The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants

Sheila E. Blumstein, Ellyn Isaacs, and John Mertus
Brown University, Department of Linguistics, Providence, Rhode Island 02912

(Received 10 December 1981; accepted for publication 9 April 1982)

This series of studies explored the extent to which the gross shape of the onset spectrum is used by the listener for the identification of place of articulation in initial stop consonants. Synthetic stimuli were generated with onset formant frequencies appropriate to the syllables [bɑ bi bu dɑ di du] and with the gross shape of the onset spectrum manipulated to be appropriate for either alveolar consonants or labial consonants. Stimuli were presented for identification and discrimination. In addition, adaptation effects of stimuli containing appropriate frequency and shape and incompatible frequency and shape were explored on a place-of-articulation onset continuum. Although identification performance was determined by onset frequency rather than gross shape of the spectrum, presentation of stimuli in which shape was inconsistent with frequency reduced identification performance. Further, subjects could discriminate stimuli which varied only in spectral shape. Finally, significantly less adaptation was found for a [dɑ] onset with a labial spectrum shape than a [dɑ] onset with an alveolar spectrum shape. These results suggest that although the invariant properties residing in the gross shape of the onset spectrum may serve as a classificatory framework for the phonetic dimensions of natural language, they may not provide the primary perceptual attributes for place of articulation in ongoing speech processing.

PACS numbers: 43.70.Dn, 43.70.Ve

INTRODUCTION

There are three principal components to recent claims that there is acoustic invariance for place of articulation in initial stop consonants (Stevens and Blumstein, 1977; Blumstein and Stevens, 1981). The first is that there are invariant properties which can be derived directly from the acoustic signal, and these properties correspond to the phonetic dimensions found in natural language; the second is that the invariant properties for place of articulation reside in the first 20-odd ms of the signal; and the third is that the invariant property for place of articulation is the gross shape of the spectrum at stimulus onset. Data from both the acoustic analysis of natural speech and results from speech perception experiments have been used to support this view. In particular, spectral analyses of the onset characteristics of natural speech utterances have indicated that labial consonants can be characterized by the diffuse-falling or diffuse-flat property, alveolar consonants by the diffuse-rising property, and velar consonants by the compact property, and that these derived properties are found independent of the particular vowel context in which the utterance occurs. These results were obtained using a relatively static representation based on the gross shape of the spectrum in a single time frame of analysis of 25.6-ms duration (Blumstein and Stevens, 1979), as well as by using a time-varying representation looking at the spectrum over a longer time frame or window of analysis (Searle *et al.*, 1979; Searle *et al.*, 1980; Kewley-Port, 1980).

Perception experiments with natural speech as well as synthetic speech stimuli have shown that listeners can identify place of articulation in stimuli containing only the release burst and the first few tens of milliseconds of glottal pulsing. Although performance of subjects improves as longer durations of the stimuli are presented (Blumstein and Stevens, 1980; Kewley-Port, 1980), performance is well above chance even with stimuli as short as 20 or so ms in duration. Blumstein and Stevens (1980) hypothesized that the perceptual decisions of the subjects for place of articulation are in fact based on the extraction of the invariant properties relating to the gross shape of the onset spectrum. However, in all of these studies, not only did the spectral shape of the stimuli vary, but so did the onset frequencies of the vowel formants. That is, each stimulus contained the appropriate invariant property as well as formant frequency characteristics associated with its particular place of articulation and vowel environment. Because both shape and frequency were always associated, it is not clear whether the subjects' responses were in fact based upon the gross shape of the onset spectrum or upon the formant-frequency characteristics of the stimuli.

It was the object of this research to determine the extent to which the invariant properties residing in the gross shape of the onset spectrum contribute to the perception of the place-of-articulation phonetic dimension. To this end, we synthesized brief stimuli containing onset formant frequencies appropriate to the syllables [bɑ bi bu dɑ di du] in which the formant transitions of the second and higher formants were constant throughout, and manipulated the onset spectrum to be appropriate for either alveolar consonants (diffuse-rising) or labial consonants (diffuse-falling). The obtained stimuli then either contained both onset formant frequencies and shapes that were appropriate to a given consonant in a particular vowel environment, or they were appropriate in frequency but inappropriate in shape. Because the higher formant transitions were steady throughout, we

 0001-4966/82/070043-08$00.80

TABLE I. Onset frequencies of the formant transitions (in hertz) for the stimuli in experiment I.

	$F1$	$F2$	$F3$	$F4$
[bɑ]	200	900	2000	3300
[bi]	180	1800	2600	3200
[bu]	180	800	2000	3200
[dɑ]	200	1700	2800	3500
[di]	180	2000	2800	3900
[du]	180	1600	2700	3200

were able to assess the perceptual effects of onset formant frequencies and spectral shape independent of formant motions. In experiment I, these stimuli were presented to listeners for both identification and discrimination. If it is the case that the gross shape of the onset spectrum provides invariant properties corresponding to the phonetic dimensions for place of articulation, then subjects should perceive all stimuli containing the diffuse-rising property as alveolar consonants and all stimuli containing the diffuse-falling property as labial consonants, regardless of the onset formant frequencies of the stimuli.

I. EXPERIMENT I

A. Methods

Ten subjects participated in this experiment. All were native English speakers and had no known hearing impairment.

Stimuli were generated using a computer simulation of a terminal analog speech synthesizer in which the tuned circuits for the vowel-generating portion were connected in parallel, and thus individual control of formant amplitudes could be effected (Klatt, 1980). The sampling rate for the synthesizer output was 10 kHz and the output was low-pass filtered with a cutoff frequency of about 4500 Hz. All stimuli were of 40-ms duration and contained five formant transitions for which the onset frequencies of the formants were appropriate to the consonants [b d] in the environment of the vowels [ɑ i u]. The frequencies of the second and higher formants were constant throughout. Table I shows the onset frequencies of the formant transitions (in hertz) for the first four formants. The fifth formant was not varied and was set at a constant 3850 Hz for all stimuli. The movement of $F1$ was retained in order to maintain the stoplike quality of the stimuli. Its onset frequency was appropriate to the consonant [b] or [d] (see Table I), and the transition moved in 20 ms to frequencies appropriate to the vowels [ɑ] at 720 Hz, [i] at 330 Hz, and [u] at 370 Hz. The fundamental frequency for all stimuli started at 103 Hz and rose linearly over the 40-ms duration to 125 Hz. Both the amplitudes and the bandwidths of the formants were then manipulated in such a way that for each of the six formant frequency configurations ([b d] in the environment of [ɑ i u]), the resulting spectrum at stimulus onset contained either the diffuse-rising property characteristic of alveolar consonants or the diffuse-falling property characteristic of labial consonants. Table II shows the input bandwidth and amplitude characteristics of the five formants used for the synthesis of the 12 stimuli. Thus there was a total of 12 stimuli, six in which both formant frequencies and shapes were appropriate to a particular place of articulation in a given vowel context, and six in which the onset frequencies were appropriate but the spectral shape was inappropriate. Figure 1 shows the short-time spectra of the stimuli containing onset frequencies appropriate to [b] in the environment of [ɑ i u] with the appropriate diffuse-falling and inappropriate diffuse-rising spectral shape. These spectra were derived by LPC analysis obtained by pre-emphasizing the high frequencies and using a 25.6-ms time window beginning at the consonantal release. Figure 2 shows the short-time spectra of the stimuli containing onset frequencies appropriate to [d] in the environment of [ɑ i u] with the appropriate diffuse-rising shape and inappropriate diffuse-falling shape.

In order to insure that the spectral shapes of the synthesized stimuli had configurations similar to those found in natural speech, the onset spectra of the stimuli were fitted to the diffuse-falling and diffuse-rising templates originally developed by Blumstein and Stevens (1979). These templates were constructed to provide an objective measure for characterizing the hypothesized invariant properties based on the

TABLE II. Bandwidth (in hertz) and amplitude (in decibels) characteristics of the five formants for the 12 test stimuli used in experiment I. The first two letters of the stimulus name indicate which syllable the formant frequencies correspond to ([bɑ bi bu dɑ di du]), and the last letter indicates what spectral shape the stimulus contained (B, diffuse-falling tilt or D, diffuse-rising tilt).

	$F1$		$F2$		$F3$		$F4$		$F5$	
	BW	A	BW	A	BW	A	BW	A	BW	A
BAB	50	50	70	50	90	50	110	35	150	40
BIB	50	70	70	70	110	60	170	50	250	70
BUB	50	65	70	72	110	75	170	55	250	30
BAD	190	32	110	30	60	40	140	40	90	40
BID	50	70	70	70	110	70	170	75	250	70
BUD	30	65	50	70	110	85	170	85	250	30
DAD	50	63	50	54	110	55	170	64	250	77
DID	50	70	70	65	110	65	170	65	250	60
DUD	50	70	70	65	110	60	170	65	250	60
DAB	50	70	50	75	110	58	170	53	250	40
DIB	50	70	70	70	110	52	170	40	250	40
DUB	50	75	70	85	110	70	170	65	250	30

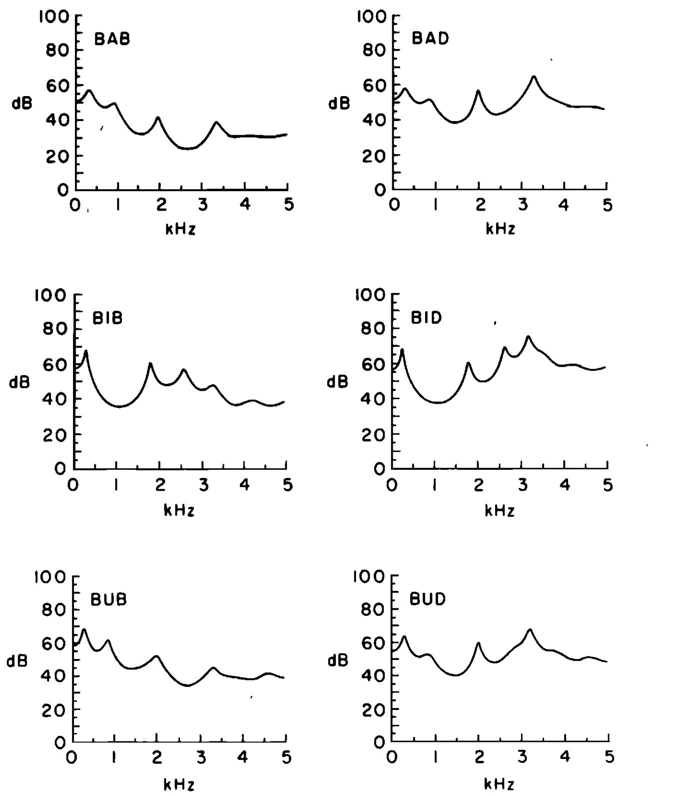

FIG. 1. Onset spectra for the test stimuli containing onset formant frequencies for [b] in the context of the vowels [ɑ], [i], and [u]. On the left are the spectra for those stimuli containing the appropriate diffuse-falling shape (BAB, BIB, BUB), and on the right are the spectra for those stimuli containing the inappropriate diffuse-rising shape (BAD, BID, BUD). These spectra were derived by means of LPC analysis using a half-Hamming window of 25.6 ms.

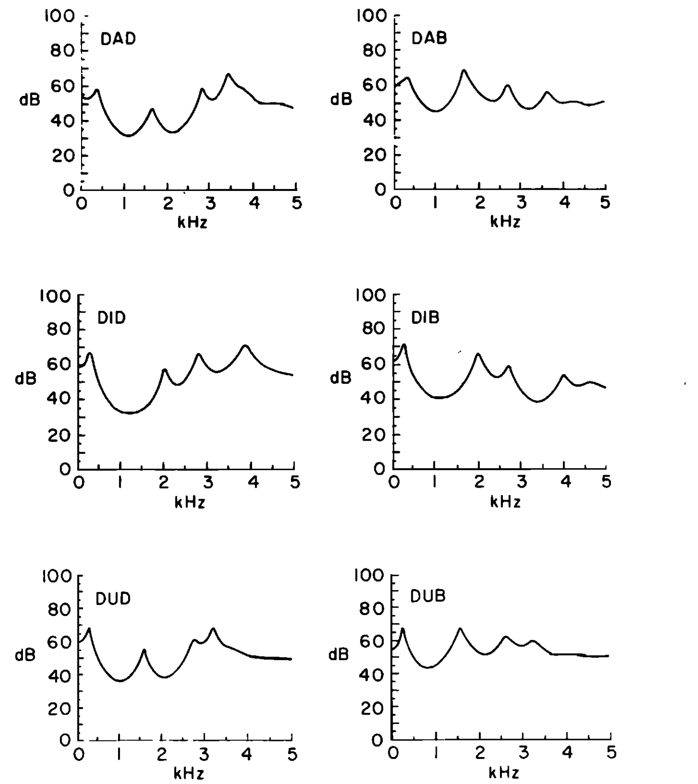

FIG. 2. Onset spectra for the test stimuli containing onset formant frequencies for [d] in the context of the vowels [ɑ], [i], and [u]. On the left are the spectra for those stimuli containing the appropriate diffuse-rising shape (DAD, DID, DUD), and on the right are the spectra for those stimuli containing the inappropriate diffuse-falling shape (DAB, DIB, DUB). These spectra were derived by means of LPC analysis using a half-Hamming window of 25.6 ms.

gross shape of the onset spectrum for place of articulation in natural speech utterances. All of the spectra of the test stimuli containing the diffuse-rising shape fit the diffuse-rising template, as did all of the spectra containing the diffuse-falling shape fit the diffuse-falling template.

Two test tapes were constructed using these stimuli—a labeling tape and a discrimination tape. The labeling tape consisted of the random presentation of ten occurrences of each of the 12 test stimuli, with a 4-s interstimulus interval and a 6-s silent interval after each block of 12 stimuli.

The discrimination (AX) tape consisted of 120 stimulus pairs, each separated by 2 s of silence with a 4-s intertrial interval. Sixty of the trials were DIFFERENT trials. These trials consisted of the random presentation of ten occurrences of each of the stimulus pairs containing the same onset formant frequencies but different spectral shapes, e.g., [bɑ] onset frequencies, diffuse-falling, or [b] spectrum shape called BAB versus [bɑ] onset frequencies, diffuse-rising, or [d] spectrum shape called BAD. The order of the members of the pairs (i.e., BAB–BAD or BAD–BAB) was also randomized across the stimulus pairs. Sixty trials were SAME trials and consisted of the random presentation of five occurrences of each of the 12 stimuli paired with itself.

Subjects were tested in groups of no larger than four. All were tested in one session and received the discrimination task followed by the labeling task. Stimuli were presented to subjects via headphones at a comfortable listening level. The perceptual quality of the test stimuli was that of a short syllable containing an initial stop followed by a very short and reduced vowel. No subjects reported that these stimuli did not sound speechlike. For the labeling test, subjects were required to identify the initial consonant as b or d and to write down their responses on the answer sheet provided. Prior to the administration of the test, subjects were presented a block of the 12 stimuli to familiarize themselves with the stimuli to be identified. They were instructed to simply listen to the stimuli, as these constituted the stimulus set, and to note the pace with which they were presented, as the test stimuli would occur at the same rate. The subjects were given no feedback on the identification of any particular stimulus, although they were told that they would hear a series of [b] stimuli followed by a series of [d] stimuli. For discrimination, subjects were required to compare each stimulus pair and write down on the answer sheet provided S if they were the same and D if they were different. Stimuli belonging to the same phonetic category but having a different quality, as well as stimuli belonging to different phonetic categories, were both to be considered different.

B. Results and discussion

Figure 3 shows the analysis of performance on the labeling task. The left panel corresponds to the number of [b]

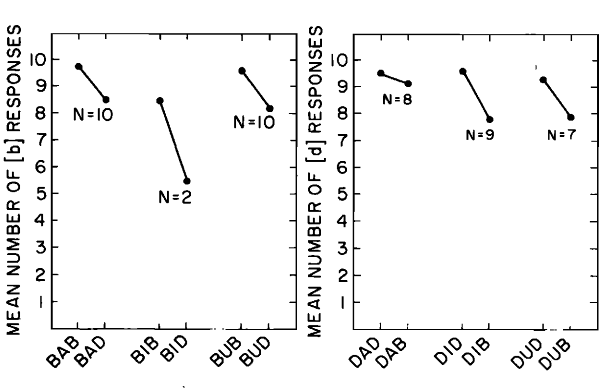

FIG. 3. Analysis of performance on the labeling test. The left panel corresponds to the number of [b] responses for stimuli in the context of the vowels [ɑ], [i], [u] with either the appropriate diffuse-falling ([b]) spectrum shape (BAB, BIB, BUB, respectively), or inappropriate diffuse-rising ([d]) spectrum shape (BAD, BID, BUD, respectively). The right panel corresponds to the number of [d] responses for stimuli in the context of the vowels [ɑ], [i], [u] with either the appropriate diffuse-rising ([d]) spectrum shape (DAD, DID, DUD, respectively), or inappropriate diffuse-falling ([b]) spectrum shape (DAB, DIB, DUB, respectively). N corresponds to the number of subjects who scored at least 70% on those stimuli containing appropriate frequencies and shape (i.e., BAB, BIB, BUB, DAD, DID, DUD).

responses for stimuli containing [b] frequencies and either appropriate diffuse-falling ([b]) spectrum tilt, or inappropriate diffuse-rising ([d]) spectrum tilt, and the right panel corresponds to the number of [d] responses for stimuli containing [d] frequencies and either appropriate diffuse-rising ([d]) spectrum tilt, or inappropriate diffuse-falling ([b]) spectrum tilt. Because we were interested in the *change* in response as a result of the manipulation of spectrum tilt, we included only those scores for subjects in which they scored at least 70% on those stimuli containing appropriate frequencies and shape (i.e., BAB, BIB, BUB, DAD, DID, DUD). Thus we insured that they could reliably identify these brief stimuli when both onset formant frequencies and shape were compatible. All ten subjects met this criterion for the BAB and BUB stimuli, nine did for DID, eight for DAD, seven for DUD, and only two for BIB (see Fig. 3). As the figure shows, when spectrum tilt is not compatible with the onset formant frequencies, the number of [b] and [d] responses is reduced. To determine if this reduction in responses was significant, two *t*-tests were conducted, one for the [b]-frequency stimuli and one for the [d]-frequency stimuli. Scores were included only for those subjects who responded at least 70% to *all* of the stimuli in either the [b] ([bɑ bi bu]) or [d] ([dɑ di du]) categories that contained the appropriate shape. Because so few subjects were able to identify the [bi] stimulus with the appropriate diffuse-falling shape, this category was eliminated from the data analysis. Results showed a significant change in responses for both the [b] stimuli ($t=2.79$, $df=9$, $p<0.05$) and the [d] stimuli ($t=2.94$, $df=5$, $p<0.05$). Nevertheless, as the figure shows, with the exception of the [bi] frequency stimuli (BIB and BID), phonetic category does *not* change with manipulation of spectrum tilt. That is, the majority of responses still corresponds to the onset formant frequencies appropriate to the particular phonetic category. Thus the perceptual decisions of subjects although affected by spectrum tilt are not

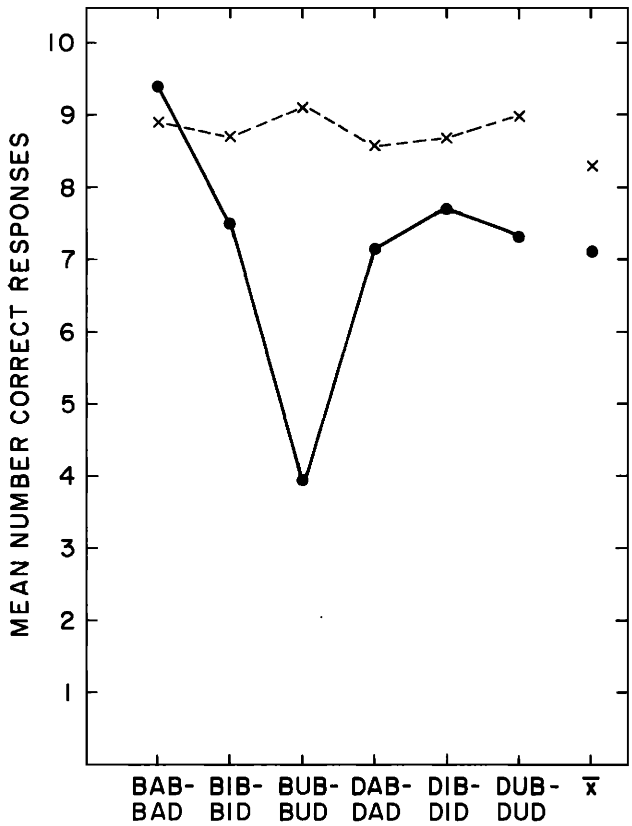

FIG. 4. Results of the discrimination task. The filled circles represent correct responses to DIFFERENT stimuli (BAB–BAD, BIB–BID, etc.). The crosses represent correct responses to pairs of SAME stimuli. Each point represents the mean SAME responses for the stimuli listed on the abscissa, e.g., the first point corresponds to the mean correct responses for BAB–BAB and BAD–BAD.

directed by it.

Figure 4 shows the results of the discrimination task. The filled circles represent correct responses to DIFFERENT stimuli and the crosses represent correct responses to SAME stimuli. With the exception of the [bu] stimuli, subjects clearly can discriminate the stimuli on the basis of spectrum tilt. Results of a d′ analysis shown in Table III indicate that with the exception of the [bu] stimuli, subjects' performance represents a reliable shift in stimulus discriminability. Thus subjects are able to discriminate contrasts based on the manipulation of spectrum tilt, although they may not perceive the stimuli as belonging to different phonetic categories.

II. EXPERIMENT

Results of experiment I indicate that while onset formant frequency was the dominant factor in determining the

TABLE III. d′ analysis of discrimination of spectrum tilt.

BAB–BAD	BIB–BID	BUB–BUD
3.64	2.58	1.47
DAD–DAB	DID–DIB	DUD–DUB
2.64	2.38	2.73

phonetic category to which the subject assigned a particular stimulus, spectral shape was playing some role in the perception process. In particular, the fact that subjects' identification responses were affected by the manipulation of spectrum tilt, and the fact that subjects could reliably discriminate between two stimuli with the same frequency characteristics but different shape configurations suggested that listeners are sensitive to changes in spectral shape. Nevertheless, it is not clear from experiment I to what extent listeners call on shape information to make decisions about the place-of-articulation phonetic dimension.

It was the object of experiment II to explore this question. To this end, we needed to use a paradigm that is sensitive enough to show the perceptual effects of subtle differences in the acoustic structure of the stimulus on subjects' categorization of the phonetic dimensions of speech. The selective adaptation paradigm seems to provide such a measure. Using this method, it has been shown that with regard to place of articulation, the perceptual system is sensitive to changes in onset frequency (Bailey, 1975; Sawusch, 1977), frequency motions (Landahl and Blumstein, 1982), and vowel contexts (Cooper, 1974; Miller and Eimas, 1976; Sawusch and Pisoni, 1978). Further, adaptation effects seem to be based primarily upon the acoustic attributes in the signal, rather than on the phonetic label given the stimulus by the subject (Blumstein *et al.*, 1977). Thus it should be possible to determine the effects of spectrum tilt on the perception of place of articulation despite the fact that stimuli containing conflicting frequency and spectrum tilt information are labeled by subjects according to their frequency characteristics.

In experiment II, we compared the effects of three adapting stimuli similar to those used in experiment I on a [bɑ–dɑ] place-of-articulation continuum. All the stimuli were similar to those used in experiment I. One adapting stimulus contained the onset frequency values associated with [dɑ] and a diffuse-rising or [d] shape, (DAD). The second also contained onset frequency values for [dɑ], but the spectrum tilt was manipulated to contain the diffuse-falling or [b] shape (DAB). The third maintained the diffuse-rising shape but the onset frequency values were scaled up relative to the DAD stimulus (HDAD). If spectral shape contributes to phonetic decisions concerning place of articulation, then one of two results should emerge: (1) adaptation with DAB should show significant [b] adaptation, whereas adaptation with DAD should show significant [d] adaptation. Such results would indicate that the acoustic information residing in the spectral shape is directing the phonetic decision of the subject; (2) in contrast to DAD, adaptation with DAB should produce no significant adaptation effects. Such results would suggest that both onset frequency and spectral shape are contributing to the phonetic decision for place of articulation, and the net effect of pitting frequencies appropriate to [d] against shape appropriate to [b] in an adaptation stimulus is no adaptation on a [b–d] continuum. The last adaptor (HDAD) was included to contrast the effect of manipulating frequency while holding spectrum tilt constant (DAD versus HDAD) to manipulating spectrum tilt while holding frequency constant (DAD versus DAB). It is generally the case that less adaptation is obtained when the adapting and test stimuli are not acoustically isomorphic than when they are (cf. Landahl and Blumstein, 1982, for discussion). Therefore, because HDAD is not acoustically isomorphic with the endpoint stimulus on the test continuum but its acoustic parameters (i.e., frequency and shape) are appropriate for an alveolar consonant, significant [d] adaptation should be obtained, but the magnitude of the effect should be significantly less than adaptation with DAD (which is acoustically isomorphic with an endpoint stimulus on the test continuum). In sum, it is hypothesized that with manipulation of frequency (HDAD), significant adaptation effects would be obtained, whereas with manipulation of shape (DAB), either the direction of adaptation would be altered or no significant adaptation effects would be obtained.

A. Methods

Eight undergraduates at Brown University served as subjects. All were native speakers of English and had no hearing impairments. One subject had previous experience listening to synthesized speech. The subjects were culled from a group of 20 listeners who were screened to identify a synthetic speech [bɑ–dɑ] continuum containing brief (40 ms) stimuli in terms of two distinct phonetic categories. Of the 20 listeners, 12 identified the test continuum in a categorical-like fashion, eight were unable to label the endpoints consistently, nor did they show evidence that they perceived two distinct phonetic categories along the test continuum. For the purposes of experiment II, we chose eight subjects from the 12 who showed categorical-like functions.

Stimuli consisted of a test continuum in which the parameter values of the endpoint stimuli were appropriate in frequency and spectral shape for [bɑ] and [dɑ], respectively. These values were the same as stimuli BAB and DAD of experiment I. The intermediate stimuli were generated by interpolating between these onset formant frequencies and formant amplitude values. Thus a continuum was produced in which both onset frequencies and spectral shapes systematically changed from a [bɑ] diffuse-falling stimulus to a [dɑ] diffuse-rising stimulus. Table IV shows the onset formant frequency values as well as the amplitudes and bandwidths of the formants for the eight stimuli on the test continuum. Previous research has shown that such a continuum containing brief stimuli produces identification functions with well defined categories (Blumstein and Stevens, 1980), and further, these brief stimuli can adapt a place-of-articulation continuum in a manner similar to those obtained for place of articulation continuum in CV syllables (Landahl and Blumstein, 1982). The three adapting stimuli consisted of (1) stimulus DAD, the [d]-endpoint stimulus on the test continuum; (2) stimulus DAB from experiment I. This stimulus contained the same onset frequencies as DAD, but the amplitudes and bandwidths of the formants were manipulated to contain the diffuse-falling spectral shape appropriate to labial consonants; and (3) stimulus HDAD, containing frequency values scaled up 10% relative to DAD. The onset frequency values for this stimulus were then 220 ($F1$), 1870 ($F2$), 3080 ($F3$), 3850 ($F4$), and 4235 ($F5$). The spectrum tilt for this stimulus, as for stimulus DAD, was diffuse-rising.

TABLE IV. Onset frequencies of the formant transitions (in hertz) as well as bandwidths (in hertz) and amplitudes (in decibels) of the formants for the stimuli comprising the [ba–ba] continuum used in experiment II.

Stimulus	$F1$ Freq. (BW, A)	$F2$ Freq. (BW, A)	$F3$ Freq. (BW, A)	$F4$ Freq. (BW, A)	$F5$ Freq. (BW, A)
1 (BAB)	200 (50, 50)	900 (70, 50)	2000 (90, 50)	3500 (110, 35)	3850 (150, 40)
2	200 (50, 52)	1010 (68, 51)	2110 (93, 51)	3500 (118, 39)	3850 (165, 46)
3	200 (50, 53)	1120 (65, 51)	2220 (96, 52)	3500 (127, 43)	3850 (179, 51)
4	200 (50, 55)	1240 (62, 52)	2340 (99, 52)	3500 (135, 47)	3850 (193, 56)
5	200 (50, 57)	1350 (59, 52)	2450 (102, 53)	3500 (144, 51)	3850 (207, 61)
6	200 (50, 59)	1470 (56, 53)	2570 (105, 53)	3500 (152, 55)	3850 (221, 66)
7	200 (50, 61)	1580 (53, 53)	2680 (108, 54)	3500 (161, 59)	3850 (235, 71)
8 (DAD)	200 (50, 63)	1700 (50, 54)	2800 (110, 55)	3500 (170, 64)	3850 (250, 77)

All subjects attended three 1-h sessions, one for each adapting stimulus. The order of presentation of the adaptation conditions was varied across the subjects. Before the first of the three sessions only, subjects listened to the continuum twice in order to familiarize them with the stimuli. Subjects were told that the first few stimuli should sound like [ba]'s and the last few like [da]'s, and that they should hear a sudden shift from one phonetic category to the other somewhere within the stimulus set. Subjects commented that the stimuli sounded unusual but none seemed to have trouble perceiving them as speechlike sounds. In all three sessions, subjects were presented a baseline tape which consisted of 80 stimuli including the random presentation of ten occurrences of each of the eight stimuli on the continuum. Three seconds separated each stimulus with 5 s after each ten stimuli. Subjects listened to the stimuli binaurally using headphones and were asked to identify each sound as either [b] or [d] on the sheet provided. After a brief rest, subjects were presented with one of the adaptation tapes. Each tape began with 2.5 min or 300 presentations of the adaptor with an interstimulus interval of 100 ms. After a 5-s gap subjects heard four test stimuli to be identified as either [b] or [d]. Three seconds separated each stimulus. Following another 5 s of silence, subjects were presented with 1 min or 150 presentations of the adaptor with an interstimulus interval of 100 ms. After another 5-s gap, subjects heard four test stimuli for identification. The tape continued in this manner until all 80 test stimuli, ten occurrences of each of the eight stimuli, were presented.

B. Results and discussion

Baseline and adaptation identification functions were obtained for all subjects. Figure 5 illustrates the mean baseline and adaptation function for each adapting condition for the eight subjects. The locus of the phonetic boundary between [b] and [d] on the stimulus continuum was determined for each identification function of each subject by transforming the percentage of identification responses to a z score and fitting a straight line to this function by a least-means-square analysis. The 50% crossover point was considered to be the locus of the phonetic boundary.

A series of repeated measures one-way analyses of variance were conducted to determine (1) whether significant adaptation occurred for each adapting stimulus; and (2) whether the magnitude of adaptation varied across adaptation conditions. The results revealed significant adaptation for the DAD adapting stimulus [$F(1,7) = 24.10$, $p < 0.01$] and HDAD adapting stimulus [$F(1,7) = 32.13$, $p < 0.01$] and no adaptation with the DAB adapting stimulus [$F(1,7) = 1.11$]. Considering individual performance of the subjects in the various adaptation conditions, with the DAD adapting stimulus, seven subjects showed a shift in the category boundary towards the [d] end of the continuum, and five of these seven showed a boundary shift greater than a

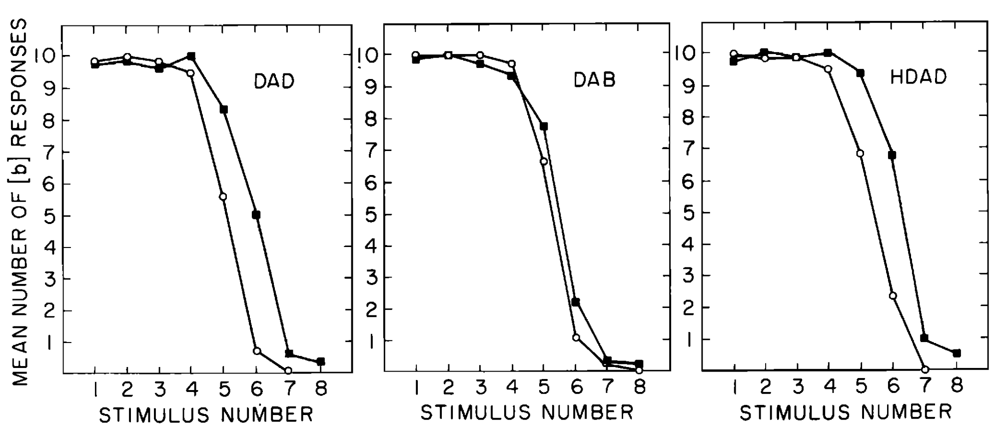

FIG. 5. Results of adaptation using three adapting stimuli on the [ba–da] onset continuum. Baseline functions are represented by the unfilled circles and adaptation functions are represented by the filled squares for each adaptation condition. On the left are the obtained functions before and after adaptation with DAD [[da] onset formant frequencies and diffuse-rising ([d]) spectrum shape]; in the center are the functions before and after adaptation with DAB [[da] onset formant frequencies and diffuse-falling ([b]) spectrum shape]; and on the right are the obtained functions before and after adaptation with HDAD [[da] onset formant frequencies scaled up 10% and diffuse-rising spectrum shape].

half a stimulus. One subject showed no change in the locus of the phonetic boundary after adaptation. With the HDAD adapting stimulus, all eight subjects showed a shift in the category boundary towards the [d] end of the continuum, and seven of the eight showed a boundary shift greater than a half a stimulus. With the DAB adapting stimulus, five subjects showed a shift towards the [d] end of the continuum, and three showed a shift towards the [b] end of the continuum. However, only four subjects showed boundary shifts greater than half a stimulus, with three showing shifts towards the [d] end of the continuum and one showing a shift towards the [b] end of the continuum. Thus the failure to find significant adaptation for the DAB adapting stimulus reflected greater variability amongst the subjects, as well as overall smaller shifts in the locus of the phonetic boundary.

The differences in the magnitude of adaptation across the three conditions was then determined by first calculating for each subject in each condition a difference score reflecting the magnitude of adaptation in each adaptation condition. This score was obtained by subtracting from each subject's obtained adaptation boundary his or her baseline boundary. Comparison of the magnitude of the adaptation effects across the three conditions revealed significantly greater adaptation with both DAD and HDAD compared to DAB [$F(1,7) = 13.55, p < 0.01$, and $F(1,7) = 6.35, p < 0.05$, respectively]. There was no significant difference in the magnitude of adaptation between DAD and HDAD [$F(1,7) = 0.58$].

The results of this experiment demonstrate that listeners do in fact process information about the gross shape of the onset spectrum and can use it to make category decisions for place of articulation. This interpretation is based on the contrasting effects of the DAD and HDAD adaptors versus the DAB adaptor. In particular, presentation of an adapting stimulus containing conflicting place of articulation information in the onset formant frequencies and spectrum tilt effectively eliminated the adaptation effects. However, that adaptation with DAB did not produce significant [b] adaptation indicates that while shape information may contribute to the phonetic decision of the subject, it did not provide the primary perceptual attribute used to identify place of articulation. The fact that scaling up the onset frequencies in the HDAD stimulus while maintaining the appropriate diffuse-rising spectral shape still produced as much adaptation as the DAD adapting stimulus indicates that acoustic isomorphism does not always determine the magnitude of the adaptation effects. Although acoustic isomorphism between the adapting and exemplar stimulus generally produces the most adaptation, several studies have shown otherwise (Sawusch, 1977; Blumstein *et al.*, 1977; Keating and Blumstein, 1978). In particular, they found no differences in the magnitude of place-of-articulation adaptation using adapting stimuli which were not isomorphic with an exemplar test continuum stimulus as compared to those that were acoustically isomorphic. These effects seem to occur either when the differences in onset properties between the adapting and test stimuli is within a limited frequency range (Sawusch, 1977) or when the onset and offset formant frequencies of the transitions are identical but the stimuli vary in either slope of the transition (Keating and Blumstein, 1978) or in the presence or absence of a burst at stimulus onset (Blumstein *et al.*, 1977).

III. GENERAL DISCUSSION

The major claim of a theory of acoustic invariance is that invariant properties may be derived directly from the acoustic signal, and these properties, corresponding to phonetic dimensions, provide an organizational framework for phonetic systems of natural language (Blumstein and Stevens, 1981). As a consequence, these invariant properties give the perceptual system a direct representation of the structural properties inherent in the phonological systems of natural language. Whether a listener uses these properties in ongoing speech perception is a separate issue. It is certainly the case that the perceptual system uses a number of acoustic attributes including the onset frequencies of formants, transitions of formants, and vowel formant frequencies as well as invariant properties to process the sounds of speech (Liberman *et al.*, 1967; Blumstein and Stevens, 1979; Landahl and Blumstein, 1982). As this study sugggests, onset formant frequencies are a critical dimension in place-of-articulation perception, and while the gross shape of the onset spectrum may contribute to place-of-articulation categorization, it is not perceptually primary in the sense that it can override the perceptual cues provided by formant frequencies. Nevertheless, the fact that changes in the gross shape of the spectrum affects both categorization and discrimination of place of articulation suggests that, at the very least, the invariant properties residing in the gross shape provide the listener with a means of organizing the sounds of language into phonetic categories or natural classes.

It may be that the gross shape of the spectrum plays a more salient role in the perception of place of articulation in ongoing speech, but only in relation to spectral changes occurring dynamically over a particular time domain (cf. Kewley-Port, 1981; Lahiri and Blumstein, 1981; Ohde and Stevens, 1981). However, the conditions in this study were not appropriate for addressing this issue. Namely, the stimuli were brief and contained no formant motions in the higher formants. As a result, the dynamic spectral changes from the onset through the transitions and into the vowel found in natural speech were not present in these stimuli. Additional research exploring the gross shape of the spectrum in relation to dynamic properties is required to determine the role of the gross shape of the spectrum in ongoing speech processing.

ACKNOWLEDGMENTS

Many thanks to Aditi Lahiri for running the subjects in experiment I and to Kenneth N. Stevens for his comments on an earlier draft of this paper. This research was supported in part by Grant NS 15123 to Brown University.

Bailey, P. J. (1975). *Perceptual Adaptation in Speech: Some Properties of Detectors for Acoustical Cues to Phonetic Distinctions*, unpublished Doctoral thesis, University of Cambridge, Cambridge, England.

Blumstein, S. E., and Stevens, K. N. (1979). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," J. Acoust. Soc. Am. **66**, 1001–1017.

Blumstein, S. E., and Stevens, K. N. (1980). "Perceptual invariance and onset spectra for stop consonants in different vowel environments," J. Acoust. Soc. Am. 67, 648–662.

Blumstein, S. E., and Stevens, K. N. (1981). "Phonetic features and acoustic invariance in speech," Cognition 10, 25-32.

Blumstein, S. E., Stevens, K. N., and Nigro, G. N. (1977). "Property detectors for bursts and transitions in speech perception," J. Acoust. Soc. Am. 61, 1301–1313.

Cooper, W. E. (1974). "Contingent feature analysis in speech perception," Percept. Psychophys. 16, 201–204.

Keating, P., and Blumstein, S. E. (1978). "Effects of transition length on the perception of stop consonants," J. Acoust. Soc. Am. 64, 57–64.

Kewley-Port, D. (1980). "Representations of spectral change as cues to place of articulation in stop consonants," research on Speech Perception, Tech. Rep. No. 3, Dept. Psychol., Indiana Univ.

Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. 67, 971–995.

Lahiri, A., and Blumstein, S. E. (1981). "A reconsideration of acoustic invariance for place of articulation in stop consonants: Evidence from cross-language studies," J. Acoust. Soc. Am. Suppl. 1 70, S39.

Landahl, K. L., and Blumstein, S. E. (1982). "Acoustic invariance and the perception of place of articulation: A selective adaptation study," J. Acoust. Soc. Am. 71, 1234–1241.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). "Perception of the speech code," Psychol. Rev. 74, 431–461.

Miller, J. L., and Eimas, P. D. (1976). "Studies on the selective tuning of feature detectors for speech," J. Phon. 4, 119–127.

Ohde, R., and Stevens, K. N.(1981). "Effects of burst amplitude on the perception of place of articulation for stops," J. Acoust. Soc. Am. Suppl. 1 70, S33.

Sawusch, J. R. (1977). "Peripheral and central processes in selective adaptation of place of articulation in stop consonants," J. Acoust. Soc. Am. 62, 738–750.

Sawusch, J. R., and Pisoni, D. B. (1978). "Simple and contingent adaptation effects for place of articulation in stop consonants," Percept. Psychophys. 23, 125–131.

Searle, C. I., Jacobson, J. Z., and Rayment, S. G. (1979). "Stop consonant discrimination based on human audition," J. Acoust. Soc. Am. 65, 799–809.

Searle, C. I., Jacobson, J. Z., and Kimerley, B. P. (1980). "Speech as patterns in the 3-space of time and frequency," *Perception and Production of Fluent Speech*, edited by R. A. Cole (Erlbaum, Hillsdale, N. J.), pp. 73–102.

Stevens, K. N., and Blumstein, S. E. (1977). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. 64, 1358–1368.