of infants to process vowel duration information. This is surprising since vowel duration is an important general subphonemic cue in English (e.g., for voicing of final obstruents) and its role in the phonemic system of other languages is even more extensive and pervasive. Thus far there has been relatively little study of the ability of infants to use vowel duration as either a subphonemic or a phonemic cue to detect contrasts in both single and multisyllabic environments.

The primary goal of the following study was to probe the infant's ability to discriminate vowel duration. A secondary goal was to provide a perspective on the possible role of vowel duration as a cue for final position stop consonant voicing contrasts in English. Infants were tested for discrimination of synthetic stimuli differing by 100-, 200-, and 300-ms vowel duration in one-, two-, and three-syllable contexts and on a final position stop voicing contrast cued by voice excitation only. In addition four naive adult listeners were tested on the same contrasts for purposes of comparison.

I. METHOD

A. Subjects

Thirty-eight full term 5-11-month-old infants from English-speaking homes were selected for the study. Some infants were excluded from the final subject sample because they failed the training criteria ($n = 3$), had conductive hearing loss ($n = 2$), or failed to finish testing because of parental difficulties in returning for the necessary number of visits ($n = 6$). The 27 infants in the final sample all met inclusion criteria of normal appearance of developmental milestones and normal hearing. The mean age of the subjects at the outset of the study was 7.7 months. Thirteen infants were male. In addition, four naive adult listeners with normal hearing (aged 20-30 years) served as subjects.

B. Stimuli

The pretraining stimulus pair consisted of five tokens each of the syllables /sa/ and /da/, produced by a male phonetician and matched pairwise for overall duration, overall amplitude, peak amplitude, mean fundamental frequency, and peak fundamental frequency.

The experimental stimuli consisted of one-, two-, and three-syllable stimuli, /mad/, /samad/, and /masamad/, synthesized using Klatt's (1980) software synthesis programs. The three-syllable stimulus /masamad/, served as the basis for synthesis of all other stimuli. The /masamad/ stimulus began with a 55-ms /m/ murmur, a 45-ms transition, and an 100-ms steady state /a/. The steady state /a/ had an $F_1$ of 745 (BW = 130), an $F_2$ of 900 (BW = 70), and an $F_3$ of 2400 (BW = 160) in all three syllables. After a 50-ms transition, the 60-ms /s/ frication began, followed by a 45-ms transition into an 100-ms /a/. The second /m/ murmur ($F_1$ = 275, $F_2$ = 900, $F_3$ = 2200), which followed a 55-ms transition was 100 ms in duration with a 45-ms transition into the steady-state portion of the vowel. The base duration stimulus had 200 ms of steady-state vowel followed by a 55-ms transition, providing a 300-ms vowel. The final consonant /d/ ($F_1$ = 150, $F_2$ = 1700, $F_3$ = 2600), had a closure duration of 115 ms followed by an approximate 20-ms burst. The burst sequence also included a 25-ms transition to schwa formant values. During the closure, the amplitude of voicing (AV) dipped to 15 dB below the vowel maximum, and was raised by 10 dB during the burst before falling off to zero. Amplitude of frication (AF) and amplitude of aspiration (AH) were also raised during the burst to values of — 10 and — 12 dB: the maximum AV value, respectively, before falling off. The amplitude of the third through the sixth formants during the burst were — 13, 0, 2, and 0 dB: the maximum AV value. The fundamental frequency contour on the final syllable involved a rising-falling pattern, starting from 100 Hz during the /m/ reaching 150 Hz midway
through the vowel, and dropping to 98 Hz at the end of the burst.

The base duration stimulus, /masamado/, had a total duration of 1045 ms. Vowel duration contrasts were constructed by adding 100-ms increments to the steady-state portion in the last syllable of /masamado/ so that the final position 300-ms vowel was contrasted with 400-, 500-, and 600-ms vowels. To provide /samad/, stimuli, 248 ms were trimmed from the front of the original /masamado/ stimuli. Similarly, /mad/ stimuli resulted from trimming the initial 505 ms of /masamado/. In this fashion, nine stimulus pairs were produced: three duration contrast pairs for /mad/, /samad/, and /masamado/, respectively. In addition, a tenth experimental pair, /mad300/ vs /mat300/, was constructed. These two stimuli had the same vowel duration and overall duration and only differed with respect to voicing during the final consonant closure. In the /mat300/ stimulus, all voicing was terminated (i.e., AV = 0) 40 ms after the end of the /a/ vowel transition. The burst for /t/ had all the same AF, AH, and A3-A6 values as the /d/ burst.

These ten testing pairs were recorded on individual audiotapes. For the vowel duration contrast tapes, channel 1 of each audiotape contained the 300-ms base duration stimulus. The stimuli on this channel were separated by an inter-stimulus interval (ISI) that varied randomly within a range of 300–1000 ms for the duration pairs, and was at a constant 650 ms for the /mat300/ vs /mad300/ pair. Channel 2 of each tape contained the longer duration, contrasting stimulus recorded so that stimuli on the two channels had synchronous onsets. The ISI range for the two channels on duration contrasts was selected to provide a 50% overlap of the distributions of ISIs on channels 1 and 2. This arrangement of variable ISIs minimized the possibility that discrimination could be based on ISI differences. The training tape, /sa/ vs /da/ was recorded with a constant 300-ms ISI. Computer control permitted switching channels without interrupting any stimuli. The preparation of stimuli and tapes was carried out using the VOCAL program of the University of Wisconsin Waisman Center (Gillman et al., 1975).

C. Apparatus

The experimental site consisted of a double-walled, sound-attenuated booth equipped with an HPM 100 speaker, four visual reinforcers (housed in a dark Plexiglas box), and a response box allowing communication with the computer. The adjoining control room housed high-fidelity playback and amplification equipment and a DEC 11/23 laboratory computer for controlling stimulus conditions and reinforcement and for recording and analyzing data.

D. Procedure

The present study employed the Visually Reinforced Infant Speech Discrimination (VRISD) paradigm (Eilers et al., 1982), in which infants are conditioned to turn their heads to a change in a repeating background auditory stimulus. During this procedure the infant was seated on a parent's lap in the booth, and an experimenter, seated to the child's left, attempted to keep the infant oriented toward the experimenter by manipulating a set of quiet toys. Both the parent and the experimenter listened to masking music through headphones during the entire experimental session. As an added precaution the experimenter wore earplugs as well. In this artificially "deafened" condition the experimenter was unable to pass the training criterion set for the infant when the experimenter played the role of subject.

The session began with the continuous presentation at 60 dB SPL of the pretraining stimuli on one tape channel (e.g., /da/). This background stimulus (Sb) remained on throughout the session, interrupted by the presentation of the change stimulus (Sd) on the other tape channel (e.g., /sa/). The particular stimulus channels assigned to Sb and Sd were counterbalanced for the training stimulus and for the /mat300/ vs /mad300/ pair. For all of the vowel duration contrasts, the 300-ms stimulus served as Sb.

The experimenter was responsible for monitoring the infant's behavior, initiating trials (6-s intervals during which a stimulus change does or does not occur) and recording head-turn responses. Each test and pretraining session began with a shaping phase during which the background stimulus (Sb) was changed to the contrasting stimulus (Sd) for approximately 6 s at an intensity of 12 dB greater than the background level. If the infant was observed to turn toward the speaker during Sd, the experimenter would activate the reinforcer by depressing a button on the response box. If the infant did not turn on the first trial, on the second trial the experimenter was activated toward the end of the 6-s period. After a few trials most infants turned at the presentation of the louder Sd. After two consecutive trials in which the infants responded correctly within the Sd window, the intensity of Sd was reduced in 4-dB steps until the infant responded correctly on two consecutive trials at each intensity level, and testing at matched intensity (60 dB) was begun.

During the test phase of all sessions, the change from Sb to Sd occurred in one-half of the trials (change trials). On the remaining trials (control trials), no change from Sb to Sd occurred. During control trials, head turns were recorded but no reinforcement was given. These no-change trials served to control for the infant's spontaneous, random head-turning behavior. The order of presentation of the change and control trials was pseudorandom with three possible combinations of change and control trials per block of ten trials: six change and four control, five change and five control, and four change and six control. The experimenter initiated a trial when the infant was engaged with toys and oriented toward the experimenter. The pseudorandom order of trial presentation coupled with the absence of auditory information during testing prevented the experimenter from knowing the nature of the test trial. Hence all observations of head-turning behavior were made free of experimenter bias.

Infants were trained to a criterion of nine out of ten correct successive equal-intensity test trials with the stimulus pair /da/ vs /sa/. This generally required between one and three sessions, following which the infant returned for four of the remaining ten experimental contrasts. Nine of the infants received the three one-syllable vowel duration contrasts and the /mat300/ vs /mad300/ voicing contrast, the second group of nine infants received the three two-syllable
vowel duration contrasts plus the voicing contrast and the last nine infants received the three three-syllable vowel duration contrast plus the voicing contrast. The subjects were tested on one contrast each week and contrasts were presented in a counterbalanced fashion.

In each of the testing sessions an abbreviated shaping phase was employed in which each infant received only two trials at each of three intensity level differences. This was followed by 30 test trials, approximately half of which were control trials. Finally, after every five test trials, the subject received a change trial in which a 4-dB difference accompanied the change from $S_1$ to $S_2$. These probe trials were included to maintain the infant's interest throughout the session, but head turns during these intensity-cued probe trials were not included in the analysis.

The four naive adults were asked to discriminate all ten of the experimental contrasts. Adults were tested using the identical VRISD procedure, but indicated a response by raising their hand rather than turning their heads. Feedback was provided on every trial.

II. RESULTS

The responses of each infant were blocked into three groups of ten test trials each and expressed in two ways, as a discriminative index (DI) score and as a percent correct (PC) score, for each trial block. The primary focus of the results reported in the analysis is on DI scores. Results with PC scores will also be presented to provide a comparison with this frequently used scoring method. The DI score is the number of hits minus the number of false positives divided by the number of change trials (see Morse et al., 1982 for a discussion of DI scores in VRISD testing). Both the DI and PC scores were subjected to both z tests of means to determine whether performance on a contrast differed from chance, and ANOVA to assess differences in performance between the different contrasts.

Table I presents the mean DI and PC scores, averaged across test blocks, for each of the stimulus contrasts within the different subject groups. Since each contrast was tested with nine subjects, each receiving 30 trials, the DI scores for the individual contrasts were tested against a population mean of zero, a population standard deviation of 0.198, and a standard error of 0.066 (Morse et al., 1982). The corresponding values for PC scores were 0.5, 0.092, and 0.031, respectively.

The results of the z tests of means are also included in Table I. These tests indicated that DI scores were significantly greater than chance for all contrasts with vowel duration differences, for all three-syllable conditions. All three infant groups failed to show evidence of discrimination of the /mad/ vs /mat/ contrast.

A split-plot ANOVA with one between subjects factor (syllable number) and two within subjects factors (stimulus contrast and trial block) was performed on the DI and PC scores. With DI scores this analysis indicated a significant effect for contrast, $F(3,72) = 8.60, p < 0.01$. Scheffe post hoc analyses revealed that discrimination of the /mad/ vs /mat/ contrast was significantly different from performance on the vowel duration contrasts, $p < 0.01$. No other main effects and no interactions were significant. Analysis of PC scores yielded the same pattern of results.

Inspection of Table I indicates that discrimination scores for the vowel duration contrasts increased as duration increased. A second ANOVA of these duration contrasts (omitting results for the voicing contrast) indicated a significant effect of vowel duration, $F(2,48) = 3.77, p < 0.05$. Scheffe analyses showed that the DI score for the pair differing by 100-ms vowel duration was significantly different from the score for the 300-ms pair ($p < 0.05$). The analysis indicated no other significant main effects and no significant interactions. Analysis of PC scores yielded the same pattern of results.

Infant subjects in each of the three groups were tested

<p>| TABLE I. Mean discriminative index (DI), mean percent correct (PC) scores, and results from z tests and t tests for each of ten stimulus pairs. |</p>
<table>
<thead>
<tr>
<th>Adult mean DI</th>
<th>Infant mean PC</th>
<th>DI</th>
<th>t</th>
<th>PC</th>
<th>t</th>
</tr>
</thead>
</table>
| Group 1
| /mad/ vs /mat/ | 0.92 0.05 0.48 0.82 (0.73) | 0.76 (0.64) |
| /mad/ vs /mad/ | 0.92 0.20 0.58 3.05 (2.02) | 2.67 (1.65) |
| /mad/ vs /mad/ | 0.90 0.22 0.57 3.28 (2.82) | 2.21 (1.68) |
| /mad/ vs /mad/ | 1.00 0.37 0.67 5.58 (5.99) | 5.47 (5.26) |
| Group 2
| /mad/ vs /mat/ | 0.11 0.51 1.60 (2.07) | 0.33 (0.60) |
| /samad/ vs /samad/ | 0.92 0.16 0.54 2.50 (3.33) | 1.32 (2.04) |
| /samad/ vs /samad/ | 0.94 0.26 0.59 3.89 (3.33) | 3.03 (2.17) |
| /samad/ vs /samad/ | 1.00 0.33 0.63 5.01 (5.31) | 4.29 (3.74) |
| Group 3
| /mad/ vs /mat/ | 0.03 0.50 0.46 (0.39) | -0.13 (0.04) |
| /masamad/ vs /masamad/ | 0.95 0.20 0.58 3.09 (4.04) | 2.31 (2.41) |
| /masamad/ vs /masamad/ | 0.97 0.29 0.59 4.33 (7.46) | 2.80 (2.95) |
| /masamad/ vs /masamad/ | 0.98 0.31 0.64 4.60 (3.85) | 4.62 (3.44) |

* $p < 0.05$

** $p < 0.01$
with the same /mad₃00/ vs /mat₃00/ contrast. A separate ANOVA of the DI and PC scores for this voicing contrast indicated no significant difference between the three subject groups. This analysis revealed no effects for trial blocks and no significant interactions. As can be seen in column I of Table I, adult performance on all ten contrasts ranged from DI scores of 0.92 to 1.0. Thus all contrasts were easily discriminable by adults. Near perfect performance on the 33% vowel duration contrast indicates that (using the VRISD procedure) the adult difference limen is well below 33%.

### III. Discussion

The present study has demonstrated infant abilities to discriminate vowel durations that could be relevant as contrastive cues in speech. Infant performance on contrasts differing only in vowel duration suggests a lawful relationship between the percent lengthening (33%, 67%, and 100%) and the obtained scores. All vowel duration contrasts were discriminated significantly better than chance with highest scores obtained with the largest vowel duration differences. However, it is important to note that with the 33% duration difference, infant scores were still relatively low (average DI score across syllable conditions was 0.18), lower than those we have obtained for vowel quality contrasts (usually in the range between 0.3 and 0.6; Eilers and Oiler, in press). With the 67% duration difference, the DI score approaches 0.3 (0.27); and it is not until 100% duration difference that the DI score exceeds 0.3 (0.34). Thus discrimination of vowel durations with difference ratios of less than 67% appears to be relatively difficult for infants.

While it is difficult to interpret differences between adult and infant performance in speech discrimination paradigms, adult performance with the VRISD task was near perfect for all three vowel durations. Other literature suggests that the difference limen for adult discrimination of vowel duration using conventional adult paradigms may be as little as 14% in vowels greater than 400 ms (Lehiste, 1970, for review). Thus infant difference limens for duration may be considerably higher than for adults.

Yet, even though infant vowel discrimination capabilities may be more limited than adult abilities, the potential for infant discrimination of adult linguistic contrasts cued by vowel duration has been demonstrated here. House and Fairbanks (1953) showed that voicing of final consonants in English is cued by a 69% increase in vowel duration preceding a voiced final consonant. Similarly Lehiste and Ivic (1963) have found an average difference of 65% between short vowels and long vowels under falling intonation in Serbo-Croatian, a language possessing vowel quantity distinctions. Since infants showed fairly good discrimination of vowel duration increments of 67%, one might expect them to be capable of phonemic discriminations of final consonant voicing in English and vowel quantity in Serbo-Croatian. At the same time, the results would suggest a more difficult situation for infant perception of English stress based on duration in final position syllables. Data from Oiler (1973) show less than 30% vowel duration differences between final stressed and unstressed vowels. In order to discriminate final position stress, infants might need to rely on the combination of cues that are stress correlates.

The present study replicates the earlier finding (Eilers, 1977) that infants fail to provide evidence of discrimination of final position stop consonant voicing contrast (cued by voice excitation only). Infants in both the present and previous studies performed significantly better on all vowel duration contrasts than on the contrast involving voice excitation alone, suggesting that for infants, the dominant cue to final consonant voicing status could be the relative duration of the pre-consonantal vowels. Such a pattern is consistent with the tendency of adult listeners to rely on vowel duration in discrimination of contrasts in final consonant voicing (Raphael, 1972; Denes, 1955). Still, it should be noted that in the present study no scaling of amplitude of voicing was attempted—only one contrast of voice excitation was tested. It is possible that with a greater degree of difference between the voicing amplitudes or aspiration amplitudes of the final consonants, infants could perform adequately in discrimination without vowel duration cues. At the same time, it is important to remember, however, that adult listeners had no difficulty identifying and discriminating the voice excitation contrast in question.

An unexpected outcome of the present study was the failure to observe a main effect for syllable number. Infants showed no tendency to discriminate differences in vowel duration differently as a function of whether that duration difference was embedded in one-, two-, or three-syllable stimuli. In this study the vowel duration cue was always in the final syllable. Since young children show strong recency effects in attending to linguistic stimuli (Eilers, 1973), it may be that such an effect masks any possible syllable number effect. Infants might be expected to perform more poorly on three-syllable rather than one- or two-syllable stimuli if the vowel duration cue is in medial rather than final position.

In summary, infants were able to detect differences in vowel duration in one-, two-, and three-syllable stimuli with differences as small as 33%. Infants did not, however, evidence discrimination of a final voiced versus voiceless stop consonant distinction (cued by voice excitation) in the absence of naturally occurring vowel duration cues. The present research suggests that vowel duration cues may be primary for voiced–voiceless distinctions in final position stop consonants for infants in the second half-year of life. Further infant work with scaling of voice excitation burst contrasts is necessary to provide a perspective on the relative contributions of vowel duration and other cues to voicing in final position stop consonants.

### Acknowledgments

This work was supported by NIH Grant 1R01 HD12701 and by the Mailman Foundation. Our special thanks to patient infants, generous parents, to Richard Urbano for his computer assistance, and to the Waisman Center for use of their facilities for stimulus preparation.

1 Although t tests appear to be the most appropriate statistic (Hays, 1963, p. 304) since the population mean and variance are known, f tests may also be
performed to assess differences from chance performance. T-test results are presented in parenthesis in Table I, next to the appropriate score type.


