Published in final edited form as:

J Acoust Soc Am. 1984 February ; 75(2): 581–589.

Infant discrimination of two- and five-formant voiced stop consonants differing in place of articulation

Amanda C. Walley, David B. Pisoni, and Richard N. Aslin

Department of Psychology, Indiana University, Bloomington, Indiana 47405

Abstract

According to recent theoretical accounts of place of articulation perception, global, invariant properties of the stop CV syllable onset spectrum serve as primary, innate cues to place of articulation, whereas contextually variable formant transitions constitute secondary, learned cues. By this view, one might expect that young infants would find the discrimination of place of articulation contrasts signaled by formant transition differences more difficult than those cued by gross spectral differences. Using an operant head-turning paradigm, we found that 6-month-old infants were able to discriminate two-formant stimuli contrasting in place of articulation as well as they did five-formant + burst stimuli. Apparently, neither the global properties of the onset spectrum nor simply the additional acoustic information contained in the five-formant + burst stimuli afford the infant any advantage in the discrimination task. Rather, formant transition information provides a sufficient basis for discriminating place of articulation differences.

INTRODUCTION

Recent models of the initial stages of speech processing have been quite successful in identifying invariant acoustic correlates of place of articulation (e.g., Blumstein and Stevens, 1979; Kewley-Port, 1982, 1983; Searle *et al.*, 1979, 1980; Stevens and Blumstein, 1978). For example, according to Stevens and Blumstein's model the relative slope and diffuseness of spectral energy in the short-term spectrum sampled at consonantal release in a CV syllable provides distinctive and contextually invariant information about place of articulation; specifically, labials are characterized by a diffuse-flat or -falling spectrum, alveolars by a diffuse-rising spectrum, and velars by a prominent mid-frequency spectral peak (see Fig. 1). The onset spectrum of a stop CV syllable is obtained by using linear prediction analysis and integrating energy over the first 25.6 ms of the syllable (Blumstein and Stevens, 1979; Stevens and Blumstein, 1978). The criterial properties of the onset spectrum are thus determined by integrating over the burst and the initial portions of the formant transitions. The same general spectral shapes are observed when only the formant transitions are present in a stimulus, but these shapes are enhanced by the presence of the stop release burst (Stevens and Blumstein, 1978). Thus, the defining properties of the onset spectrum are global ones, the presence of which does not depend on fine details of the CV waveform, such as the direction and extent of formant transitions.

Sensitivity to these invariant properties of the short-term spectrum could underlie the adult listener's ability to differentially identify place of articulation for stop CV syllables in the face of the contextual variability inherent in formant transition information (Blumstein and Stevens, 1979, 1980; Stevens and Blumstein, 1978, 1981). Moreover, the human auditory

system may be innately endowed with detectors that are sensitive to these *primary*, context-independent properties of the onset spectrum and thus later provide the basis for place of articulation perception. By this view, formant transition information (e.g., formant starting frequency and direction, formant frequencies of the following vowel) constitutes a *secondary*, context-dependent cue which may be invoked for place of articulation identification by adults in the absence or distortion of the primary, invariant spectral cues (Blumstein and Stevens, 1980; Stevens and Blumstein, 1978). Thus, adults are able to differentially identify the place of articulation of two-formant stimuli (Cooper *et al.*, 1952; Delattre *et al.*, 1955) that lack primary invariant spectral properties (Stevens and Blumstein, 1978; Blumstein and Stevens, 1980). This ability to use secondary formant transition cues is one that is presumably acquired in development through experience with the co-occurrence of primary and secondary cues (Blumstein and Stevens, 1979; Stevens and Blumstein, 1978, 1981). Thus, the properties of the onset spectrum are asserted to be developmentally primary for the perception of place of articulation in that they are used prior to formant transition information and sensitivity to these properties is innate.

The assumption that human listeners are innately sensitive to properties of the onset spectra for different places of stop consonant articulation could, as Stevens and Blumstein (e.g., 1978; Blumstein and Stevens, 1979) suggest, account for the ability of prelinguistic infants to discriminate place of articulation differences in synthetic three-formant and three-formant + burst stimuli (Eimas, 1974; Leavitt *et al.*, 1976; Miller and Morse, 1976; Moffitt, 1971; Morse, 1972; Till, 1976; Williams and Bush, 1978). Although previous demonstrations of the infant's ability to discriminate place of articulation differences in stop consonants lend support to Stevens and Blumstein's argument, it is not known whether the onset spectra of the stimuli employed in these studies possessed the critical properties described by Stevens and Blumstein. Nor can it be concluded that these properties, if present, mediated discrimination. Moreover, it is not clear that the existence of invariant acoustic cues need be assumed in order to account for previous findings concerning the infant's perception of place of articulation; i.e., the majority of such studies have merely shown that the infant is able to *discriminate* place of articulation *differences* within a particular vowel context (see, however, Fodor *et al.*, 1975).¹ Presumably *any* difference (i.e., not necessarily in the onset spectra) between stimuli of different places of articulation could account for such discrimination and the precise basis for discrimination could vary with the particular place contrast and vowel context.

The present study was undertaken to evaluate in a direct manner the relative salience of potential place of articulation cues for young infants. If innate sensitivity to global, contextually invariant properties of the onset spectrum typically mediates discrimination of place of articulation contrasts (and thus provides the initial basis for "parsing" acoustic information into the appropriate phonetic categories), then young infants should have no difficulty discriminating synthetic five-formant + burst stimuli possessing the appropriate onset spectra. It should be noted that Stevens and Blumstein do not explicitly propose any time frame within which infants come to be able to use secondary, formant transition information. However, they *do* suggest (e.g., Stevens and Blumstein, 1978) that the infant, even up to the age of 6 months, relies on onset spectra differences in discriminating place of articulation. Because infants have had only limited exposure to speech, they are unlikely to have learned which variations in formant transitions co-occur with the invariant spectral properties of a particular place category. Therefore, they might be less able to discriminate synthetic two-formant stimuli in which place of articulation differences are signaled *only* by

¹Although Fodor *et al.* (1975) report evidence of perceptual constancy in infants for voiceless stop consonants differing in place of articulation, their evidence is rather weak (see Aslin *et al.*, 1983) and their findings have not been replicated (cf., Katz and Jusczyk, 1980).

secondary, formant transition variations. Still, in order to learn to categorize context-conditioned cues appropriately, infants must be sensitive to the acoustic differences inherent in two-formant patterns. A further reason for expecting a difference in the discriminability of five- and two-formant stimuli is that five-formant patterns simply contain much more acoustic information than do two-formant ones.

Previous work by Eimas (1974) has indicated that infants as young as 2–3 months of age can, in fact, discriminate a two-formant /bae/ vs /dae/ contrast. The onset spectra of these two stimuli were probably quite different, although they could not have possessed the global, invariant properties proposed by Stevens and Blumstein; discrimination could, therefore, have been mediated by sensitivity to differences in *either* the onset spectra *or* the formant transitions of the test stimuli. Thus, it remains unclear precisely what acoustic cues are necessary and/or sufficient for mediating infants' discrimination of place of articulation. Moreover, no conclusions about the relative discriminability of onset spectra and formant transition cues can be made on the basis of this work. A study by Williams and Bush (1978) is suggestive in this respect. With the high-amplitude sucking paradigm, these investigators studied infants' discrimination of synthetic /da/–/ga/ contrasts in which the stimuli either did or did not contain release bursts. Using pre- versus post-decrement criterion sucking rate as a measure of discrimination, ordered differences according to level of significance were observed for the two experimental groups (burst, no burst) relative to the control group; specifically, infants' performance in the experimental plus burst group differed from the control infants' performance at a level of significance greater than did infants' performance in the experimental — burst group from control infants' performance. Williams and Bush tentatively conclude that this difference, in terms of level of significance, indicates that the contrasts between different place stimuli containing bursts are more discriminable than are those without bursts—presumably because the distinctive, invariant spectral properties of these stop consonants are enhanced by the presence of the burst (see Stevens and Blumstein, 1978). The problem with accepting this conclusion, of course, is that the level of significance attained with a statistical test does not allow any statement as to “how significant” a result is either in an absolute or a relative sense. Moreover, as Williams and Bush themselves note, the results of their study are only suggestive since the critical comparison between the two experimental groups of infants failed to reach statistical significance.

In the present study, we were therefore interested in obtaining further evidence concerning the relative discriminability for infants of stimuli with varying amounts of acoustic (specifically, spectral) information. This was achieved by examining 6-month-old infants' discrimination of five-formant + burst and two-formant place of articulation contrasts using an operant head-turning paradigm. The study is also the first to use this paradigm in a comprehensive investigation of the infant's perception of place of articulation—i.e., to study the infant's discrimination of all three of the place contrasts (/b/ vs /d/, /b/ vs /g/, and /d/ vs /g/) that are used in English. If infants' discrimination of place of articulation contrasts depends on global, invariant spectral properties, one might expect that performance for the five-formant + burst stimuli would be superior to that for the two-formant patterns. However, if the five-formant + burst stimuli contain no additional information for discrimination and relatively local details of the waveform, such as formant transition differences, are sufficient to mediate place of articulation discrimination, then no differences in discrimination performance for the two types of stimuli should be observed.

I. EXPERIMENT

A. Subjects

A total of 69 infants participated in the present experiment (mean age on the first day of testing = 6.6 months; range = 5.5–8.0 months). Fifteen (22%) of the infants failed to show any reliable evidence of acquiring the head-turn response after two sessions and were dropped from the study. Fourteen of these subjects were ones who had been assigned to the two-formant condition; the greater attrition rate for this group of infants is probably not due to the nature of the stimulus condition (i.e., two-formant versus five-formant + burst) or the particular place contrast being tested since nine of the 14 infants failed to respond to an intensity difference between the target and background stimuli during the initial shaping phase. None of the subjects, according to their parents' report, had any known chronic hearing disorder. Infants were typically tested in several sessions on successive days.

B. Stimuli

A set of five-formant stimuli with both bursts and transitions (FF stimuli) and a set of two-formant stimuli with transitions only (TF stimuli) were constructed for the present experiment. The stimuli in each set, which were perceived by adults as the syllables /ba/, /da/, and /ga/, were generated using a modified version of Klatt's (1980) software digital speech synthesizer (see Kewley-Port, 1978). The FF stimuli were modeled after stimuli constructed by Stevens and Blumstein (1978) on an earlier version of the same synthesizer (Klatt, 1977). Except where otherwise noted, all of the parameter values used in the synthesis of our FF stimuli were identical to those used by Stevens and Blumstein.

All of the FF and TF stimuli were synthesized at a 10-kHz sampling rate and low-pass filtered at 4.8 kHz on output through a 12-bit D–A converter. The duration of voicing for all stimuli was 255 ms. The excitation source for the vowel began abruptly, such that the first glottal pulse coincided with the beginning of the formant transitions. The amplitude was constant for 220 ms and fell to 0 dB over the last 35 ms. The fundamental-frequency contour began at 103 Hz, rose to 125 Hz in 35 ms, fell first to 94 Hz in 180 ms, and then to 50 Hz in 40 ms.

1. Five-formant + burst stimuli (FF)—The three FF stimuli were each synthesized with the digital resonators connected in series. These stimuli were modeled after the best exemplars, according to subjects' identification functions, of each place of articulation category (i.e., after stimulus 1, 8, and 12) in the Stevens and Blumstein (1978) full cue /ba–da–ga/ series. The center frequencies (and bandwidths) of the formants were appropriate for the steady-state vowel /a/ and were set at 720 (50), 1240 (70), 2500 (110), 3600 (170), and 4500(250) Hz, for F_1 , F_2 , F_3 , F_4 , and F_5 , respectively. All of the stimuli had the same starting frequency (220 Hz) for F_1 , but had varying transition durations 20, 35, and 45 ms for /ba/, /da/, and /ga/, respectively. The starting frequencies of F_2 and F_3 were 900 and 2000 Hz for /ba/, 1700 and 2800 Hz for /da/, and 1640 and 2100 Hz for /ga/. The transition durations of F_2 and F_3 were 40 ms. F_4 and F_5 were steady-state formants and thus had no formant transitions.

Consistent with the construction of the Stevens and Blumstein stimuli, the peak of the burst spectrum for a given stimulus was located at the starting frequency of a particular formant of the following vowel. This procedure was adopted, according to Stevens and Blumstein (1978), in order to satisfy the condition of continuity between release burst and vowel formants—a condition which arises from the fact that, in natural speech, the same vocal tract resonances are excited first by noise and later, at voicing onset, by the glottal source (although with different relative amplitudes). A burst, consisting of a single resonance peak,

was located at 900 Hz (BW = 70) for the labial stimulus, at 3600 Hz (BW = 170) for the alveolar stimulus, and at 1640 Hz (BW = 70) for the velar stimulus. The bursts were produced by exciting these formants with 5 ms of random noise, which began 5, 10, and 15 ms prior to voicing onset for /ba/, /da/, and /ga/, respectively. Therefore, total stimulus duration for the labial, alveolar, and velar stimuli was 265, 270, and 275 ms, respectively. The spectro-temporal specifications of the three FF stimuli are shown in the top of Fig. 2.

Because of differences in the two versions of the Klatt synthesizer used for stimulus construction, the amplitude of the burst relative to the vowel for the stimuli of the present experiment could not be set with the same synthesis parameter values as the Stevens and Blumstein stimuli. Linear prediction analysis was, therefore, employed to match spectral sections of the original Stevens and Blumstein stimuli with the waveforms of the stimuli of the present experiment and to thus determine the appropriate amplitudes for the frication of the bursts. The original Stevens and Blumstein stimuli were digitized from audiotapes provided by Professor Stevens, and SPECTRUM (see Kewley-Port, 1979), a general purpose digital signal processing program, was used for the spectral analysis. Because greater energy was observed in the region of F_4 and F_5 relative to F_1 in the steady-state vowel of the present stimuli, the tilt of the vowel was adjusted slightly by modifying the parameters controlling the bandwidth of the glottal antiresonator and the lip radiation characteristic (see Klatt, 1977).

Shown at the top of Fig. 3 are the onset spectra of the FF stimuli. These onset spectra were obtained using linear prediction analysis similar to that used by Blumstein and Stevens (1979). SPECTRUM was used to perform the analysis. Specifically, each stimulus waveform was pre-emphasized, windowed, and analyzed with 14 linear prediction coefficients using the autocorrelation method. The 25.6-ms window, which was positioned at stimulus onset, was an extended half-Hamming window; the first 12.8 ms of the window was rectangular, the following 12.8 ms was a half-Hamming window. This window, therefore, differed slightly from the extended half-Kaiser window employed by Stevens and Blumstein. Template fitting was achieved, for purposes of the present experiment, by first obtaining hardcopies from a Tektronix 4010-1 plotter of the calculated onset spectra and then visually comparing these hardcopies with transparencies of Stevens and Blumstein's templates. The onset spectrum of each FF stimulus is accepted by the appropriate Stevens and Blumstein template, and is rejected by the other two templates (for a detailed description of these templates and matching criteria, see Blumstein and Stevens, 1979).

2. Two-formant stimuli (TF)—Each of the TF stimuli was synthesized with the digital resonators connected in parallel. The parameter values of the first two formants of the FF stimuli were followed as closely as possible in constructing the corresponding labial, alveolar, and velar TF stimuli. (Of course, the TF stimuli did not have any upper formants, nor did they contain any release bursts.) Thus, the starting frequencies of F_1 and F_2 for the TF /ba/ and /da/ and the steady-state values for the formants and bandwidths of F_1 and F_2 for all of the TF stimuli were identical to those of their FF counterparts. Several modifications were made, however, to enable naive adult listeners to identify these stimuli with 90% or greater accuracy. Specifically, the starting frequency of F_2 in the velar stimulus was extended to 1940 Hz. In addition, the transition duration of F_1 for the alveolar stimulus was reduced to 20 ms. Finally, the formant amplitudes were altered so that the relative amplitudes of F_1 and F_2 in the FF and TF stimuli were the same. SPECTRUM was employed to set the amplitudes and obtain the best spectral matches between the TF stimuli and their appropriate FF counterparts. The TF stimuli were each 255 ms in duration. Their spectro-temporal specifications are shown at the bottom of Fig. 2.

The onset spectrum of each TF stimulus is shown in the bottom of Fig. 3. These onset spectra do not possess the global, invariant properties which Stevens and Blumstein have proposed to characterize different places of articulation and therefore the stimuli are not correctly and differentially classified by their spectral templates.

C. Procedure

A variant of the operant head-turning (OHT) procedure developed by Moore *et al.* (1977) was employed in the present experiment to measure discrimination of the FF and TF contrasts (see Aslin and Pisoni, 1980a). Basically, the procedure entails the repetitive presentation of a “background” stimulus that is interrupted by several presentations of a novel, “target” stimulus. The discriminative response is a unidirectional, 45°-head-turn toward a visual reinforcer (a motorized toy animal) and loudspeaker, which in our laboratory were located 90° to the infant’s left. Evidence of discrimination in this procedure consists of a head-turn on stimulus-change (experimental) trials and the absence of a head-turn on no-change (control) trials. The entire experimental procedure, including stimulus presentation, response recording and coding, and presentation of the visual reinforcer, was controlled online by a PDP-11/34 computer.

Each subject was tested on one of the FF or TF place of articulation contrasts in several experimental sessions conducted on successive days. Inside a single-walled sound-attenuated booth (IAC model 402), an assistant sat facing the infant seated on the parent’s lap. The assistant used small, silent toys to attract the gaze of the infant. The repeating background stimulus and target stimulus were presented to the infant with a constant interstimulus interval of 1200 ms over a single loudspeaker (Radio Shack MC-1000). Both the assistant and the parent listened to masking music (originating from a cassette tape deck located in the booth) over headphones during the entire experimental session to prevent any biasing of the infant’s responses (see Aslin and Pisoni, 1980b).

The procedure in each experimental session involved two phases: (1) shaping of the head-turn response and (2) criterion testing. In the shaping phase, the infant learned to make a directional head-turn in anticipation of the activation of the visual reinforcer whenever a change in the speech stimulus (i.e., from the continuously presented background stimulus to the target stimulus) occurred. The experimenter (located outside the booth) monitored the infant’s behavior via closed-circuit TV. When the infant’s gaze was steadily directed toward the toys being manipulated by the assistant, the experimenter initiated an experimental trial which consisted of three presentations of the target stimulus. If the infant responded to the stimulus change with a head-turn toward the speaker from which the speech stimuli were delivered, the experimenter immediately activated the visual reinforcer beside the speaker. During activation of the reinforcer, the inside of a smoked Plexiglas enclosure containing the toy animal was illuminated and the toy became animated (drummed, clapped, or jumped). If the infant did not spontaneously respond to the sound change on experimental trials, the visual reinforcer was presented to elicit a head-turn coincident with the speech stimulus change. In addition, the target stimulus was initially presented at 10 dB above the level of the background stimulus (70 dB–A scale) to facilitate shaping of the head-turn response. When the infant had responded correctly on two out of three trials for a given background/target intensity difference, the difference was attenuated by 5 dB. If the infant responded consistently to a 0 dB difference between the background and target stimuli, an immediate transition to the testing phase of the procedure was made. Throughout the shaping phase, the experimenter could hear the stimuli being presented to the infant and was therefore able to judge how well the infant was performing. A maximum of 30 shaping trials was typically given.

In the testing phase of the procedure, the experimenter continued to initiate trials whenever the infant's gaze was directed toward the toys being manipulated by the assistant in the booth. However, trials now consisted of two types: experimental (change) trials and control (no-change) trials. These two trial types were presented according to one of eight pseudorandom orders (Fellows, 1967) with the constraint that no more than two control trials were presented successively. The observation interval for scoring the infant's head-turn response began with the onset of the first target repetition and ended 2 s after the third (and final) target repetition (total scoring interval = 5.285 s). If the experimenter judged that a head-turn had occurred during the scoring interval on an experimental trial, a button was pressed and the computer immediately presented the visual reinforcer for 3 s. If the head-turn button was pressed on a control trial, the computer did not present the visual reinforcer. (Of course, if the experimenter judged that no head-turn occurred within the scoring interval, the head-turn button was not pushed.) During testing, the experimenter wore headphones through which a tone was presented in synchrony with the background and target stimuli to ensure that the experimenter was "blind" to the specific stimulus conditions and reinforcement contingencies on any given trial. The computer program controlling the experimental procedure kept a running tally of the infant's percentage of correct responses on the previous five experimental and five control trials. An infant was required to achieve 80% correct responding on a minimum of five experimental and five control trials in order to successfully complete testing on a contrast. This 80% criterion, applied to the five experimental and five control trials, results in a probability of 0.055, based on the binomial expansion, $p = 0.5$, $q = 0.5$, of falsely accepting the hypothesis that the infant was responding above chance.

If, after approximately 30 trials, an infant had not met discrimination criterion, but continued to show interest in the visual reinforcer and still appeared attentive, the experimenter terminated criterion testing and reinitiated the shaping phase of the procedure. Otherwise, the session was terminated. If, in a subsequent testing session, the infant again progressed beyond the shaping phase, discrimination testing was undertaken once more. If, after another 30 trials, the infant failed to meet criterion or appeared inattentive or bored with the visual reinforcer, the session was terminated. Infants who did not meet discrimination criterion within two sessions after the first session in which shaping was passed and who showed little interest in the reinforcer were dropped from the study. These infants were considered to have failed discrimination testing. Infants who met the minimum discrimination criterion were tested on an additional block of trials (either in the same session that criterion was first met or in a subsequent session) to determine whether or not they could meet this criterion a second time.

D. Design

Subjects were randomly assigned to one of six experimental groups, with the constraint that subjects were tested until six met discrimination criterion in each group. Groups 1, 2, and 3 were tested on the FF contrasts /ba-/da/, /ba-/ga/, and /da-/ga/, respectively. For a given contrast, the assignment of stimuli to target versus background status was randomly determined. Groups 4, 5, and 6 were tested on the TF contrasts /ba-/da/, /ba-/ga/, and /da-/ga/, respectively, and the assignment of the target and background stimuli was also randomly determined.

E. Results

In order to obtain six subjects who were able to meet the minimum discrimination criterion (80% correct responding on five experimental and five control trials in a row) for each of the FF and TF contrasts, it was necessary to test 32 infants in the FF condition and 37 infants in the TF condition. Thus, 56% and 49% of the infants initially assigned to the FF and TF

contrasts met this criterion. Twelve (38%) of the infants in the FF condition and 18 (49%) of the infants in the TF condition met this discrimination criterion for a second block of trials.² Table I shows the proportion of subjects who met discrimination criterion for each particular FF and TF contrast. The proportion of subjects who were able to meet criterion for a second block of trials is also shown in this table.

An examination of the mean proportion of correct responses by subjects on all trials prior to, but within the same session(s) that discrimination criterion was met (see Fig. 4), indicated that performance was above chance (according to two-tailed *t* tests) for all contrasts. For the first block of trials in which criterion was met, $t_{(5)} = 5.67, 4.67, \text{ and } 5.80$, for the FF contrasts /b–d/, /b–g/, and /d–g/, respectively; $t_{(5)} = 6.00, 4.60, \text{ and } 8.33$ for the same TF contrasts; $p < 0.01$ in all cases. For the second block of trials in which criterion was met, $t_{(4)} = 7.40, t_{(3)} = 3.63, t_{(2)} = 4.50$, for the FF contrasts /b–d/, /b–g/, and /d–g/; $p < 0.05$ in all cases. For the second block of trials in which criterion was met, $t_{(5)} = 5.17, 5.67, \text{ and } 6.20$ for the corresponding TF contrasts; $p < 0.01$ in all cases.

The results of a three-way repeated measures ANOVA (with formant and contrast as between-subjects factors and block as a within-subjects factor) on the proportion of correct responses indicated no main effect of formant condition or place contrast ($F_{(1,24)} = 0.04$ and $F_{(2,24)} = 0.89$, respectively; n.s.). Thus, there were no differences in performance by this measure either within or across stimulus conditions. For example, the performance of subjects tested on the FF /b/ vs /d/ contrast did not differ from that of subjects tested on the FF /b/ vs /g/ contrast or on the TF /b/ vs /d/ contrast. A main effect of block was observed ($F_{(1,24)} = 5.88$; $p < 0.05$), indicating that the proportion of correct responses did increase between the two blocks in which subjects met discrimination criterion. No interaction between this and the other two factors was obtained.

The number of trials required by subjects in each stimulus condition to meet discrimination criterion within a session (see Table II) were also submitted to a three-way (formant $\times$ contrast $\times$ block) repeated measures ANOVA. There was no main effect of formant or contrast ($F_{(1,24)} = 0.05$ and $F_{(2,24)} = 0.09$, respectively; n.s.) —indicating again that there were no significant differences in subjects' performance on the place contrasts within or across stimulus conditions. This analysis did reveal a main effect of block ($F_{(1,24)} = 13.15$; $p < 0.002$) and this factor did not interact significantly with any of the others.

Finally, the infants' performance on the FF and TF contrasts was assessed by examining the number of sessions the infants required to meet criterion (see Table III). (Note that this measure of discrimination performance includes trials from sessions not incorporated in the trials to criterion measure.) A three-way (formant $\times$ contrast $\times$ block) repeated measures ANOVA indicated no main effect of either formant or contrast with respect to the number of sessions required to meet discrimination criterion in each stimulus condition ($F_{(1,24)} = 0.01$ and $F_{(2,24)} = 0.10$, respectively; n.s.). However, a main effect of block was again obtained ($F_{(1,24)} = 87.69$; $p < 0.001$), indicating that having once met discrimination criterion, subjects required relatively fewer sessions to do so again. This factor did not interact with the formant and/or contrast conditions.

II. DISCUSSION

The results of the present experiment indicated that infants of about 6 months of age were able to discriminate two-formant labial versus alveolar, labial versus velar, and alveolar versus velar contrasts, as well as the corresponding five-formant + burst contrasts. Of the

²Three of the 18 infants who met discrimination criterion for a first block of trials could not be rescheduled for further testing.

infants tested on these two-formant contrasts, six infants in each group met discrimination criterion twice. Moreover, their discrimination performance was, like that of the infants tested on the five-formant contrasts, significantly above chance for all trials within the session(s) that discrimination criterion was met. No differences (in terms of proportion of correct responses, trials, or sessions to criterion) were observed in the performance of subjects either within or across the five-formant and two-formant conditions. Therefore, not only can two-formant place of articulation contrasts be discriminated by young infants, but they also appear to be *as* discriminable as the five-formant ones.

The failure to observe any differences in performance within and across the five-formant and two-formant conditions could, of course, be due to the insensitivity of the operant head-turning procedure and/or of our dependent measures. However, we did find that performance improved between the blocks in which discrimination criterion was met. In this respect, our data indicate that some learning is taking place, but, it occurs irrespective of the particular formant condition or place contrast. To the extent that this improvement in performance was observed both within and across subjects, the head-turning procedure and the measures we employed appear to have been sufficiently sensitive to have detected any existing differences in the discriminability of the contrasts. Furthermore, although the three measures we examined are not completely independent, they do provide converging evidence as to the lack of any difference in discriminability for the contrasts.

Because two-formant stop CV stimuli lack the global, context-independent spectral properties of their five-formant + burst counterparts and because they may be considered relatively acoustically impoverished, one might expect that infants would be less capable of discriminating two-formant place of articulation contrasts. Our failure then to observe any differences in discrimination performance for the two types of stimuli suggests that infants are not any more sensitive to global, invariant properties of the CV syllable onset spectrum than they are to local details of the waveform, such as formant transition information or even quite small spectral differences. This result suggests that Williams and Bush's (1978) conclusion concerning infants' superior discrimination of five-formant + burst stops relative to stops without bursts may have been inaccurate—particularly in light of the interpretative problems associated with their study as we noted earlier.

That infants do not appear to be any less able to discriminate two-formant versus five-formant + burst contrasts is an interesting finding. However, given the results of previous studies, our finding that infants *are* capable of discriminating two-formant CV syllables is not a surprising one nor is it inconsistent with the developmental aspects of Stevens and Blumstein's (1978) account of place of articulation perception. By this account, global, invariant properties of the CV syllables onset spectrum are primary for both adult perception and the immature perceptual system; context-dependent properties of stop consonants, such as formant transition information, constitute secondary cues to place of articulation that are learned through their co-occurrence with the primary ones. In order to learn which formant transition variations are associated with which primary cues and thus with place categories, infants must presumably be sensitive to differences in formant transitions. However, the fact that infants discriminate place contrasts that are signaled by "secondary" cues as well as they do those that are signaled by "primary" cues is consistent with the notion that the appropriate partitioning of the acoustic variations corresponding to different stop consonants can be achieved by virtue of the existence of natural psychophysical boundaries that are based on quite simple acoustic information—i.e., without learning co-occurrences between primary and secondary cues. In support of the existence of such boundaries is Kuhl and Padden's (1983) recent observation that macaques show enhanced discrimination at the (human) place category boundaries for a continuum of CV syllables varying only in the starting frequency of their second formant transitions.

Nevertheless, such a psychophysically based account cannot fully account for adults' perception of place of articulation at a phonetic level—particularly in light of the nondistinctiveness of formant transition information *within* certain vowel contexts in natural speech (e.g., Dorman *et al.*, 1977; Kewley-Port, 1982; Ohde and Sharf, 1977). Awareness of such nondistinctiveness, together with the contextual variability of formant transitions and release bursts, is what has recently prompted several researchers to seek new and invariant descriptions of the acoustic correlates of stop consonants (in addition to the work of Stevens and Blumstein, see Kewley-Port, 1980, 1982, 1983; Searle *et al.*, 1979, 1980). Such descriptions would seem to be required in order to account for the perceptual constancy of place of articulation in adults. Still, it is important to realize that only distinctive and not invariant descriptions are necessary to account for the *infant's* perception of place of articulation contrasts. Although there may be theoretical arguments for the necessity of invariant descriptions, there is, at present, a lack of any convincing empirical evidence for the perceptual constancy of stop consonants across vowel context in infants (i.e., under the conditions in which stops exhibit their greatest variability). For example, in a recent study, Katz and Jusczyk (1980) used a version of the OHT procedure in an attempt to determine whether or not infants could reliably discriminate the stops /b/ and /d/ in the face of variations across four different vowel contexts. These investigators found that infants were only able to so “categorize” the stops across two vowel contexts. When the vowel set was expanded, the infants performed at chance.

In summary, the results of the present study indicate that young infants discriminate two-formant CV syllables differing in place of articulation as well as they do corresponding five-formant + burst place contrasts. Thus, any global, invariant spectral properties or additional cues contained in the latter stimuli do not appear to provide any basis for enhanced discriminative performance. Rather, it appears that the presence of more simple acoustic cues is sufficient to mediate discrimination.

Acknowledgments

This research was supported in part by NIMH research grant MH-24027 and NIH research grants NS-12179 and HD-11915 to Indiana University, Bloomington and by a predoctoral fellowship awarded to the first author by the Research Council of Canada. A preliminary report of this research was presented at the 100th Meeting of the Acoustical Society of America in Los Angeles, California in November, 1980. We wish to thank Tom Carrell, Cathy Kubaska, and Howard Nusbaum, as well as our reviewers, for their many helpful comments on this work. We also thank Diane Kewley-Port for her assistance with the synthesis of the speech stimuli, Jerry Forshee and Dave Link for their technical assistance, and Kathy Mitchell for her help with the data collection process. In addition, we thank Professor K. Stevens for providing audio recordings of the stimuli on which we modeled those used in the present study.

References

- Aslin, RN.; Pisoni, DB. Some developmental processes in speech perception. In: Yeni-Komishan, G.; Kavanagh, J.; Ferguson, CA., editors. *Child Phonology: Perception*. Vol. 1. Academic; New York: 1980a. p. 67-96.
- Aslin RN, Pisoni DB. Effects of early linguistic experience on speech discrimination by infants: A critique of Eilers, Gavin, and Wilson (1979). *Child Dev.* 1980b; 51:107–112. [PubMed: 7363730]
- Aslin, RN.; Pisoni, DB.; Jusczyk, PW. Auditory development and speech perception in infancy. In: Haith, MM.; Campos, JJ.; Mussen, PH., editors. *Infancy and Developmental Psychobiology*, Vol. 2 of Carmichael's Manual of Child Psychology. 4. Wiley; New York: 1983. p. 573-687.
- Blumstein SE, Stevens KN. Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *J Acoust Soc Am.* 1979; 66:1001–1017. [PubMed: 512211]
- Blumstein SE, Stevens KN. Perceptual invariance and onset spectra for stop consonants in different vowel environments. *J Acoust Soc Am.* 1980; 67:648–662. [PubMed: 7358906]

- Cooper FS, Delattre PC, Liberman AM, Borst JM, Gerstman LJ. Some experiments on the perception of synthetic speech sounds. *J Acoust Soc Am.* 1952; 24:597–606.
- Delattre P, Liberman AM, Cooper FS. Acoustic loci and transitional cues for consonants. *J Acoust Soc Am.* 1955; 27:769–773.
- Dorman MF, Studdert-Kennedy M, Raphael LJ. Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Percept Psychophys.* 1977; 22:109–122.
- Eimas PD. Auditory and linguistic processing of cues for place of articulation by infants. *Percept Psychophys.* 1974; 16:513–521.
- Fellows BJ. Chance stimulus sequences for discrimination tasks. *Psychol Bull.* 1967; 67:87–92. [PubMed: 6045339]
- Fodor JA, Garrett MF, Brill SL. Pi ka pu: The perception of speech sounds by prelinguistic infants. *Percept Psychophys.* 1975; 18:74–78.
- Katz, J.; Jusczyk, PW. Do six-month-olds have perceptual constancy for phonetic segments?. paper presented at the International Conference on Infant Studies; New Haven. April, 1980; 1980.
- Kewley-Port, D. Res Speech Percept: Prog Rep No 4. Dept. of Psychol., Indiana University; 1978. KLTEXC: Executive program to implement the KLATT software speech synthesizer; p. 235-245.
- Kewley-Port, D. Res Speech Percept: Prog Rep No 5. Dept. of Psychol., Indiana University; 1979. SPECTRUM: A program for analyzing the spectral properties of speech; p. 475-492.
- Kewley-Port, D. Res Speech Percept: Tech Rep No 3. Dept. of Psychol., Indiana University; 1980. Representations Of spectral change as cues to place of articulation in stop consonants; p. 263
- Kewley-Port D. Measurement of formant transitions in naturally produced stop consonant-vowel syllables. *J Acoust Soc Am.* 1982; 72:379–389. [PubMed: 7119280]
- Kewley-Port D. Time-varying features as correlates of place of articulation in stop consonants. *J Acoust Soc Am.* 1983; 73:322–335. [PubMed: 6826902]
- Klatt DH. A cascade-parallel terminal analog speech synthesizer and a strategy for consonant-vowel synthesis. *J Acoust Soc Am Suppl 1.* 1977; 61:S68.
- Klatt DH. Software for a cascade/parallel formant synthesizer. *J Acoust Soc Am.* 1980; 67:971–995.
- Kuhl PK, Padden DM. Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *J Acoust Soc Am.* 1983; 73:1003–1010. [PubMed: 6221040]
- Leavitt LA, Brown JA, Morse PA, Graham FK. Cardiac orienting and auditory discrimination in 6-week infants. *Develop Psychol.* 1976; 12:514–523.
- Miller CL, Morse PA. The ‘heart’ of categorical speech discrimination in young infants. *J Speech Hear Res.* 1976; 19:578–589. [PubMed: 979219]
- Moffitt AR. Consonant cue perception by twenty-four-week-old infants. *Child Dev.* 1971; 42:717–731. [PubMed: 5121099]
- Moore JM, Wilson WR, Thompson G. Visual reinforcement of head-turn responses in infants under twelve months of age. *J Speech Hear Disord.* 1977; 42:328–334. [PubMed: 881813]
- Morse PA. The discrimination of speech and nonspeech stimuli in early infancy. *J Exp Child Psychol.* 1972; 14:477–492. [PubMed: 4658485]
- Ohde RN, Sharf DJ. Order effect of acoustic segments of VC and CV syllables on stop and vowel identification. *J Speech Hear Res.* 1977; 20:543–554. [PubMed: 904315]
- Searle, CL.; Jacobson, JZ.; Kimberly, BP. Speech patterns in the 3-space of time and frequency. In: Cole, RA., editor. *Perception and Production of Fluent Speech.* Erlbaum; Hillsdale, NJ: 1980. p. 73-102.
- Searle CL, Jacobson JZ, Rayment SG. Stop consonant discrimination based on human audition. *J Acoust Soc Am.* 1979; 65:799–809. [PubMed: 447910]
- Stevens KN, Blumstein SE. Invariant cues for place of articulation in stop consonants. *J Acoust Soc Am.* 1978; 64:1358–1368. [PubMed: 744836]
- Stevens, KN.; Blumstein, SE. The search for invariant acoustic correlates of phonetic features. In: Eimas, PD.; Miller, J., editors. *Perspectives on the Study of Speech.* Erlbaum; Hillsdale, NJ: 1981. p. 1-38.

- Till, JA. Infants' discrimination of speech and nonspeech stimuli. paper presented at the Annual meeting of the American Speech and Hearing Association; Houston, Texas. November 1976; 1976.
- Williams M, Bush L. Discrimination by young infants of voiced stop consonants with and without release bursts. *J Acoust Soc Am*. 1978; 63:1223–1226.

\$watermark-text

\$watermark-text

\$watermark-text

FIG. 1.
Theoretical onset spectra of labial, alveolar, and velar stop consonants (after Stevens and Blunstein, 1978).

FIG. 2.
Spectro-temporal specifications of the five-formant + burst and the two-formant stimuli.

FIG. 3. Onset spectra of the five-formant + burst and the two-formant stimuli. Note that the amplitude scale in this figure differs from that in Fig. 1.

FIG. 4.

Group data for infants tested on each of the five-formant + burst and the two-formant place of articulation contrast. Shown is the mean proportion of correct headturns for all trials prior to the point at which criterion was met for two blocks of trials.

\$watermark-text

\$watermark-text

\$watermark-text

TABLE I

Percentage of subjects who met discrimination criterion.

Contrast	FF stimuli (N = 32)			TF stimuli (N = 37)		
	b/d	b/g	d/g	b/d	b/g	d/g
Block 1	67% (6/9)	67% (6/9)	43% (6/14)	40% (6/15)	75% (6/8)	43% (6/14)
Block 2	56% (5/9)	44% (4/9)	21% (3/14)	40% (6/15)	75% (6/8)	43% (6/14)

TABLE II

Mean trials to criterion.

Contrast	FF stimuli			TF stimuli		
	b/d	b/g	d/g	b/d	b/g	d/g
Block 1	17.33	17.33	16.17	19.50	20.33	18.00
Block 2	13.20	15.25	14.67	12.50	14.67	13.50

TABLE III

Mean sessions to criterion.

Contrast	FF stimuli			TF stimuli		
	b/d	b/g	d/g	b/d	b/g	d/g
Block 1	2.80	2.75	2.33	3.17	2.17	2.77
Block 2	0.80	1.00	0.33	1.67	1.92	1.67