Auditory property detectors and processing place features in stop consonants

DAVID B. PISONI and JEFFREY TASH
Indiana University, Bloomington, Indiana 47401

The effects of selective adaptation on the perception of consonant-vowel (CV) stimuli varying in place of production was studied under two conditions. In the first condition, repeated presentation of a CV syllable produced an adaptation effect resulting in a shift in the locus of the phonetic boundary between [ba] and [da]. This result replicated previously reported findings. However, in the second condition, an adaptation effect was obtained on this same test series when the critical acoustic information (i.e., formant transitions) was present in final position of a VC speech-like syllable. These latter results support an auditory account of selective adaptation based on the spectral similarity of the adapting stimuli and test series rather than a more abstract linguistic account based on phonetic identity.

A number of recent studies have reported evidence for the existence of feature detectors in the perception of speech sounds with the use of selective adaptation techniques (Cooper, 1974a; Eimas, Cooper, & Corbit, 1973; Eimas & Corbit, 1973; for a review, see Cooper, 1975). The aim of this research has been to demonstrate that the recognition of certain phonetic feature contrasts in speech sounds is mediated by specific linguistic feature detectors, each sensitive to a restricted range of the acoustic properties of these sounds. In the earliest work, Eimas and Corbit (1973) studied the phonetic feature of voicing which distinguishes the voiced stops [b, d, g] from the voiceless stops [p, t, k]. The articulatory dimension distinguishing these sounds is called voice onset time (VOT) and refers to the interval between the release of the articulators and the onset of laryngeal pulsing (Lisker & Abramson, 1964). VOT manifests itself acoustically in terms of a complex of features which includes variations in the onset time of the first formant (F1) relative to the second and third formants (F2, F3). In addition, during the time period when energy is not present in the F1 region, F2 and F3 are excited by noise rather than harmonics.

To assess the degree to which selective adaptation alters voicing perception, Eimas and Corbit constructed a continuum of synthetic consonant-vowel (CV) syllables by systematically varying the stimuli in equal acoustic steps of VOT, between [ba] and [pʰa]. Identification functions were obtained for listeners in an unadapted state and then again after a period of adaptation. Adaptation was accomplished by

repeatedly presenting a CV syllable selected from either extreme end of the stimulus continuum before obtaining identification responses. The initial results showed that adaptation with [ba] caused the locus of the phonetic boundary between [ba] and [pʰa] to shift toward the [ba] end of the continuum. Thus, stimuli near the boundary which were identified as [ba] when the listener was in an unadapted state were subsequently labeled [pʰa] after adaptation. Similar findings were obtained when [pʰa] was the adapting stimulus; the locus of the phonetic boundary shifted toward the [pʰa] end of the continuum. In the same experiment, Eimas and Corbit also showed that the adaptation effects were not specific to the syllable or phonemes in the test series but rather were due to a subcomponent or feature of the consonant segments. For example, adaptation with the voiceless bilabial stop [pʰ] produced approximately equivalent effects on the identification functions for a series of alveolar stop consonants (i.e., [d] and [tʰ]) as it did for a series of bilabial stops ([b] and [pʰ]). In both cases, the locus of the phonetic boundary shifted toward the voiceless end of the continuum.

Eimas and Corbit (1973) interpreted these initial results as support for the hypothesis that the selective adaptation effects are phonetic rather than auditory. That is, the effects were due to fatiguing of detectors that are specialized for processing linguistic features as opposed to processing the auditory information that may underlie these distinctions. In another study, Eimas et al. (1973) attempted to provide additional support for the phonetic hypothesis. They reasoned that, if the information concerning VOT is extracted by detector mechanisms operating at the auditory level of processing, repeated presentation of only the essential acoustic information required to specify VOT should produce a shift in the phonetic boundary that is equivalent to that obtained when voicing information is presented in a speech context. However, if adaptation of the VOT detectors occurs

only when a speech pattern is the adapting stimulus, it would imply that the voicing detectors are responding exclusively to phonetic rather than auditory information.[1]

To test these predictions, listeners were adapted with the initial 50 msec of the [da] syllable, which, according to Eimas et al., presumably contained the same acoustic information concerning VOT as the whole [da] syllable, although it did not sound like speech. Adaptation with these stimuli produced no significant change in the locus of the phonetic boundary, thus supporting the phonetic interpretation of selective adaptation of voicing. Eimas et al. (1973) argued that since subjects did not perceive the truncated syllable as speech, they did not process the F1 and F2 timing as formant timing. However, since VOT is a temporal cue for voicing involving the relative timing of the onsets of F1 and F2, we feel that this was not an adequate control stimulus to warrant a phonetic interpretation of the adaptation results. It is not unreasonable to suppose that different auditory processes may be employed in the analysis of full syllables than in the analysis of truncated syllables. Moreover, information about the following vowel appears to be necessary for the analysis of VOT, since differences in voicing among the stops are a function of both place of articulation of the consonant and the vowel context (see Lisker & Abramson, 1964; also Klatt, Note 1). There is also strong evidence that the acoustic information, for a particular phonetic segment may be distributed across a whole syllable rather than being restricted to the initial 50 msec of the stimulus (see Liberman, 1970).

Nevertheless, based on these adaptation results, Eimas and his colleagues argued that the perception of voicing contrasts involves two distinct types of feature detectors organized as opponent pairs, a voiced detector (+V) and a voiceless detector (-V). Each detector is selectively tuned to a restricted range of partially overlapping VOT values. Repeated presentation of a stimulus containing a specific feature was assumed to fatigue that detector and therefore reduce its sensitivity. Thus, after adaptation, the opponent or unadapted detector provides more information to the decision process than the adapted detector and ultimately results in a shift in the locus of the phonetic boundary in identification.

A number of recent experiments have also used the selective adaptation paradigm to study the phonetic feature of place of articulation. In English, the place feature serves to distinguish among the voiced stop consonants [b, d, g] and the voiceless stops [p, t, k]. The major acoustic cues which carry place information are the transitions of the second and third formants (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Examination of the place feature provides an important test of the hypothesis

that the selective adaptation effects found previously are due to fatiguing of phonetic rather than auditory feature detectors. For example, the adaptation effects previously observed for the voicing feature might be due to fatiguing of an auditory mechanism sensitive to timing differences in VOT rather than a mechanism that recognizes phonetic features such as +V or -V. Moreover, since VOT is a timing cue, one could argue that the acoustic cues for voicing are more nearly invariant with the following vowel context than the acoustic cues that underlie the place distinctions for stop consonants which show much more context-conditioned variation (however, see Cooper, 1974b; Pisoni, Sawusch, & Adams, Note 2).

Selective adaptation of the place feature has been studied by Cooper (1974a), who used a set of stimuli that ranged perceptually from [bae] to [dae] to [gae]. These stimuli varied systematically in terms of the starting frequencies of the second- and third-formant transitions. Listeners were adapted on either one of the end-point stimuli (i.e., [bae] or [gae]) or the midpoint stimulus [dae] in order to evaluate the effects of the adaptation procedure on the loci of the two phonetic boundaries (i.e., the b-d boundary and the d-g boundary). Cooper's results showed selective adaptation effects for the feature of place of articulation. Repeated presentation of [bae] produced a shift in the locus of the [bae]-[dae] boundary toward the [bae] end of the place continuum, although no change was produced in the [dae]-[gae] boundary. Similarly, adaptation with [gae] caused the [dae]-[gae] boundary to shift toward [gae]. Finally, adaptation with [dae], which was the midpoint stimulus, produced a shift in the loci of both the [bae]-[dae] and the [dae]-[gae] boundaries. Cooper (1974a) suggested that these results indicate that place information is decoded in terms of three separate feature modes, bilabial, alveolar, and velar, since adaptation with the end-point stimuli produced shifts in only the immediately adjacent boundary. Moreover, Cooper interpreted these results as favoring a phonetic, rather than auditory, interpretation of the adaptation effects. However, as we shall argue below, the results are also consistent with an auditory account of selective adaptation.

Cooper (1974a) obtained further support for the phonetic interpretation in another experiment where the adapting and test stimuli differed acoustically. In one condition, the vowel context of the adapting stimulus was manipulated in order to determine if adaptation operates primarily at an auditory feature level of perceptual processing or a more abstract phonetic feature level. In the previous studies, the same vowel was used in both the adaptor and test series so that auditory and phonetic features were isomorphic. For example, when the vowel is the same, the acoustic cues that underlie a particular phonetic feature of the consonant are also the same. In contrast, when the vowel is different, the acoustic cues

for the same phonetic feature in the consonant are also different and thus the separate contributions of auditory and phonetic information can be studied independently (see, for example, Pisoni, 1975; Studdert-Kennedy, Shankweiler, & Pisoni, 1972).

Since the frequencies of the second- and third-formant transitions are dependent on the following vowel and since they provide the essential information for the place feature in stop consonants, Cooper's "cross-vowel" condition represented one way of testing the auditory and phonetic accounts of selective adaptation. No boundary shifts would be anticipated in the cross-vowel condition if selective adaptation operated exclusively at the auditory feature level where common spectral information was processed. However, if boundary shifts are obtained in this condition, the results would suggest that the locus of selective adaptation effects is either at a more abstract level of auditory analysis or at the phonetic feature level or both.

Cooper (1974a) tested listeners with the [bae]-[dae]-[gae] series after adaptation with the real speech syllable [bi] and found a significant shift in the locus of the [bae]-[dae] boundary in the predicted direction. However, the magnitude of the shift was considerably less than that obtained after adaptation with the real speech syllable [bae]. In another condition, the real speech syllable [bae] was compared with the synthetic speech syllable [bae]. Both stimuli produced a shift in the [bae]-[dae] phonetic boundary, although the shift produced by the synthetic syllable was significantly larger than that produced by the real speech syllable. Finally, in a third condition, Cooper used a "cross-consonant" adaptation procedure in which listeners were adapted with the real speech bilabial voiceless stop CV syllable [pʰae] and then tested for identification with the [bae]-[dae]-[gae] series. Although the results indicated that [pʰae] was effective in producing a shift, the magnitude of the effect represented a significant decrement when compared with the shifts obtained with the real speech syllable [bae]. Similar findings have also been reported by Cooper and Blumstein (1974), who varied the manner class of the adaptor.

The results of Cooper's (1974a) experiments on the place feature are basically ambiguous with regard to answering the question of whether the selective adaptation effects have either an auditory or phonetic basis. Cooper (1974a) chose one way out of this dilemma by assuming that selective adaptation is a multicomponent process that operates at several different levels of perceptual analysis in the speech processing system. Thus, according to Cooper, one component of the adaptation effect may be attributable to fatiguing of mechanisms at an auditory level whereas another component may be due to fatiguing of mechanisms at a more abstract phonetic level of analysis. The observed magnitude of the

adaptation effect is therefore assumed to be the sum of the two components. The real speech syllable [bi] may have produced less adaptation than [bae] because it was fatiguing only a feature-specific component of adaptation operating at the phonetic level. Similarly, the synthetic speech syllable [bae] may have produced more adaptation at an auditory level than the real speech syllable [bae] because real speech may have contained additional acoustic information that was not relevant to the perception of phonetic distinctions based solely on differences in the second- and third-formant transitions.

In another recent experiment, Diehl (1975) studied selective adaptation of the place feature in a [bɛ] to [dɛ] continuum. However, in this study the adapting stimuli were chosen so that although they shared identical phonetic features with the test stimuli, they presumably had little in common with them acoustically. This was accomplished by using the voiceless stops [pɛ] and [tɛ] as adaptors in which the place features were cued by bursts rather than formant transitions. The adaptation results showed a significant shift in the predicted direction for six subjects with the burst cued [tɛ]. However, only four of the six subjects showed the expected shift for the burst cued [pɛ] adaptor. Diehl argued that since the burst-cued adaptors had virtually nothing in common acoustically with the acoustic cues specifying place in the test series, at least some component of selective adaptation effects must occur at the phonetic feature level. However, Diehl's results could be explained as an auditory effect by assuming that a burst indicates the initiation of some rapid spectral change (cf. Stevens, Note 3). Thus, bursts may be considered to be somewhat analogous to transitions, since both involve rapid changes in the acoustic spectrum immediately following the release of the consonant. Stevens has suggested that there may be auditory property detectors which respond to the presence or absence of a change in the spectrum.

Several recent studies have also reported evidence which favors an auditory interpretation of the selective adaptation effects. For example, Bailey (Note 4) specifically tested for the presence of spectral overlap on selective adaptation by constructing two sets of [ba] to [da] stimuli. In one set, the second formant was fixed and all information about the place feature was carried by the third formant transitions. In the other set, no third formant was present and all place cues were carried by the second formant. Bailey's results showed cross-adaptation in only one direction from the second formant cue series to the fixed second formant series. But no cross-adaptation was obtained when the adaptor contained third formant place cues and the test series only two formant cues. Thus, these findings argue strongly for an auditory interpretation of selective adaptation based on spectral similarity.[2]

In another experiment, Ades (1974) studied the effect of selective adaptation of consonants in initial

and final position by using a VC syllable adaptor on a CV syllable test series, and vice versa. As in the previous studies, Ades again found that adaptation with either [bae] or [dae] produced positive adapting effects on a [bae] to [dae] test series. Similarly, adaptation with either [aeb] or [aed] produced a positive shift in the locus of the [aeb] to [aed] phonetic boundary. However, when the adapting and test stimuli were drawn from different continua, no differential adaptation effects were observed. Thus, the phonetic segments [b] and [d] in final position failed to produce shifts in the locus of the phonetic boundary between [b] and [d] in initial position and vice versa. At first glance, these results indicate that selective adaptation appears to be dependent on the position of the consonant in the syllable and thus could not be due to fatiguing of detectors at the phonetic feature level, since phonetic segments are assumed to be identical linguistically in any position. However, it is also possible that this latter result was due to the cancellation of opposing effects that occurred at both auditory and phonetic stages. We will return to this again below.

The results of the experiments we have reviewed on selective adaptation have been used by several investigators to argue for the existence of specific phonetic feature detectors in speech perception. Although most of the studies have shown adaptation effects in terms of a shift in the locus of the identification boundary between two phonetic segments, many of the results are ambiguous with regard to specifying where these effects are occurring in the speech processing system or resolving whether there are specific, linguistically oriented, feature detectors in speech perception. For example, Eimas and Corbit (1973) and Eimas et al. (1973) have uniformly interpreted their results in favor of a strict phonetic account. Cooper (1974a, 1975) and Cooper and Blumstein (1974) have argued for both auditory and phonetic components to selective adaptation, whereas the results of Ades (1974) and Bailey (Note 4) suggest an auditory interpretation.

In our view, most of the adaptation results are consistent with an account that assumes at least two levels of auditory processing before phonetic categorization. For example, Stevens (Note 3) has shown that place distinctions among stop consonants can be characterized in terms of rapid spectrum changes which show a general rise in frequencies of the spectrum for labial consonants, a fall for dentals and a diverging for velars in canonical CV syllable environments. Stevens argues that place distinctions could be achieved with three different types of property detectors each sensitive to these particular acoustic attributes. Thus, Cooper's results on place in the cross-vowel and cross-consonant conditions could be accounted for by assuming the existence of auditory property detectors which respond to simple spectral differences among the formant transitions in

these stimuli. That is, there may be auditory detectors that selectively respond to rising, falling, or diverging frequencies. The results of Diehl's study with burst cues would have to be handled in a slightly more complex way by assuming auditory detectors that respond to the presence or absence of a rapid change in the spectrum at a particular frequency. Thus, at this level, a burst would be considered somewhat analogous to a formant transition, since both initiate a rapid change in the spectrum (see Stevens, Note 3).

Given the ambiguous findings of most of the previous studies, we carried out the present study to determine the locus of selective adaptation effects for the place feature. To accomplish this goal, it was necessary to construct an adapting stimulus that would preserve all of the auditory information underlying the place feature while dissociating any corresponding phonetic information. The sound pattern chosen for this task was what we have called a "speech-embedded chirp." A speech-embedded chirp consisted of a CV syllable's initial 50-msec segment (i.e., a chirp) preceded by a steady-state vowel whose formant values were fixed equal to the starting frequencies of these formant transitions. The advantage of the speech-embedded chirp is that the formant transitions occur in final position of the syllable. This contrasts with the role of the corresponding acoustic segment in the CV syllable where the same acoustic information occurs in initial position. Note also that when identical F1, F2, and F3 transitions are placed at the end of the syllable, F1 shows an articulatorily impossible rise instead of fall.

By using a speech-embedded chirp, we were also interested to see whether any adaptation effect could be obtained with VC stimuli as adaptors and, if so, whether the obtained effect had an auditory or phonetic basis. That is, would adaptation follow the spectral similarity or phonetic identity of the adaptor and test series? We anticipated that any positive adapting effects with the VC speech-embedded chirps would represent adaptation operating only at the site of auditory processing since Ades' (1974) results indicated that a phonetic segment in final position (e.g., [ab]) was unable to affect the same phonetic segment in initial position, and vice versa.

METHOD

Subjects

The listeners were five paid volunteers who responded to an advertisement. All subjects were right-handed native speakers of American English with no known history of a hearing or speech disorder. The subjects were paid at the rate of $2/h. No subject had had any prior experience with the selective adaptation procedure, although two subjects had had some previous experience with synthetic speech stimuli.

Stimuli

The stimuli used in this study were three-formant speech patterns which were prepared on the parallel resonance speech synthesizer at

Table 1
Starting Frequencies of the Second- and Third-Formant
Transitions for the Synthetic CV Test Stimuli

Stimulus	Starting Frequencies (in Hz)	
	F2	F3
1	996	2180
2	1075	2348
3	1155	2525
4	1232	2694
5	1312	2862
6	1386	3026
7	1465	3195

Haskins Laboratories. All stimuli were recorded on magnetic tape for later playback.

The test stimuli consisted of a series of seven synthetic CV syllables, 300 msec in duration. The stimuli ranged perceptually from [ba] to [da] and differed from one another only in the direction and extent of the second- and third-formant transitions. The differences in starting frequency of the F2 and F3 transitions are shown in Table 1. F1 always started at 412 Hz. All transitions were 50 msec in duration and linear. The final 250 msec of the CV syllables consisted of steady-state-formants appropriate for the English vowel [a] with center frequencies at 769 Hz (F1), 1,232 Hz (F2), and 2,525 Hz (F3).

In addition to the seven test stimuli, two additional stimuli were synthesized, a b-"speech-embedded chirp" (b-SEChirp) and a d-"speech-embedded chirp" (d-SEChirp). These stimuli were constructed in the following manner: First, a 250-msec three-formant steady state was inserted at the beginning of the two end-point stimuli in the test series (i.e., Stimulus 1, or [ba], and Stimulus 7, or [da]). The frequency values of these new steady-state formants were set equal to the starting values of formant transitions of Stimuli 1 and 7 (i.e., 412 Hz (F1), 996 Hz (F2), and 2,180 Hz (F3) for the b-SEChirp, and 412 Hz (F1), 1,465 Hz (F2), and 3,195 Hz (F3) for the d-SEChirp). Finally, the original 250-msec steady state (i.e., the vowel [a]) was deleted. Thus, the b-SEChirp and d-SEChirp stimuli may be characterized as 300-msec sound patterns with 250-msec steady states in initial position and 50-msec transitions in final position. Schematized versions of Stimulus 1, Stimulus 7, b-SEChirp, and d-SEChirp are displayed in Figure 1.

Stimuli 1 and 7 and the two speech-embedded chirps were used to construct four separate adaptation sequences. Each adaptation sequence consisted of 100 repetitions of the adapting stimulus with a 300-msec interstimulus interval.

Apparatus

All experimental materials were recorded on audio tape and reproduced on an Ampex AG-500 two-track tape recorder. The stimuli were presented binaurally through Telephonics (TDH-39) matched and calibrated headphones. The gain of the tape-recorder playback was adjusted to give a voltage across the headphones equivalent to 80 dB SPL re 0.0002 dynes/cm² for the vowel [a]. Measurements were made on a VTVM (Hewlett-Packard Model 400) prior to the presentation of each experimental tape. All five subjects were run together in a small experimental classroom used for speech perception experiments.

Procedure

Baseline identification functions were obtained for all listeners in the unadapted state by presenting 20 random sequences of the seven test stimuli with an interstimulus interval of 3 sec. The subjects were instructed to identify each stimulus as either [ba] or [da] by writing the appropriate consonant letter on prepared response sheets. The listeners were told to respond to every stimulus during the identification tests even if they had to guess.

On the day after the initial identification test, adaptation was begun. Each adaptation test lasted roughly 1 h, and took place at 24 h intervals on 4 consecutive days. The order of presentation of

the adapting sessions was as follows: Session 1, [b] adaptation; Session 2, [d] adaptation; Session 3, b-SEChirp adaptation; and Session 4, d-SEChirp adaptation.

Each of the four adaptation tests was conducted in the following manner: Listeners were first presented with two consecutive adaptation sequences (200 presentations) of the selected adapting stimulus. Following this "warm-up" period of adaptation, 10 adaptation trials were administered. Each adaptation trial was composed of 100 presentations of the adapting stimulus with 225 msec between repetitions (i.e., one adaptation sequence). This was followed by 2 sec of silence and then the presentation of the five middle stimuli from the original test series (Stimuli 2-6). The subjects were instructed to identify each of these five test stimuli as either [ba] or [da] by placing the appropriate consonant letter on response sheets. The five middle test stimuli occurred in random order with a 4-sec ISI. After the fifth stimulus was presented for identification, there was a 5-sec interval before the onset of the next adaptation trial. Each of the 10 adaptation trials had a different random order of the five test stimuli. Each stimulus occurred in each of the five test positions twice. After one presentation of the experimental adaptation tape, the subjects were given a short break, after which the same tape was rewound and played again. Thus, each of the five middle stimuli in the test series was presented for identification a total of 20 times within a single adaptation session on a given day.

RESULTS

Individual and mean phonetic boundaries for each of the five experimental sessions, one identification session without adaptation and the four adaptation sessions, are shown in Table 2. Each phonetic boundary was determined by a computer program that found the point along the stimulus scale which would, by linear interpolation, receive 50% [ba] responses and 50% [da] responses. In all, there were
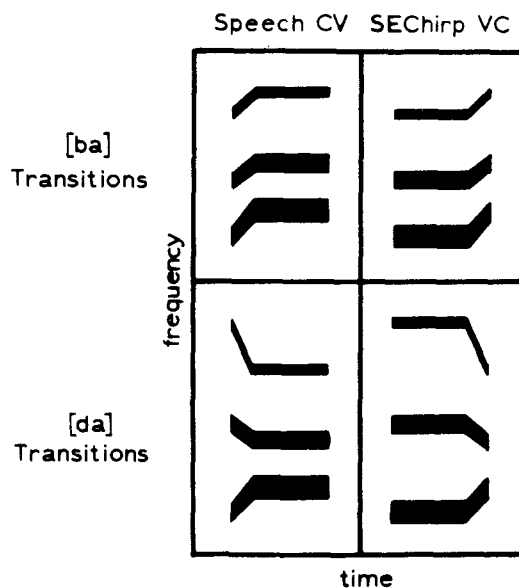
Figure 1. Schematized spectrographic patterns illustrating the speech adapting stimuli, [ba] and [da], and the speech-embedded chirps, b-SEChirp and d-SEChirp. The speech-embedded chirps contained the same formant transitions as the speech stimuli to their left in the figure.

Table 2
Individual and Mean Loci of Phonetic Boundaries
for Each Test Condition

Subjects	Without Adapta-tion	Speech Stimuli		Speech-Embedded Chirps	
		Adaptation With [ba]	[da]	Adaptation With b-SEChirp	d-SEChirp
1	4.10	3.25	5.25	3.56	4.50
2	3.59	2.60	4.36	3.41	3.59
3	3.57	2.63	4.63	3.47	4.17
4	3.95	3.10	5.64	3.56	4.29
5	4.10	2.59	5.64	3.71	4.50
Mean	3.86	2.83	5.10	3.54	4.21

20 instances of attempted adaptation, four adapting conditions for each of five subjects. In all but one instance there was a shift in the locus of the [ba]-[da] phonetic boundary in the predicted direction. The only exception was Subject 2, who showed virtually no shift at all in the d-SEChirp adaptation test.

The group identification and adaptation functions, averaged over all five subjects, are plotted in Figure 2. From inspection, it may be observed that the shifts produced by selective adaptation were not accompanied by a decline in the steepness of the slopes of the identification functions.

All shifts in the locus of the phonetic boundaries obtained after adaptation were found to be significantly different from the baseline identification functions. Since the predicted direction of the boundary shift was toward the category of the adapting stimulus, one-tailed correlated t tests were used.

For the [ba] and [da] syllable adaptation conditions, the respective mean shifts in the locus of the [ba]-[da] phonetic boundary of -1.03 and +1.24 were found to be highly significant (t = 8.36 for [ba] adaptor, p < .005; t = 7.80 for [da] adaptor, p < .005). These shifts are comparable in direction and magnitude to those found in previous studies which investigated adaptation effects on the place feature (e.g., Ades, 1974; Cooper, 1974a; Cooper & Blumstein, 1974).

For the speech-embedded chirp adaptation condition, the mean boundary shifts of -.32 (b-SEChirp) and +.35 (d-SEChirp) were also found to be significant (t = 4.02 for b-SEChirp adaptor, p < .01; t = 3.63 for the d-SEChirp adaptor, p < .025). Under these conditions, the direction of the shift was toward the phonetic category from which the adaptor's formant transitions were originally obtained. Thus, after adaptation with the b-SEChirps, the subjects made fewer [ba] identification responses. Similarly, after adaptation with the d-SEChirps, the listeners assigned fewer identification responses to the [da] category.

Larger shifts were obtained with the speech adaptors than with the speech-embedded chirps. The differences between the [ba] adaptor and the

b-SEChirp adaptor and the [da] adaptor and d-SEChirp adaptor were also highly significant (t = 4.83, p < .005, and t = 5.65, p < .005, respectively).

DISCUSSION

The results of the first condition that employed speech stimuli as adaptors replicated previous findings. Repetitive listening to a CV syllable from either end of the stimulus continuum produced a shift in the locus of the phonetic boundary. The shift was toward the adapting stimulus, indicating that a greater number of identification responses were assigned to the unadapted category. The obtained shifts were comparable to those found previously by other investigators.

However, turning to the second condition, we find that the nonspeech VC stimuli also produced shifts in the perception of a CV place series. The obtained shifts were in the direction predicted by an auditory account based on spectral similarity of the adaptor and test series. Although small in magnitude, these results provide additional support for the argument that the selective adaptation effects reflect the fatiguing of some acoustically oriented mechanism that operates during the auditory stage of analysis in speech perception. This conclusion is based on the assumption that no phonetic information contributed to the observed adaptation effects and follows from Ades' (1974) results which showed that repeated presentation of a consonant in final position was unable to produce shifts in the locus of the phonetic boundary for the consonant in central position. Thus, the adaptation found with the speech-embedded chirps in the present experiment appears to be due exclusively to interaction at an auditory level.

Our conclusions are strengthened, however, if we consider the possibility that some phonetic adaptation occurred in the speech-embedded chirp condition. A phonetic account of adaptation with speech-embedded chirps would have predicted shifts in the loci of the [ba]-[da] boundary in the opposite direction to those predicted by an auditory account. Thus, a phonetic interpretation would predict fewer responses assigned to the [d] category after adaptation with the b-SEChirp and fewer [b] responses after adaptation with the d-SEChirp. Indeed, this is precisely the
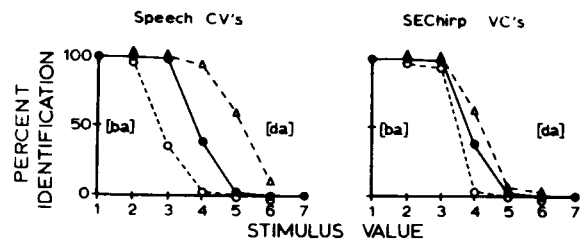
Figure 2. Group identification functions obtained under conditions of adaptation with CV speech syllables (left panel) and VC speech-embedded chirps (right panel). The data are based on the means of all six subjects.

account rejected by Ades when he failed to obtain any shifts on a CV continuum with a VC adaptor and vice versa. But his null results are ambiguous for two reasons. First, the possibility of an auditory effect still remained, because Ades' initial and final consonants, although phonetically identical, were not acoustically identical. Thus, the absence of an adaptation effect might therefore have been due to the lack of acoustic similarity between adaptor and test series. But a second and perhaps more interesting possibility exists. Ades' failure to find a clear-cut auditory or phonetic effect may have been due to cancellation of the combined effects of adaptation at both auditory and phonetic levels. Since the shifts are predicted to go in opposite directions when each component is taken separately, these effects may have simply combined to produce a null result when auditory and phonetic information is simultaneously placed in conflict.

Consider the differences in the acoustic cues underlying the phonetic segments [b] and [d] in initial and final position. Given the same vowel environment such as [a], the phonetic segment [b] has rising transitions for all three formants when it is in initial position of a CV syllable but falling transitions when it is in final position of a VC syllable. On the other hand, the phonetic segment [d] has falling second- and third-formant transitions and a rising first-formant transition in initial position and just the reverse configuration in final position. That is, the [d] in final position has rising second- and third-formant transitions and a falling first-formant transition. In this experiment, all three formant transitions were rising in the b-SEChirp stimulus. In the d-SEChirp, the first-formant transition was rising whereas the second- and third-formant transitions were falling. Since the first formant does not convey any information about the place feature, its contribution can be assumed to be irrelevant in this situation. Thus, if selective adaptation occurred at a phonetic level, we should have found a shift in the phonetic boundary toward the [da] category after adaptation with the b-SEChirp, since it contains the appropriate formant transitions for [d] in final position. Similarly, we should have found a shift in the boundary toward [ba] after adaptation with the d-SEChirp, since its transitions are appropriate for [b] in final position. Instead, our findings showed precisely the opposite effects, indicating that, at least under the present conditions, selective adaptation appears to be more a function of spectral similarity than of phonetic identity. Thus, one component of selective adaptation has an auditory basis in terms of the similarity of the spectral properties of the adapting stimuli and test series. Recall that the speech-embedded chirps had the same formant transitions as the speech adaptors, except that these transitions now occurred in final position in the syllables. It is difficult for us to see how the outcome of this study could be reconciled by appeal to some type of phonetic explanation. Indeed,

these results provide good support for an account that postulates auditory property detectors that respond to rising or falling spectral information. Although Ades (1974) rejected this auditory explanation on the grounds that he could find no evidence of adaptation effects of initial [b] on final [d] and vice versa, we think we have good evidence for resurrecting this account in the present context. Indeed, a careful reexamination of Ades' data (Table 1, p. 63) in the crossed syllable conditions shows some trend in the direction predicted by the auditory account.

One additional finding of the present study should be considered since it is relevant to several previous results. Although we found a selective adaptation effect for the VC speech-embedded chirps, the magnitude of the effect was considerably smaller than that typically found with the speech CV syllables. This is a result that has also been reported by Cooper in the cross-consonant and cross-vowel conditions and by others, and appears to be obtained whenever the adaptor and test series differ acoustically. However, the presence or absence of an adaptation effect should clearly be distinguished from the magnitude of the effects obtained under various experimental conditions. Thus, it appears that the greater the spectral similarity between adapting stimuli and test continuum, the greater will be the adaptation effects. It is also possible that an additional level of auditory processing exists where more abstract properties of the stimulus are analyzed prior to phonetic categorization. For example, there may be auditory property detectors analogous to those proposed by Stevens which respond neither to rising nor falling spectral information but only to the presence or absence of energy in a particular frequency region.

We should also point out here that, within the context of this experiment, we have tried to argue for an auditory as opposed to a phonetic account of the adaptation results. However, the possibility of some additional phonetic contribution to adaptation cannot be ignored. For example, Cooper (1974a, 1975) recently proposed that selective adaptation is a multicomponent process that operates at several different levels of the speech proceessing system. According to Cooper, one component of adaptation may be attributable to fatiguing of mechanisms at an auditory level whereas another component may reflect the contribution of adaptation at a phonetic level. Cooper (1975) has summarized several perceptual experiments in which adaptation appears to take place at a phonetic feature level, since acoustic commonality could not be easily found by an examination of some simple invariant property of the stimuli. Cooper (1975) also reported the results of several studies showing perceptuo-motor adaptation effects. Assuming these findings are reliable, it would be quite difficult to maintain a general auditory account of all the selective adaptation effects.

However, results supporting our conclusions for a

strong auditory basis to selective adaptation also have been reported recently by a number of other investigators. For example, Cooper (1974b) studied the effects of vowel environment on adaptation of the voicing feature. Using a contingent adaptation procedure, he presented listeners with an alternating sequence of two different adaptors, [da] and [tʰi], and then tested along two continua, [ba] to [pʰa] and [bi] to [pʰi]. If the adaptation effect occurred at a phonetic level, no adaptation would be anticipated, since the opponent effects of voicing in [da] and [tʰi] should cancel each other out. However, Cooper found just the opposite result, namely, a different effect on each of the identification functions. The boundary shifted toward [ba] in the [ba]-[pʰa] series but toward [pʰi] in the [bi]-[pʰi] series. These results indicate that adaptation of voicing occurs on an auditory level which is sensitive to vowel context. Recently, we replicated and extended the contingent adaptation procedure for both place and voicing features and found basically the same results, which are also consistent with an auditory account of selective adaptation (Pisoni, Sawusch, & Adams, Note 2).

More recently, Tartter (1975) has shown that the greater the spectral similarity between adaptor and test sequence, the greater the adaptation effects. In this study, she also used various components of the speech signal (e.g., an isolated second formant, formant transitions, etc.) and found that these are also sufficient to produce adaptation effects, although, as in previous studies, the magnitudes of these shifts were smaller than those obtained with the full speech stimuli.

In summary, we have found evidence for selective adaptation of the place feature in initial consonant-vowel syllables when the critical acoustic information is contained in the final position of a speech-like VC syllable. Although these results contrast with those reported recently by Ades (1974), they are consistent with an auditory account of selective adaptation which assumes some spectral interaction of adaptor and test stimuli at an auditory stage of perceptual processing prior to phonetic categorization.

REFERENCE NOTES

1. Klatt, D. H. *Voice-onset time, frication and aspiration in word-initial consonant clusters.* Quarterly Progress Report of the Research Laboratory of Electronics, M.I.T., 1973, 109, 124-136.

2. Pisoni, D. B., Sawusch, J. R., & Adams, F. T. *Simple and contingent adaptation of place and voicing features in stop consonants.* Paper to be presented at the Acoustical Society of America, San Francisco, California, November 1975.

3. Stevens, K. N. *The potential role of property detectors in the perception of consonants.* Paper presented at the Symposium on Auditory Analysis and Perception of Speech, Leningrad, August 1973.

4. Bailey, P. *Perceptual adaptation for acoustical features in speech.* Speech Perception, 1973, Series 2, No. 2, 29-34, Department of Psychology, The Queen's University of Belfast.

REFERENCES

ADES, A. How phonetic is selective adaptation? Experiments on syllable position and vowel environment. *Perception & Psychophysics*, 1974, 16, 61-67.

COOPER, W. E. Adaptation of phonetic feature analyzers for place of articulation. *Journal of the Acoustical Society of America*, 1974, 56, 617-627. (a)

COOPER, W. E. Contingent feature analysis in speech perception. *Perception & Psychophysics*, 1974, 16, 201-204. (b)

COOPER, W. E. Selective adaptation to speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory* (Vol. I). Potomac, Md: Erlbaum, 1975.

COOPER, W. E., & BLUMSTEIN, S. E. A labial feature analyzer in speech perception. *Perception & Psychophysics*, 1974, 15, 591-600.

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

EIMAS, P. D., COOPER, W. E., & CORBIT, J. D. Some properties of linguistic feature detectors. *Perception & Psychophysics*, 1973, 13, 247-252.

EIMAS, P. D., & CORBIT, J. D. Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 1973, 4, 99-109.

LIBERMAN, A. M. The grammars of speech and language. *Cognitive Psychology*, 1970, 1, 301-323.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, 74, 431-461.

LISKER, L., & ABRAMSON, A. S. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 1964, 20, 384-422.

MATTINGLY, I. G., LIBERMAN, A. M., SYRDAL, A. K., & HALWES, T. Discrimination in speech and non-speech modes. *Cognitive Psychology*, 1971, 2, 131-157.

PISONI, D. B. Dichotic listening and processing phonetic features. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory* (Vol. I). Potomac, Md: Erlbaum, 1975. Pp. 79-102.

STUDDERT-KENNEDY, M., SHANKWEILER, D., & PISONI, D. B. Auditory and phonetic processes in speech perception: Evidence from a dichotic study. *Cognitive Psychology*, 1972, 3, 455-466.

TARTTER, V. C. *Selective adaptation of acoustic and phonetic detectors.* Unpublished MA thesis, Brown University, June 1975.

NOTES

1. The rationale used here for the control stimuli is similar to that employed by Mattingly, Liberman, Syrdal, and Halwes (1971) in their study of the identification and discrimination of speech and nonspeech sounds. First, nonspeech stimuli are created which differ in the same acoustic parameters as the speech stimuli under study but do not sound like speech. The control stimuli are then used in the same experimental design as the speech stimuli and the results compared. If the outcome is the same in both cases, one attributes the results to the acoustical properties of the stimuli and not to their being heard as speech per se. On the other hand, if differences are obtained, it is assumed that they are due to the nature of the stimuli being perceived as speech and not to their physical properties. This conclusion is, of course, predicted on the assumption that the context in which the speech cues occur is irrelevant to its perception, which is highly questionable to begin with.

2. One of the reviewers pointed out the the Bailey experiment must be interpreted with some caution. For example, if the fixed F2 was a stronger cue for [d] than [b] and if F2 cues are stronger than F3 cues for the place feature, then both the [b] and [d] fixed F2 cues should produce an adaptation effect that was [d]-like, with the latter stimulus producing a somewhat larger effect. This is exactly what Bailey found in his experiment. Thus, the absence of a specific F2 transition does not imply that it fails to contribute to the observed adaptation effects.