

Selective Adaptation of Auditory  
Feature Detectors in Speech Perception

by

Jeffrey B. Tash

Submitted to the Faculty of the Graduate School in  
partial fulfillment of the requirements for the  
degree of Master of Arts in the  
Department of Psychology  
Indiana University  
August, 1974

## ACKNOWLEDGEMENTS

I wish to express my deepest appreciation to Dr. Lloyd R. Peterson who served as a member of my committee, and to Dr. David B. Pisoni, chairman of the committee, whose criticisms and suggestions were invaluable in the design of this study and in the preparation of this thesis.

In addition, I would like to thank Dr. James Cutting, Yale University and Haskins Laboratories, for constructing the experimental materials used in this investigation.

I also wish to express my gratitude to Dr. David Pisoni for providing the funds (NIMH Research Grant MH 24027-01) and the facilities without which this study could never have been done.

Above all, I wish to thank my wife Lois for her invaluable contributions. Not only did she assist with many technical and clerical aspects of this thesis, but more importantly, she provided encouragement, understanding and loving toleration throughout many frustrating moments.

JBT

## INTRODUCTION

It is generally well established that the conversion of a continuously varying acoustic waveform into a string of discrete phonetic segments involves a number of distinct stages or levels of perceptual analysis (Fant, 1967; Stevens and Halle, 1967; Stevens and House, 1972; Studdert-Kennedy, 1974). At the lowest level the physical signal is analyzed into a set of time-varying auditory dimensions such as timbre, pitch and loudness. This low-level auditory information is then operated upon by the next higher stage of processing for the extraction of abstract phonetic features. This distinction between an auditory and a phonetic level of speech processing has been experimentally supported by a number of different researchers (e.g., Studdert-Kennedy and Shankweiler, 1970; Studdert-Kennedy, Shankweiler and Pisoni, 1972; Wood, 1974; Pisoni and Tash, 1974). The present study intends to examine the role of this dichotomy in terms of a feature detector analysis of speech perception.

The notion of a feature detector originally comes from the electrophysiological investigations of single cell neurons. Lettvin, Maturana, McCulloch, and Pitts (1959), studying the visual system of the frog, discovered that specialized neural receptors in the frog's eye extract relatively restricted patterns of information from the visual signal. In their experiment, four classes of

feature detectors were observed: edge detectors, moving detectors, dimming detectors and convex edge detectors. Edge detectors responded whenever a border between light and dark occurred within a specific receptive field. Moving detectors were activated by the presence of a moving edge. Dimming detectors reacted to an overall decrease in illumination. And convex edge detectors responded whenever a small, dark, moving object appeared in the visual field (i.e., a bug). The nature of these detector mechanisms suggest that they provide exactly the visual information necessary for the frog to survive in its sensory-restricted environment.

In a series of later experiments, Hubel and Wiesel developed a technique which enabled them to record from single cell units in the visual cortex of the cat (1962, 1965) and of the monkey (1968). What they discovered was the existence of highly specialized neural mechanisms in the cortex capable of extracting abstract features or patterns from the visual signal. Furthermore, Hubel and Wiesel noted that this extraction process involves multiple stages of analysis, where depth of processing is directly related to the complexity of feature abstraction. For example, detector mechanisms at the lowest level of perceptual analysis effectively function as abstract pattern recognizers by extracting basic features from small, specific receptive fields. These low-level feature detectors which are located on the retina correspond to

the pattern analyzers discovered by Lettvin et. al. These processes, however, represent only the first levels of perceptual analysis. The neural signals generated by these peripheral feature detectors are then sent to the visual cortex for more advanced stages of processing. At the lowest levels of cortical analysis, feature detectors exist which monitor the output of the retinal detector mechanisms. Since the information being monitored is already abstract in nature, the task of the cortical pattern analyzers is to detect abstract features from an abstract message. Similarly, information extracted by these low-level cortical detectors is then sent to higher-level detector mechanisms for subsequent pattern analysis. In this hierarchical manner, the system is able to continually extract relatively more and more abstract features from the original physical signal. Thus, a perceptual system based on increasingly complex levels of detector mechanisms, provides the power required for the recognition of abstract patterns and features.

Evidence for the existence of feature detectors involved in the perception of speech signals was originally presented in a study by Eimas and Corbit (1973). Their intention was to demonstrate by means of a selective adaptation procedure that the perception of voicing contrasts in speech is mediated by linguistic feature detectors, each sensitive to a restricted range of voice onset times.

Voiced onset time (VOT) is a major acoustic cue under-

lying the perceived phonetic distinction between voiced and voiceless stop consonants. For example, in English it distinguishes /b/ from /p/, /d/ from /t/, and /g/ from /k/. In terms of production, VOT has been defined as the interval between the release of the articulators and the onset of laryngeal pulsing (Lisker and Abramson, 1964). Acoustically, it refers to the delay in the onset of the first-formant relative to the second- and third-formants. Additionally, when the first-formant is absent, the second- and third-formants are noisy rather than voiced (Lisker and Abramson, 1970). The amount of delay in VOT required for a stop to be heard as voiceless rather than voiced is normally about 30 - 40 msec (Abramson and Lisker, 1970).

In order to assess the degree by 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 steps of VOT. Identification functions were then obtained for listeners in the unadapted state and after adaptation. Adaptation was accomplished by repeatedly presenting a CV syllable selected from either extreme end of the VOT continuum.

Eimas and Corbit reasoned that if a given detector is selectively sensitive to a particular feature in a stimulus pattern, then repeated presentation of that feature should fatigue the detector and reduce its sensitivity. As such, they predicted that adaptation of the

voicing feature should cause a shift in the locus of the phonetic boundary in the direction toward the adapting stimulus' end of the VOT continuum. The results confirmed their predictions. Adaptation with /ba/ caused the phonetic boundary between /ba/ and /pa/ to shift toward the /ba/ end of the continuum. In other words, stimuli near the boundary which were identified as /ba/ when the listener was in an unadapted state, were subsequently labeled /pa/ after adaptation. Similar results were obtained when /pa/ was the adapting stimulus.

The results discussed thus far do not conclusively demonstrate that the observed effects are due to the selective adaptation of "linguistic" feature detectors. Alternatively, it may be that the sound patterns corresponding to the phonetic units are being adapted, and as such, the feature detectors may be auditory rather than phonetic in nature.

Eimas and Corbit rejected this alternative explanation by pointing out that the effects of adaptation are not class specific as indicated by the presence of crossed-consonant shifts. For example, adaptation with the voiceless bilabial stop /p/, produced approximately equivalent effects on the identification functions for a series of alveolar (/d/ and /t/) stop consonants as it did for a series of bilabial (/b/ and /p/) stops. In both cases the locus of the phonetic boundary shifted toward the voiceless end of the continuum, indicating that a greater number of identification responses belonged to the voiced or unadapted

category. From these results taken together with some additional findings based on a discrimination task, Eimas and Corbit concluded that the perception of voicing contrasts involves two distinct classes of feature detectors, each class being specifically tuned to a restricted range of VOT values.

In a subsequent study reported by Eimas, Cooper, and Corbit (1973), an attempt was made to corroborate the earlier interpretation that the effects of adaptation are the result of phonetic rather than auditory perceptual analysis. They reasoned that if the information concerning VOT is extracted by detector mechanisms operating at the auditory level of processing, then repeated presentation of just the essential acoustic information required to specify VOT should produce an equivalent shift in the phonetic boundary as that incurred when voicing information is presented in a speech context. If, however, adaptation of the VOT detectors occurs only with a speech pattern as the adapting stimulus, then it would appear reasonable to infer that the voicing detectors are operating only during phonetic processing.

In order to test these assertions, listeners were repeatedly exposed to a synthetic CV syllable /da/, selected from the extreme end of the VOT continuum. The results from this condition replicated the findings obtained earlier in the Eimas and Corbit experiments. The locus of the phonetic boundary shifted toward the voiced end of the

continuum. In a different experimental condition, identification functions were obtained after adaptation with d-chirps. These adapting stimuli consisted of only the initial 50 msec of the /da/ syllable. Although these initial portions of the speech pattern carry the same acoustic information concerning VOT as the entire /da/ syllable, they do not sound at all like speech, but rather like glissandos or the chirps of a bird (Lieberman, 1970). Adaptation with d-chirps produced no significant changes in the phoneme boundary. These results lend additional support to the notion of linguistic feature detectors specifically sensitive to the distinctive features of speech which are engaged by the language processing system during the phonetic stage of perceptual analysis.

All of the studies discussed thus far have dealt exclusively with the phonetic feature of voicing. More recent experiments have examined the effects of adaptation along the phonetic dimension place of articulation. In English, the feature of place serves to distinguish among the voiced stop consonants /b,d,g/ and the voiceless stops /p,t,k/. The major acoustic cues which carry this information are the transitions (i.e., relatively rapid changes in the frequency of the formants) of the second- and third-formants (Lieberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967). However, unlike the feature of voicing, there exists no invariant range of acoustic cues which determine place distinctions. Instead, the major acoustic

cues for place are highly subject to contextual variation. This is so, because the formant transitions, at every instant, simultaneously provide information about both the stop consonant and its vowel environment. The classic example illustrating this point comes from an examination of the syllables /dɪ/ and /du/. In the case of /dɪ/, the acoustic cue for the perception of the consonant /d/ is a rising second-formant transition. It begins at 2200 Hz and climbs up to 2600 Hz. In /du/, the phoneme /d/ is cued by a second-formant transition that falls from about 1200 to 700 Hz. Thus, the same perceived phoneme is cued, in different contexts, by auditory features that are vastly different in acoustic terms.

In a study by Cooper (1974a) an attempt was made to determine whether the selective adaptation procedure can produce alterations in perception along the acoustically non-invariant place feature in a manner analogous to that observed for voicing. It was argued that if adaptation operates only on invariant acoustic information, then perceptual shifts should not occur for the feature of place. If, on the other hand, adaptation is a function of phonetic processing, then perceptual shifts should be observed. To examine these issues, Cooper constructed a set of thirteen synthetic speech syllables ranging perceptually from /bae/ to /dae/ to /gae/, by systematically varying the starting frequency of the second- and third-formant transitions. By then presenting listeners with repeated

presentations of either one of the endpoint stimuli or the midpoint stimulus, it was possible to evaluate the effects of the adaptation procedure on the loci of the two phonetic boundaries. The obtained results indicate that a selective adaptation effect can be induced for the feature place of articulation. Thus, repeated presentation of /bae/ produced a shift in the locus of the /bae/-/dae/ boundary toward the /bae/ end of the place continuum. However, since place of articulation is a tri-valued dimension, it is necessary to also examine the /dae/-/gae/ phonetic boundary.

If adaptation is operating on stable feature modes such as bilabial, alveolar, and velar, then no shift is predicted for the alveolar-velar phonetic boundary when the adapting stimulus is bilabial. On the other hand, if it is found that this boundary does shift toward the bilabial end of the place continuum, such that stimuli that were originally identified as alveolar were, after adaptation, identified as velar, then this would indicate that adaptation was effecting a purely relative feature analyzing system. The results suggest that the former is correct. Adaptation with /bae/ produced no significant change in the /dae/-/gae/ boundary. Similarly, adaptation with /gae/ caused the /dae/-/gae/ boundary to shift toward /gae/ but had no effect on the /bae/-/dae/ phonetic boundary. Finally, adaptation with /dae/ (the midpoint stimulus) produced a shift in the loci of both the /bae/-/dae/ and

the /dae/-/gae/ boundaries. These results, then, suggest that place information is decoded in terms of three separate analyzer modes, and that the locus of a phonetic boundary corresponds to the point along the place continuum where two adjacent analyzer modes respond with equal strength.

After having obtained results consistent with Eimas et. al.'s original "phonetic" interpretation of feature detectors for the feature place of articulation, Cooper set out to examine the extent to which the adapting stimulus could be varied. In one condition, a test was conducted to determine the role of vowel context on the adapting stimulus. It was argued that if adaptation operates primarily at an acoustic-invariant level of perceptual processing, then no boundary shifts should occur in a "crossed-vowel" adaptation test since the frequency values of second- and third-formant transitions are highly vowel dependent. On the other hand, if boundary shifts do occur, then this would suggest that the phonetic feature detectors fatigued during adaptation are specifically sensitive to the distinctive features of speech. To examine this problem, listeners were tested along the /bae/-/dae/-/gae/ identification series after adaptation with the real speech syllable /bi/. The obtained identification functions indicate the presence of a significant shift in the locus of the /bae/-/dae/ phonetic boundary. However, the magnitude of this shift was considerably less than those obtained after adaptation with the real speech syllable /bae/.

In a second experimental condition, the real speech syllable /bae/ was compared with the synthetic speech syllable /bae/ in terms of their effectiveness as an adapting stimulus for the synthetic /bae/-/dae/-/gae/ test series. The results showed that both stimuli were able to produce a shift in the /bae/-/dae/ phonetic boundary, but that the shift incurred by the synthetic speech syllable was significantly larger than that incurred by the real speech syllable.

Finally, a third experimental condition was run in which a "crossed-consonant" adaptation strategy was employed. Listeners in this condition were adapted with the real speech voiceless bilabial stop CV syllable /p<sup>h</sup>ae/ and then tested for identification with the synthetic speech series /bae/-/dae/-/gae/. Although the results indicate that /p<sup>h</sup>ae/ was effective in producing a perceptual shift along the /bae/-/dae/-/gae/ continuum, the magnitude of this shift represented a significant decrement when compared with those obtained when the adapting stimulus was the real speech voiced bilabial stop CV syllable /bae/.

The results from these three different experimental conditions strongly suggest that Eimas' interpretation may be too simplistic. Perhaps instead, it would be more beneficial to view adaptation as a multi-component process which operates during multiple levels of perceptual analysis, such that part of the adaptation effect is attributable to the fatiguing of feature detectors at the auditory level,

while another part is due to the fatiguing of feature detectors at the phonetic level. Thus, /bi/ may have been a less effective adapting stimulus than /bae/ because it was only fatiguing the feature-specific component of adaptation operating at the phonetic level of processing. Similarly, the synthetic speech syllable /bae/ may have produced more auditory adaptation than the real speech syllable /bae/ because the real speech syllable contained more acoustic information irrelevant to the task of perceiving phonetic distinctions based solely on differences in the second- and third-formant transitions. However, this interpretation still does not explain why the real speech syllable /bae/ should have been a more effective adapting stimulus than the real speech syllable /p<sup>h</sup>ae/. At the auditory level, the information specifying place should have been nearly identical for both of these syllables. Also, at the phonetic level both should have been classified as bilabial. Thus, it appears that some additional explanatory component is required to account for the observed difference.

Cooper's explanation for the significant difference in shift magnitude between the real speech /bae/ and /p<sup>h</sup>ae/ adaptation conditions was based on the postulation of a high-level "phonetic unit" component of adaptation. This phonetic unit component presumably operates on consonantal sounds as a unit rather than on individual distinctive features. Therefore, a larger shift occurred for the adapting stimulus /bae/ because this syllable's

consonant was represented directly in the /bae/-/dae/-/gae/ test series whereas the syllable /p<sup>h</sup>ae/'s consonant was not. Thus, it can be seen that Cooper's interpretation of the selective adaptation results assumes not only that the processing of speech sounds takes place at more than one level of perceptual analysis, but also that more than one such level is capable of being adapted (Cooper, 1974b). This approach closely resembles Hubel and Wiesel's account of a perceptual system based on increasingly complex levels of detector mechanisms. However, this is not the only possible explanatory model of the selective adaptation phenomena.

A different explanation can be derived from Helson's Adaptation Level Theory (1964). According to this theory, listeners should partition the /bae/-/dae/-/gae/ test continuum into three equivalent categories when each stimulus occurs with an equal probability. However, when one stimulus occurs more often than any of the other stimuli, the theory predicts that the phonetic boundary should shift toward the more frequently occurring stimulus' end of the continuum. Thus, for the selective adaptation procedure, both a feature detector theory and Helson's adaptation level theory predict that the locus of a phonetic boundary should shift toward the adapting stimulus' end of the test continuum. For the feature detector theory, the effect is due to a generalized decrease in the sensitivity of a feature detector across its entire response range incurred by repeated presentation of its adequate

stimulus. For the adaptation level theory, the effect is attributed to simple response bias. That is, during adaptation the listener is exposed to many more instances of the category from which the adapting stimulus was drawn, and as a result, he has a tendency to identify sounds as belonging to a category other than the one assigned to the adapting stimulus. What Helson's theory is describing is a phenomenon commonly referred to in the psychophysical literature as the contrast effect.

In an attempt to distinguish between these two differing interpretations of the selective adaptation results, Sawusch and Pisoni (1973) and Sawusch, Pisoni, and Cutting (1974) examined the effects of unbalanced probabilities of occurrence of stimuli on the identification functions of a voicing (/ba/-/pa/) and a place (/bae/-/dae/) test continuum. By manipulating the distribution of probabilities such that one of the endpoint stimuli appeared more often than any of the other members of the test series, it was possible to determine whether frequency of occurrence can, by itself, alter phonetic boundaries. According to adaptation level theory, the boundary should shift toward the more frequently occurring stimulus. Feature detector theory, on the other hand, makes no such prediction since relatively no adaptation (fatiguing) is operating within this test procedure. The results indicate that for the class of stop consonants phonetic boundaries do not shift as a function of probability of stimulus occurrence.

It was thus concluded that a response bias explanation cannot account for the boundary shifts found with the selective adaptation paradigm. This conclusion cannot, however, be considered definitive. First of all, there exists a problem related to the degree of probabilistic asymmetry. In the Sawusch et. al. experiments, the more frequently occurring stimulus appeared only twice (voicing) or four times (place) as often as each of the rest of the test stimuli. In contrast, the adapting stimulus in a typical selective adaptation experiment can occur as much as 100 - 200 times more frequently than any other test item. Given the relatively small shifts obtained after adaptation, it is not surprising to find that no shifts resulted from such minor probability manipulations. A second, and perhaps more serious problem involves the assumption that a contrast interpretation and a feature detector interpretation are mutually exclusive. It is possible that a response bias component is confounded with the effects of feature detector fatigue in the selective adaptation procedure's single response measure. If so, then some type of response bias explanation may account for why, in Cooper's experiment, /bae/ was a more effective adapting stimulus than /p<sup>h</sup>ae/.

Regardless of whether a response bias component is involved in the selective adaptation procedure, the evidence from Cooper's crossed-vowel and crossed-consonant conditions clearly indicate the presence of a feature-specific component. In order to obtain further information concerning

the organizational properties of these feature-specific analyzing mechanisms, Cooper and Blumstein (1974) investigated the adaptation effects when the adapting and test stimuli belonged to different consonant categories. It was reasoned that perceptual shifts should be obtained if the detectors which extract place information operate irrespective of manner information. In contrast, the failure to obtain perceptual shifts would tend to indicate that the adaptation effects operate on analyzers of a more restricted nature which decode place information for each of the major consonant classes individually. To test this, identification functions were obtained for the /bae/-/dae/-/gae/ test series with five different adapting stimuli. The five adapting syllables selected were the voiced stop /bae/, the voiceless stop /p<sup>h</sup>ae/, the nasal /mae/, the voiced fricative /vae/, and the semiconsonant /wae/. Since they all contained a labial initial segment, a positive finding would be indicated by a shift in the locus of the /bae/-/dae/ phonetic boundary toward the /bae/ end of the place continuum. This prediction was partially confirmed. The results obtained with the adapting syllables /bae/ and /p<sup>h</sup>ae/ replicated the findings of Cooper (1974a). They showed a significant shift in the /bae/-/dae/ phonetic boundary, the shift being directed toward the /bae/ category. Similarly, the adapting stimuli /mae/ and /vae/ produced the expected perceptual shifts, thus demonstrating that place information is extracted from

consonants independently of their particular manner of articulation. The adapting stimulus /wae/, however, displayed only a slight, non-significant mean shift in the /bae/-/dae/ boundary, although it was in the predicted direction. Since /w/ is a semiconsonant, it was suggested that this single discrepant finding may indicate that the defining limit of the "place" analyzers is that they extract information only for the class of true consonants.

In a more recent experiment, Cooper (1974c) has investigated the effects of vowel environment on the feature-specific component of adaptation. To accomplish this, listeners were adapted with an alternating sequence of two different adapting syllables, /da/ and /t<sup>h</sup>i/, and then tested for identification with stimuli selected from two different VOT continua. One VOT series ranged from /ba/ to /p<sup>h</sup>a/ and the other from /bi/ to /p<sup>h</sup>i/. Since /da/ and /t<sup>h</sup>i/ represent both extremes of the VOT dimension, their repetitive alternating presentation should simultaneously adapt both the voiced and voiceless VOT detectors, and as such, produce no perceptual shifts along either test series. This, of course, is assuming that the effects of adaptation are operating solely on the consonant feature of voicing. The results, however, demonstrated a differential effectiveness for the two adapting stimuli as reflected by the occurrence of opposite shifts in the identification functions of the /ba/-/p<sup>h</sup>a/ and /bi/-/p<sup>h</sup>i/ series. For the /ba/-/p<sup>h</sup>a/ series, the phonetic boundary shifted toward

the /b/ category. Conversely, the /b<sup>h</sup>/ series exhibited a shift toward the /p/ category. Thus, perception of the /ba/-/p<sup>h</sup>a/ stimuli was primarily influenced by the adapting stimulus /da/, while the perception of the /b<sup>h</sup>/ series was mainly influenced by the adapting stimulus /t<sup>h</sup>/. These results indicate that adaptation operates on voicing perception in a vowel-contingent manner.

The results from Cooper's contingent-adaptation experiment makes it necessary to reevaluate the claim that the feature-specific component of adaptation is selectively sensitive to a set of "phonetic distinctive features" in the Chomsky - Halle (1968) sense. Alternatively, it is possible that the feature-specific component reflects the operation of high-level auditory detector mechanisms that are specifically sensitive to the acoustic cues which underlie the phonetic distinctive features of speech. Taking this latter interpretation, Stevens and Klatt (1974) have shown how the perception of voicing contrasts can be accounted for strictly in terms of an acoustically-based feature detector model. Similarly, the findings obtained with the selective adaptation procedure for the feature place of articulation have not distinguished phonetic similarities from the acoustic similarities that underlie them. For example, in both Cooper (1974a) and Cooper and Blumstein (1974), it was found that /p<sup>h</sup>ae/ was a less effective adapting stimulus than /bae/, when the test series

ranged from /bae/ to /dae/ to /gae/. This result can be understood in terms of an acoustic pattern interpretation by noting that /p<sup>h</sup>ae/ contained second- and third-formant transitions that were relatively weak in energy, whereas the adapting stimulus /bae/ and the /bae/ members of the test series both contained strong second- and third-formant transitions. Recall also that in the Cooper and Blumstein study, relatively strong adaptation effects were obtained for the adapting stimuli /mae/ and /vae/, but not for the adapting stimulus /wae/. This finding can be explained by examining the slopes of the formant transitions. For /mae/ and /vae/, the second- and third-formant transitions were, like the /bae/ members of the identification series, both rising and steeply sloped. The syllable /wae/, on the other hand, contained rising transitions which sloped only gradually. Thus, from the presently reported data, it is impossible to determine whether the fatigued detectors are specific for distinctive features, or rather for the acoustic cues that underlie these features.

In an attempt to distinguish between an acoustic and a phonetic interpretation, Ades (1974) examined the effects of a CV adapting stimulus on a VC identification series, and vice versa. To accomplish this, two different test continua were constructed. One series ranged from /bae/ to /dae/, and the other (its mirror-image) ranged from /aeb/ to /aed/. The set of adapting syllables consisted of the four endpoint stimuli, /bae/, /dae/, /aeb/, and /aed/.

Listeners were run in eight different experimental sessions; four sessions with the /bae/-/dae/ test series and four with the /aeb/-/aed/ series. In each session a different adapting stimulus was used. It was reasoned that if adaptation operates at a truly phonemic level, then repetition of any sound containing a /b/ should shift the ID boundary toward /b/. Likewise, repeated presentation of a sound containing the phoneme /d/ should shift the ID boundary toward the /d/ category. The results showed that adaptation with either /bae/ or /dae/ produced positive adapting effects on the /bae/-/dae/ test series. Similarly, adaptation with either /aeb/ or /aed/ produced a positive shift in the locus of the /aeb/-/aed/ phonetic boundary. However, when the adapting and test stimuli were drawn from different continua, there were no differential adaptation effects. Thus, /b/ and /d/ in final position were unable to fatigue the detectors responsible for the perception of /b/ and /d/ in initial position, and vice versa. This finding, although not strong enough to refute a phonetically-based feature detector model, clearly specifies the limits of such a model on adaptation.

If the feature detector systems responsible for the perception of speech sounds are sensitive to individual auditory features rather than for entire "linguistic features" (alla Chomsky and Halle), then it should be possible to demonstrate this fact experimentally. Recall, however, that in the Eimas, Cooper and Corbit (1973) study,

it was found that the initial 50 msec of the syllable /da/ (d-chirp) was ineffective in producing alterations in the perception of voicing contrasts. As Ades(1973) has pointed out, though, this finding can be accounted for if it is assumed that the hypothetical "voicing detectors" are only sensitive to the relative onset times of fairly sustained components. As such, the d-chirps would not convey enough information about voice to affect the "voicing detectors." With regard to the feature place of articulation, Ades(1973) has provided results which demonstrate a substantial adaptation effect with chirps (chirps were, however, significantly less effective adapting stimuli than syllables). Furthermore, since listeners found it relatively easy to identify the chirps as speech-like (e.g., b-like or d-like), adaptation was also attempted with an even less speechlike stimulus, the tweet. Tweets consisted of only F2 and F3 transitions (exactly that part of the sound that distinguishes it from the others in the place series). Again, significant adaptation effects were observed, although the effect was relatively small in comparison to those obtained with syllables and chirps.

The question of concern now is how to interpret Ades' results. Clearly, they do not suggest that adaptation is operating solely at the auditory level of perceptual analysis. Had this been the case, then syllables, chirps and tweets should all have produced equivalent adaptation effects since each contained identical acoustic information

relevant to the perception of the feature place of articulation. Especially interesting is a comparison of the effectiveness of chirps and tweets as adapting stimuli. Since the only difference between these two classes of sound patterns is the presence or absence of F1, then this component of the adapting stimulus should account for a relatively large proportion of the auditory adaptation. This conclusion, however, is unreasonable since the first formant can in no way help distinguish individual members of a place series. Perhaps, instead, Ades' findings can best be understood in terms of a strictly phonetic model of feature analyzers. It is possible to view phonetic features as purely relative events. For example, Ades' Ss reported that chirps were b-like or d-like. In this respect syllables can be considered as being more b-like or d-like than chirps, whereas tweets can be thought of as being less b-like or d-like. Thus, it is possible to construct a continuum for the phonetic features bilabial and alveolar in which the syllables /bae/ and /dae/, the b-chirps and d-chirps, and the b-tweets and d-tweets represent strong, moderate, and weak exemplars, respectively. Such an approach can easily account for the observed data. Alternatively, Ades' results are generally consistent with Cooper's (1974b) multi-component model of adaptation which states that part of the adaptation effects are attributable to the fatiguing of feature detectors at the auditory level while another part is due to the fatiguing

of feature detectors at the phonetic level. Within the framework of this approach, the tweet adaptation condition can be viewed as reflecting only the adaptation of auditory detector mechanisms. Adaptation with chirps and syllables, on the other hand, involves both the auditory and the phonetic feature detector systems. The difference in effectiveness between chirps and syllables can be explained in terms of depth of processing. For instance, syllables may also be adapting higher-level "syllable" feature detectors in addition to the auditory and phonetic feature detectors. The finding that Sg were able to identify the chirps as b-like or d-like tends to indicate that they are processed at least during the early stages of phonetic processing.

The purpose of the present experiment was to provide additional information concerning the nature of the detector mechanisms operated upon during selective adaptation. Specifically, this study attempted to resolve the issue of whether the effects of adaptation are the result of feature detector fatigue at both the auditory and the phonetic level of perceptual analysis, or alternatively, whether they only reflect the adaptation of feature detectors operating during the phonetic stage of speech processing. To accomplish this, it was necessary to construct an adapting stimulus which would preserve all of the acoustic properties underlying the perception of place information while dissociating any corresponding

phonetic information. The sound pattern chosen for this task was the "speech-embedded chirp."

The 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 the formant transitions. The major advantage of the speech-embedded chirp is that the formant transitions occur in final position. This contrasts with the role of the corresponding acoustic segment in the CV syllable where the same acoustic information occurs in initial position.

By representing the same acoustic information in different serial positions in the adapting and test stimuli, the possibility of phonetic adaptation is greatly reduced. Since the acoustic cues which underlie place information are highly context-dependent, it is very unlikely they would produce identical phonetic transformations in both initial and final position. However, even if they did, it is still unlikely that phonetic adaptation would occur. This assumption is based on Ades' (1974) findings that a phonetic segment in final position was unable to fatigue the same phonetic segment in initial position, and vice versa. Thus, it would appear reasonable to infer that any positive adapting effects obtained with speech-embedded chirps would represent adaptation operating solely during auditory analysis. The failure to find such effects, however, does not rule

---

out the possibility of auditory adaptation. A negative finding could indicate that auditory adaptation is, like phonetic adaptation, dependent upon serial position.

## METHOD

### Subjects

The listeners were five paid volunteers, all of whom responded to an advertisement in the Indiana University student newspaper. All Ss were right-handed, native speakers of American English with no known history of a hearing or speech disorder. Ss were paid at the rate of \$2.00 per hour. No S had had any prior experience with the selective adaptation procedure, although two Ss had had some previous experience with synthetic speech stimuli.

### Stimuli

All of the stimuli used in this experiment were three-formant speech patterns constructed on the parallel resonance speech synthesizer at Haskins Laboratories, and recorded on magnetic tape.

The test stimuli consisted of a series of seven synthetic CV syllables 300 msec in duration. These stimuli ranged perceptually from /ba/ to /da/. Stimuli differed from one another only in the starting frequency and direction of the second- and third-formant transitions. The different starting frequencies of F2 and F3 are displayed 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/. These fixed steady-state formants were centered at 769 Hz (F1), 1232 Hz (F2), and 2525 Hz (F3).

-----

Insert Table 1 about here

-----

In addition to the seven test stimuli, two additional stimuli were synthesized, the b-"Speech-Embedded Chirp" (b-SEChirp), and the 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 endpoint 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 Stimulus 1's and Stimulus 7's formant transitions i.e., 412 Hz (F1), 996 Hz (F2), and 2180 Hz (F3) for the b-SEChirp stimulus, and 412 Hz (F1), 1465 Hz (F2), and 3195 Hz (F3) for the d-SEChirp stimulus. Finally, the original 250 msec steady-states (i.e., the vowel /a/) were deleted. Thus, the b-SEChirp and d-SEChirp stimuli can be characterized as 300 msec sound patterns with 250 msec steady-states in initial position and 50 msec transitions in final position. Stimulus 1, Stimulus 7, b-SEChirp, and d-SEChirp are displayed schematically in Figure 1.

-----

Insert Figure 1 about here

-----

TABLE 1

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

---

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

---

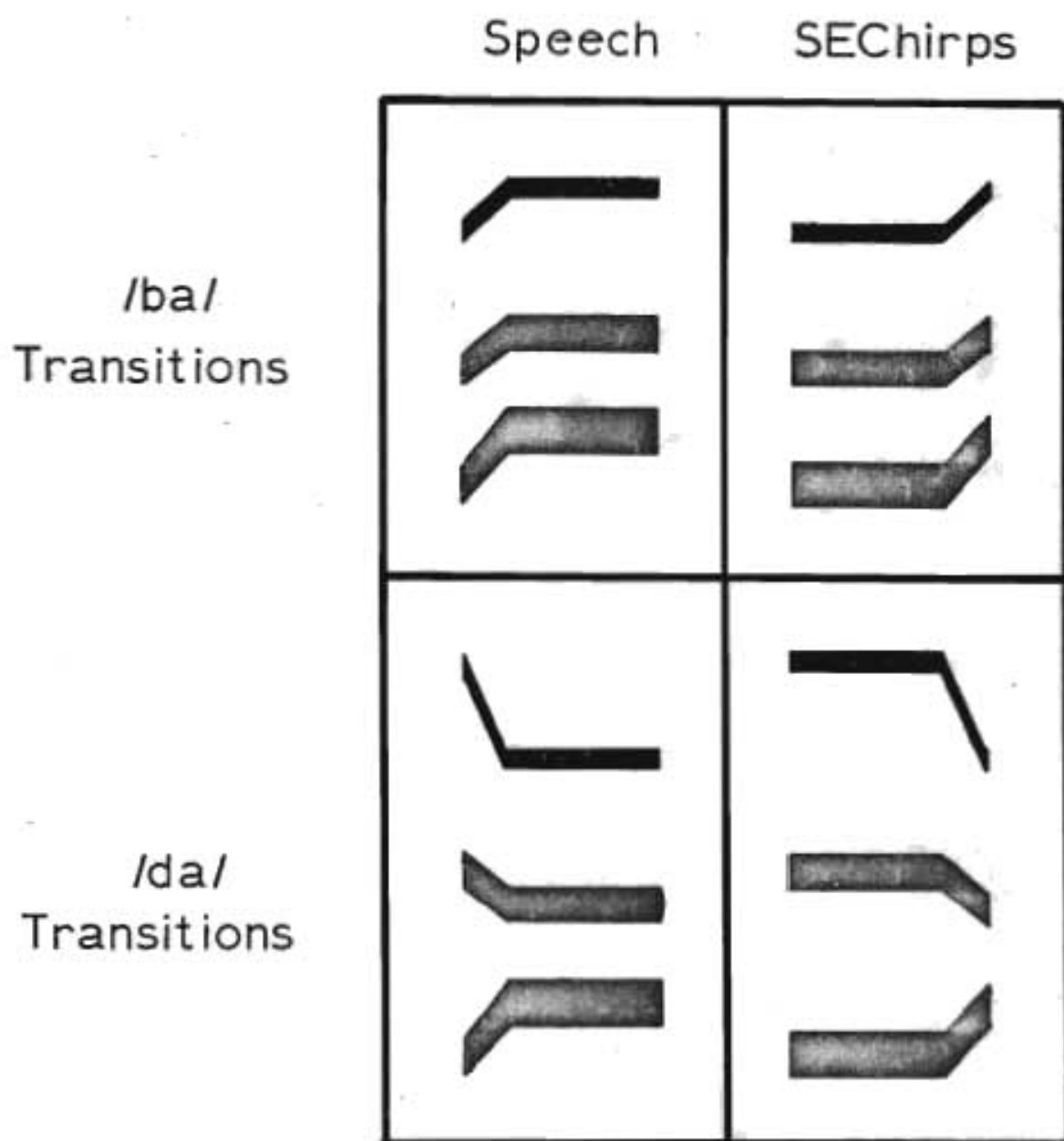


Figure 1. Schematized sound spectrograms of the adapting stimuli */ba/*, */da/*, b-SEChirp, and d-SEChirp.

Stimuli 1 and 7, and the two speech-embedded chirps were used to construct four different adaptation sequences. Each adaptation sequence consisted of 100 repetitions of the adapting stimulus with a 225 msec interstimulus interval.

### Apparatus

All experimental materials were recorded on audio tape and reproduced on an Ampex AG-500 two-track tape recorder and were presented diotically 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/cm<sup>2</sup> for a vowel-like /a/ calibration signal. Measurements were made on a Hewlett-Packard VTVM (model 400) prior to the presentation of each experimental tape. All five Ss were run together in a small experimental room.

### Procedure

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

On the day after the initial identification test, a series of four different adaptation tests was conducted, each test lasting roughly 1 hour, and taking place at 24 hour intervals. The order of presentation of the adapting sessions was as follows: Session 1. /b/ - Adapt; Session 2. /d/ - Adapt; Session 3. b-SEChirp Adapt; and Session 4. d-SEChirp Adapt.

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 (/ba/, /da/, b-SEChirp, or d-SEChirp). Following this "warm-up" period of adaptation (after Cooper, 1974a), ten 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 Ss were instructed to identify each of these five test stimuli as either /ba/ or /da/ by writing the appropriate consonant letter on response sheets. The five middle test stimuli occurred in random order with 4 sec between each. After the fifth stimulus was presented for identification, 5 sec intervened before the onset of the next adaptation trial. Each of the ten 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, Ss were given a short break after which the same tape was rewound and played again. In this manner, each of the five middle stimuli in the test series was presented for identification a total of twenty times within a single adaptation session.

## RESULTS

Table 2 shows the individual and mean phonetic boundaries for each of the five experimental sessions: one identification session without adaptation and four adaptation sessions. Each phonetic boundary was computed by finding the point along the stimulus scale which would, by extrapolation, receive 50% /ba/ responses and 50% /da/ responses. In all, there were twenty 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 S 2 who showed virtually no shift at all in the d-SEChirp adaptation test.

-----  
Insert Table 2 about here  
-----

In Figure 2 the group identification and adaptation functions, averaged over all five Ss, are plotted. It should be noted that the shifts produced by selective adaptation were not accompanied by a decline in the steepness of the response function slopes.

-----  
Insert Figure 2 about here  
-----

TABLE 2

Individual and Mean Loci of Phonetic Boundaries  
for Each Test Condition

## Syllable Conditions

Subjects	Without Adaptation	Adaptation with	
		/ba/	/da/
1	4.095	3.250	5.250
2	3.594	2.600	4.357
3	3.571	2.625	4.625
4	3.950	3.100	5.643
5	4.095	2.588	5.643
$\bar{X}$	3.797	2.766	5.133

## Speech-Embedded Chirp Conditions

Subjects	Without Adaptation	Adaptation with	
		b-SEChirp	d-SEChirp
1	4.095	3.556	4.500
2	3.594	3.412	3.588
3	3.571	3.474	4.167
4	3.950	3.556	4.286
5	4.095	3.714	4.500
$\bar{X}$	3.797	3.534	4.228

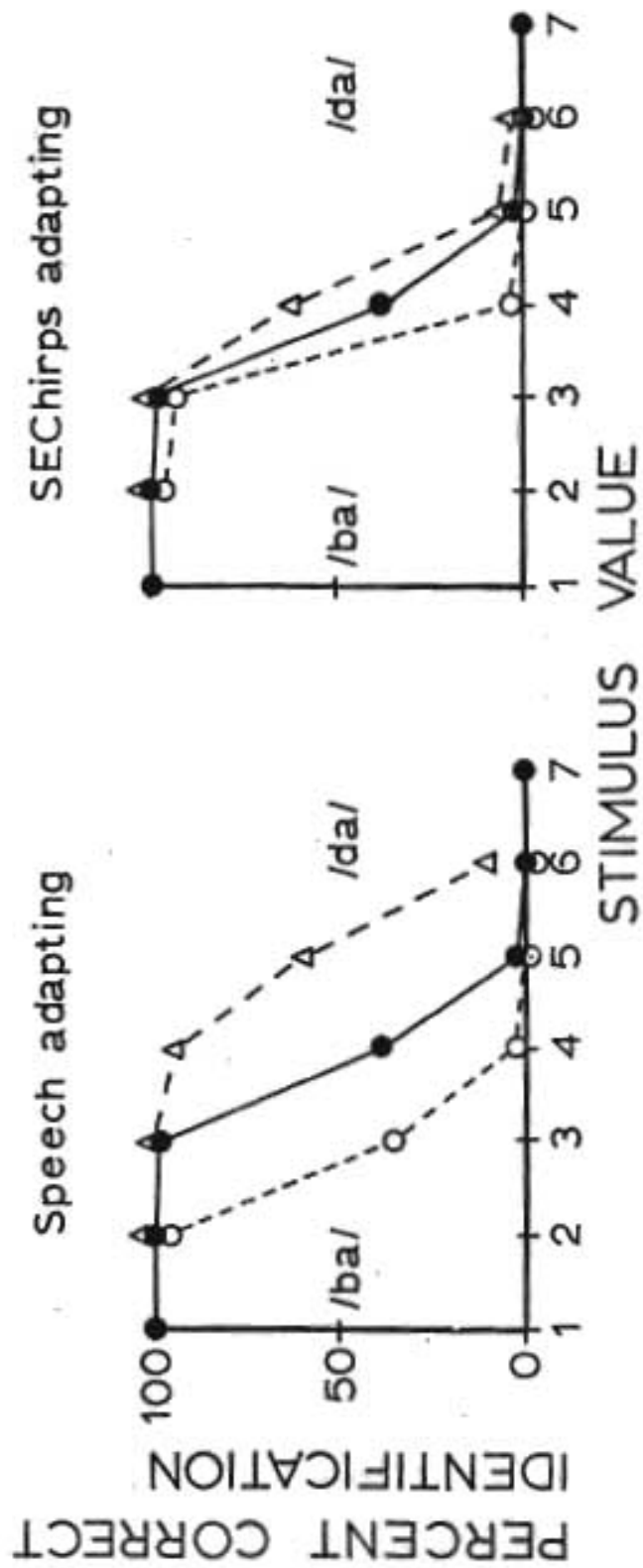


Figure 2. Group identification and adaptation functions.

Shown in Table 3 are the results of tests of significance for the shifts in the locus of the /ba/-/da/ phonetic boundary. Only one-tailed tests of significance were applied, the predicted direction of the boundary shift being toward the category of the adapting stimulus. All shifts in the phonetic boundary incurred by adaptation were found to be significant in comparison to the baseline identification data.

-----  
Insert Table 3 about here  
-----

For the /ba/ and /da/ syllable adaptation conditions, the respective mean shifts in the locus of the /ba/-/da/ phonetic boundary of 1.031 and 1.336 stimulus value units were found to be highly significant (Correlated t-tests:  $t=8.36$  for /ba/ adaptor,  $p<.005$ ;  $t=7.36$  for /da/ adaptor,  $p<.005$ ). Moreover, the magnitude of these shifts are comparable in direction and consistency to those found in previous studies which investigated adaptation effects on a place series (e.g., Cooper, 1974a; Cooper and Blumstein, 1974; Ades, 1973; Ades, 1974).

For the speech-embedded chirp adaptation conditions, the mean boundary shifts of .263 (b-SEChirp) and .431 (d-SEChirp) stimulus value units were also found to be significant (Correlated t-tests:  $t=4.02$

TABLE 3

Tests of Significance for Shifts in Phonetic Boundary  
Loci Between the Unadapted Condition  
and Each Condition of Adaptation

## Syllable Conditions

Adaptation with:	/ba/	/da/
p-value:	$p < .005$	$p < .005$

## Speech-Embedded Chirp Conditions

Adaptation with:	b-SEChirp	d-SEChirp
p-value:	$p < .01$	$p < .02$

Note: All tests were one-tailed. The direction of each significant shift was toward the category of the adapting stimulus.

for b-SEChirp adaptor,  $p < .01$ ;  $t = 3.53$  for the d-SEChirp adaptor,  $p < .02$ ). Within these test conditions, the direction of the significant shift was toward the phonetic category from which the adapting stimulus' formant transitions were originally obtained. Thus, after adaptation with the b-SEChirps, Ss made fewer /ba/ identification responses. Similarly, after adaptation with the d-SEChirps, the listeners assigned fewer identification responses to the /da/ category.

## DISCUSSION

The fact that speech-embedded chirps were able to produce alterations in the perception of a place series strongly supports the notion that one component of the selective adaptation process reflects the fatiguing of acoustically-oriented feature detectors that operate during the auditory stage of perceptual analysis. The reasoning underlying this conclusion is based on the assumption that any phonetic information conveyed by the speech-embedded chirps was irrelevant in the present task. This assumption was derived from Ades' (1974) results which demonstrated that repeated presentations of a phonetic segment in final position were unable to fatigue the feature detector mechanisms responsible for the perception of the same phonetic segment in initial position.

The validity of this study's conclusion, however, does not depend on the validity of its underlying assumption. In fact, if we reject Ades' results as the product of statistical error, and consider the possibility that phonetic adaptation was operating within the present experimental paradigm, then the results point even more definitively toward an acoustic interpretation. This is so because a phonetic account of this study's design would have predicted shifts in the locus of the /ba/ - /da/ phonetic boundary opposite

those predicted by the acoustic model for the speech-embedded chirp adaptation conditions. Thus, a phonetic interpretation would predict fewer responses assigned to the /d/ category after adaptation with the /b/ - speech-embedded chirps and fewer /b/ responses after adaptation with the /d/ - speech-embedded chirps. The logic behind these somewhat counter-intuitive predictions can best be understood by examining the acoustic structures characteristic of the phonemes /b/ and /d/ in final position. Given either the steady-state formants assigned to the b-SEChirps or the steady-state formants assigned to the d-SEChirps, the phoneme /b/ in final position would have been characterized by falling transitions in all three formants. The phoneme /d/, on the other hand, would have been characterized by rising second- and third-formant transitions, and a falling first-formant transition. In the actual b-SEChirps, all three formants were rising. In the actual d-SEChirps, the first-formant transition was rising while the second- and third-formant transitions were falling. However, since F1 does not convey any information concerning place of articulation, it would be expected that if adaptation was operating during the phonetic stage of perceptual processing then, based solely on the slopes of the second- and third-formant transitions, the listener's /ba/ - /da/ phonetic boundary should have shifted toward the /d=/ category after adaptation with b-SEChirps, and

toward the /ba/ category after adaptation with d-SEChirps. Of course, the obtained findings displayed the opposite effects indicating that adaptation was operating during the auditory stages of perceptual processing.

An interesting follow-up to the present study could attempt to replicate the findings obtained here using speech-embedded chirps with falling first-formant transitions. Then, assuming that Ades' results are not the product of statistical error, it should be possible to shift a listener's /ba/ - /da/ phonetic boundary toward the /ba/ category using an adapting stimulus with a d-like quality in final position (b-SEChirp with falling F1 transition). Similarly, it should be possible to shift the /ba/ - /da/ boundary toward the /da/ category using a b-like adapting stimulus (d-SEChirp with falling F1 transition).

Of particular interest in the present investigation was the finding that the magnitude of the shifts incurred during the speech-embedded chirp adaptation conditions closely resemble the shift magnitudes obtained by Ades (1973) after adaptation with tweets (only F2 and F3 transitions). If you recall, it was suggested that tweets represented the adaptation of only the auditory component of adaptation. As such, it would appear that both results reflect the adaptation of the same underlying perceptual mechanisms.

The existence of detector mechanisms operating at the auditory level of speech analysis has two very important implications. First, it provides the missing link in a hierarchical feature detector network such as that proposed by Hubel and Wiesel. And second, it corroborates the findings of several different paradigms which have demonstrated a distinction between auditory and phonetic stages of processing.

## REFERENCES

- Abramson, A. S. and Lisker, L. Discriminability along the voicing continuum: Cross-language tests. In Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967. Prague: Academia, 1970, 569-573.
- Ades, A. E. Some effects of adaptation on speech perception. Quarterly Progress Report of the Research Laboratory of Electronics, M. I. T., 111, 121-129.
- Ades, A. E. A study of acoustic invariance by selective adaptation. Perception & Psychophysics, 1974, in press.
- Chomsky, N. and Halle, M. The Sound Pattern of English. New York: Harper & Row, 1968.
- Cooper, W. E. Adaptation of phonetic feature analyzers for place of articulation. Journal of the Acoustical Society of America, 1974a, in press.
- Cooper, W. E. Selective adaptation to speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, and D. B. Pisoni (Eds.), Cognitive Theory, Volume I. Potomac, Maryland: Erlbaum, 1974b, in press.
- Cooper, W. E. Contingent feature analysis in speech perception. Perception & Psychophysics, 1974c, in press.

- Cooper, W. E. and Blumstein, S. E. A "labial" feature analyzer in speech perception. Perception & Psychophysics, 1974, in press.
- Eimas, P. D. and Corbit, J. D. Selective adaptation of linguistic feature detectors. Cognitive Psychology, 1973, 4, 99-109.
- Eimas, P. D., Cooper, W. E., and Corbit, J. D. Some properties of linguistic feature detectors. Perception & Psychophysics, 1973, 13, 247-252.
- Pant, G. Auditory patterns of speech. In W. Watson-Dunn (Ed.), Models for the Perception of Speech and Visual Form, Cambridge: M. I. T. Press, 1967, 111-125.
- Helson, H. Adaptation-level Theory. New York: Harper & Row, 1964.
- Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Journal of Physiology (London), 1962, 160, 106-154.
- Hubel, D. H. and Wiesel, T. N. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. Journal of Neurophysiology, 1965, 28, 228-289.
- Hubel, D. H. and Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. Journal of Physiology (London), 1968, 195, 215-243.

- Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., and Pitts, W. H. What the frog's eye tells the frog's brain. Proceedings of the Institute of Radio Engineers, New York, 47, 1959, 1910-1951.
- Lieberman, A. M. Some characteristics of perception in the speech mode. In D. A. Hamburg (Ed.), Perception and its disorders, Proceedings of the A. R. N. M. D. Baltimore: Williams and Wilkins, 1970.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-461.
- Lisker, L. and Abramson, A. S. A cross-language study of voicing in initial stops: Acoustical measurements. Word, 1964, 20, 384-422.
- Lisker, L. and Abramson, A. S. The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences, 1970, 563-573.
- Pisoni, D. B. and Tash, J. Reaction times to comparisons within and across phonetic categories. Perception & Psychophysics, 1974, 15, 2, 285-290.
- Sawusch, J. R. and Pisoni, D. B. Category boundaries for speech and non-speech sounds. Paper presented at the 86th meeting of the Acoustical Society of America, November, 1973, Los Angeles.

- Sawusch, J. R., Pisoni, D. B. and Cutting, J. E.  
Category boundaries for linguistic and non-  
linguistic dimensions of the same stimuli.  
Paper presented at the 87th meeting of the  
Acoustical Society of America, April, 1974,  
New York City.
- Stevens, K. N. and Halle, M. Remarks on analysis  
by synthesis and distinctive features. In  
W. Watson-Dunn (Ed.), Models for the Perception  
of Speech and Visual Form, Cambridge: M. I. T.  
Press, 1967, 88-102.
- Stevens, K. N. and House, A. S. Speech Perception.  
In J. V. Tobias (Ed.), Foundations of Modern  
Auditory Theory, New York: Academic Press,  
1972, 1-62.
- Stevens, K. N. and Klatt, D. H. Role of formant  
transitions in the voiced-voiceless distinction  
of stops. Journal of the Acoustical Society  
of America, 1974, 55, 3, 653-659.
- Studdert-Kennedy, M. The perception of speech.  
In T. A. Sebeok (Ed.), Current Trends in Lin-  
guistics, Volume XII, The Hague: Mouton, 1974  
(Also appeared in Haskins Laboratories Status  
Report on Speech Research, SR-23, 1970, 15-48).

- Studdert-Kennedy, M. and Shankweiler, D. Hemispheric specialization for speech perception. Journal of the Acoustical Society of America, 1970, 48, 579-594.
- Studdert-Kennedy, M., Shankweiler, D. and Pisoni, D. B. Auditory and phonetic processes in speech perception: Evidence from a dichotic study. Cognitive Psychology, 1972, 3, 455-466.
- Wood, C. C. Parallel processing of auditory and phonetic information in speech perception. Perception & Psychophysics, 1974, in press.