COGNITIVE PSYCHOLOGY 11, 38–57 (1979)

Adaptation of the Category Boundary between Speech and Nonspeech: A Case Against Feature Detectors

ROBERT E. REMEZ
Haskins Laboratories and University of Connecticut

Two experiments were performed employing acoustic continua which change from speech to nonspeech. The members of one continuum, synthesized on the Pattern Playback, varied in the bandwidths of the first three formants in equal steps of change, from the vowel /a/ to a nonspeech buzz. The other continuum, achieved through digital synthesis, varied in the bandwidths of the first five formants, from the vowel /æ/ to a buzz. Identification and discrimination tests were carried out to establish that these continua were perceived categorically. Perceptual adaptation of these continua revealed shifts in the category boundaries comparable to those previously reported for speech sounds. The results were interpreted as suggesting that neither phonetic nor auditory feature detectors are responsible for perceptual adaptation of speech sounds, and that feature detector accounts of speech perception should therefore be reconsidered.

Recent explanations of speech perception have been heavily influenced by neurophysiological lore. In a general sense, this influence is entirely proper, for acoustic signals which possess biological significance are felt to be natural entities which owe their existence, as do hands or eyes, partly to hereditary factors (Lenneberg, 1960; Hoy & Paul, 1973; Lieberman, 1975). Because the genetically ordained neural mechanisms which support language generation and comprehension may exercise powerful constraints on the structure of communication (Fodor, 1966), the psychology of speech perception must at least be compatible with the underlying biological principles if it hopes to embrace *natural* language phenomena. Nevertheless, the argument of this paper is that the modeling of speech perception after the specific example of single-unit electrophysiology (originally, Eimas & Corbit, 1973) has been an unsuccessful

Address correspondence to Robert E. Remez, who is now at the Department of Psychology, Indiana University, Bloomington, IN 47401.

During the course of this research and report the author was supported by an NICHD Predoctoral Training Grant in Language and Psychology awarded to the University of Connecticut, and by NICHD Program Project Grant HD-01994 awarded to Haskins Laboratories.

It is a pleasure to thank Tom Casola, Harriet Greisser, Alvin Liberman, Ignatius Mattingly, David Pisoni, Diane Kewley-Port, Philip Rubin, Michael Studdert-Kennedy, Quentin Summerfield, Michael Turvey, Hal Tzeutschler, and Robert Verbrugge for their aid, comment, and criticism during the development of this study and report. Also, a special thank you goes to Agnes McKeon, who showed me how to paint Playback-style.

0010-0285/79/010038-20$05.00/0

albeit very popular attempt at explaining the perception of speech in a biologically relevant way. In this kind of perceptual explanation, the primitives of speech perception are assumed to be a set of specialized neurons whose sensitivities are phonetically appropriate extensions of the restricted "stimulus preferences" of sensory neurons of animals (e.g., Lettvin, Maturana, McCulloch, & Pitts, 1959; Hubel & Wiesel, 1965; Whitfield & Evans, 1965; Capranica, Frishkopf, & Nevo, 1973; Winter & Funkenstein, 1973). Although objections to similar pandemonium renditions of perception have been raised on grounds of analytic effectiveness (Neisser, 1967) and physiological appropriateness (Pribram, 1971), the common notion that speech signals use distinctive features provided some immunity from those general criticisms. On the independent validity of the binary distinctions of phonetic analysis (Jakobson, Fant, & Halle, 1963) it seemed reasonable to describe speech perception as the detection, by opponent-process means, of a phonetic message carried by an acoustic signal. However, several other considerations contributed as well to the initial attractiveness of the phonetic feature detector notion.

First, the correspondence between abstract phonetic segments and their acoustic realization appeared to be other than a straightforwardly isomorphic one (Halle & Stevens, 1962; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The very response of the perceptual system seemed to be nonmonotonic with respect to changes in the acoustic pattern (Liberman, Harris, Hoffman, & Griffith, 1957; Pisoni, 1971). The extraction of meaning from the rapidly fluctuating pressure wave appeared to require the resort to heuristic processes which took place at levels of patterning of a higher order than the simply acoustic; translation from acoustic pattern to phonetic message was portrayed as involving elaborate inference of phonological or physiological content. A device "tuned" to the phonetic structure of the pressure wave would obviate much of the hypothetical mental effort required by the motor theory or the analysis by synthesis models, yet it would pursue its course at the requisite higher level.

Second, a century of effort directed toward delineating the cortical loci of mental faculties established that speech and language are mediated in a restricted cortical area (Penfield & Roberts, 1959; Geschwind & Levitsky, 1968; Witelson & Pallie, 1973). Perhaps the perception of speech, as well as syntactic language, was managed by these specialized regions of the brain; the implication of feature detectors in phonetic perception is that the neurology of speech is as specialized as the neurology of language, a possibility encouraged by the findings involving receptive field preferences in animals (e.g., Capranica et al., 1973).

Third, the nativist position of the generative grammarians argued that rather detailed linguistic "knowledge" might be prewired into every human infant (Chomsky, 1965; 1968). Although the particular contribution of the grammarians was to begin the formalization of the intuitions of native speakers, thereby characterizing the particular combinatorial preferences of the natural language system, their work implicitly supported the notion that human language was "species specific;" this lent credibility to the proposal of a specialized speech neurophysiology. The fact that human languages make use of a small, common set of sound contrasts can be taken as evidence both for specific inherited phonetic dispositions and for the specialization of the neurological mechanism which handles perception in the linguistic mode.

Fourth, perceptual experiments on neonates seemed to show that infants were sensitive to some phonetic distinctions before any relevant experience (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Because these infants were perceptually acute before they were comparably refined in speech production, a dissociation between production and perception was evident; an explanation involving a specialized neural mechanism which required neither entrainment nor reafferent information appeared to be warranted. A battery of genetically pretuned hypercomplex cells (i.e., feature detectors) which translates auditory events into phonetic descriptions addressed this, and the foregoing, issues in a parsimonious and elegant fashion.

Following the demonstration of McCulloch (1965) and the rationale of Weisstein (1969), the experimental test of the hypothesized detectors involved the adaptation of phonetic category boundaries.[1] On the premise that the output of an array of detectors could be biased through the fatigue of selected detectors within it, the difference between fatigued and unfatigued performance would reveal the sensitivities of the fatigued portion of the detector array. If the response of the system was found to alter along the dimensions of distinctive feature analysis, then the hypothesis of phonetically tuned detectors would be supported.[2] The actual data on fatigue-induced change in the speech perception system, however, only marginally confirm the original description of phonetic feature detectors. Although certain adaptation effects have required the explanation to include a phonetic level of analysis at which particular acoustic-auditory values are less perceptually significant (Ades, 1974; Diehl, 1975; Miller, 1975; Remez, Cutting, & Studdert-Kennedy, Note 2), other research has

[1] Haggard (1967) first described the paradigm of after-effect research employing speech sounds, but his rationale was completely independent of neurophysiological claims.

[2] It should be noted that evidence taken to support the reality of distinctive *features* [e.g., Miller & Nicely (1955), and Fromkin (1971)] does not also constitute evidence for the reality of feature *detectors*. Were this not the case, to take a ludicrous instance, then evidence for the reality of UFO's would also be evidence for the existence of UFO detectors. Clearly, the issues of existence and of physiology of detection are separate, empirical issues in both cases.

produced evidence of nonphonetic adaptation which is fully compatible with any of the previously obtained phonetic effects (Ades, 1974; Pisoni & Tash, 1975; Bailey, Note 1; Diehl, 1976). In short, the present situation is paradoxical. While the underlying detectors are phonetic by the original intention as well as by occasional necessity, some of them may suffer acoustic fatigue, and all suffer from inexplicit specification of their tuning curves. [In tonotopically organized cells, the dimension for measuring sensitivity is frequency (Woolsey & Walzl, 1942); in phonetically organized detectors, the dimensions of analysis must correspond to those of vocal production, and many of these have yet to be defined.] Additionally, and perhaps fatally, the passive filtration method of phonetic feature extraction in speech assumed that there is a simple relation between aspects of the acoustic pattern and the phonetic segments to which they correspond perceptually. But it is the fundamental point of many speech perception studies that the information which specifies segmental identity is typically carried by the sound pattern distributed across the entire syllable (e.g., Cooper, Delattre, Liberman, Borst, & Gerstman, 1952). This requires, in essence, that each feature detector be a little homunculus, omniscient on the nature of the context conditioned variation of its favorite feature of speech; a detector could not be passive, but would have to creatively anticipate the portion of the signal containing the information specific to its judgment. Such feature-by-feature recognition schemes ignore the well-documented facts of parallel, integral transmission of information which gives the speech wave its robustness and efficiency (Liberman et al., 1967; Liberman & Studdert-Kennedy, 1978). For these reasons, phonetic feature detection can approximate the phenomena of speech perception only by sacrificing the simplicity of the conception in favor of multiple tuning arrangements. This modification permits the sensitivities of feature detectors to vary with context, although, since the context determines the information for phonetic identity, and the interacting phonetic segments determine the context, this variant of feature detection may *assume* most of the problem of recognition via the legerdemain of "context-sensitivity;" surely, it can be no simpler to *identify* a context of this kind than to *identify* an element within it. The hypothesis of multiple tuning, then, is equally promissory as the simpler, original conception, and it, too, does no more than restate the classic problem, of translating from the acoustic domain to the phonetic, in terms of conceptual physiology. Further, Halwes and Jenkins (1971) have argued that this kind of associative mechanism, even if it could account for the phenomena of concern, would be computationally unwieldy, and unnatural.

In the light of the reservations expressable about the hypothesis of phonetic feature detectors, it would be appropriate to inquire whether the experimental confirmations of that hypothesis may have been partially misconstrued. Perceptual adaptation of a category boundary might not necessarily reflect an underlying process of tuned detectors. One test of the hypothesis of a linguistic organization for the detectors would be to demonstrate adaptation outside the set predictable from the inventory of distinctive features of spoken languages. The experiments reported here, which use acoustic continua from speech to nonspeech sounds, were an attempt to satisfy this condition. Further, it is argued that the observed adaptation effects cannot simply be reduced to a process of auditory detector fatigue, suggesting a perceptual basis for the effects rather than a sensory one. Perceptual adaptation observed under these conditions, it is argued, is compatible with a phonetic level of analysis beyond the auditory; that is, it would not conflict with the body of speech perception research which motivates abstract analysis to describe perceptual constancies. However, phonetic and auditory feature detectors would have to be ruled out as an explanation of speech perception and of phonetic category boundary adaptation.

EXPERIMENT Ia

The manufacture of a synthetic acoustic continuum, in which one endpoint was the vowel /ɑ/ and the other endpoint a nonspeech buzz, is described in Experiment I. Before testing for adaptation effects on the placement of the category boundary, in Experiment Ib, identification and discrimination data were obtained in Experiment Ia to determine the status of this continuum with regard to categorical perception. This was done to provide assurance that any boundary shifts occasioned by adaptation would be due to a perceptual process rather than to changes in a judgmental criterion (Sawusch, Pisoni, & Cutting, 1974).

Methods

Subjects. Sixteen University of Connecticut undergraduates, whose participation fulfilled the introductory psychology course requirement, served as listeners in this part of the study. All were native English speakers with no known speech or hearing disorder or psychopathology. None had any experience with synthetic speech sounds before the listening session.

Stimuli. The Haskins Laboratories Pattern Playback (Cooper, Liberman, & Borst, 1951) was used to synthesize the basic materials.[3] This device uses a tone wheel to generate the harmonics of 120Hz in light intensities arrayed in a frequency scale. A graphic pattern selectively reflects portions of this scale, and this reflection, through capture by a photocell, is transduced to a frequency by amplitude by time acoustic signal. Figure 1 displays the pattern painted (Liquitex Acrylic Titanium White; Grumbacher #4) on the acetate belt (Eastman Kodak), and the frequency values of the transduced signal. The pattern changes from a vowel /ɑ/, with formant frequency values of 600Hz, 1200Hz, and 2400Hz, respectively, to a nonspeech buzz, by modifying the bandwidths of the formants; initially, the

[3] Neither the Haskins Parallel Resonance Synthesizer nor the Ove III were suitable for this study because of hardware-imposed limits on formant bandwidth. These devices are devoted speech synthesizers, and this study required a full-frequency synthesizer (i.e., one with no such restriction).

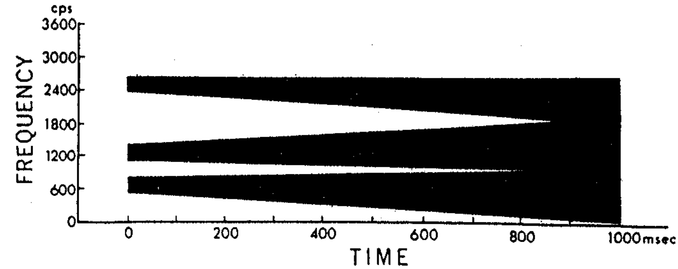

FIG. 1. The painted pattern which the Playback reproduced.

bandwidths of all three formants were 100Hz, and they increase to effectively infinite width at the end of the pattern. Figure 2a presents one spectral section through each of the endpoints.

This 1-sec sound was transferred to audiotape and then digitized by the Haskins Pulse-Code Modulation System (PCM) (Cooper & Mattingly, 1969), sampling at 10KHz with low pass filtering at 5KHz. A 10-step stimulus continuum was then made by editing the original digitized waveform. Nine cuts were made, one every 100 msec; the oscillographic patterns were equated for amplitude, producing 10 tokens, each of 12 pitch periods which vary in formant bandwidth as does the overall pattern.

Two test tapes were then created using the PCM system. The identification sequence contained 10 occurrences of each of the 10 continuum items, for a test of 100 trials, with 5 sec between trials, and 9 sec following each decade. The discrimination sequence consisted of ABX triads with 1 sec between items, 5 sec between trials, and 9 sec separating the decades. The four permutations of each ABX comparison were represented: ABA, ABB, BAB, and BAA. In a one-step discrimination, the comparisons are Items 1 and 2, 2 and 3, 3 and 4, and so on; at four trials per comparison, and nine comparisons, there were a total of 36 trials in this test.

Procedure and apparatus. The 16 listeners were tested in four groups of four subjects each. Sounds were presented binaurally over Grason–Stadler earphones activated by a Crown 820-144 tape recorder through a junction box so that several listeners could listen simultaneously. Each session commenced with a briefing sequence in which the endpoints of the continuum were repeated 10 times each in alternation. At that time, listeners were asked to signify that they had a good idea of what the sounds were; their instructions were to consider the buzz a machine noise, and the vowel a synthetic speech sound. The identification test was then administered. Identifications were scored on a response sheet as speech or buzz (S or B). After a short intermission, listeners were given sample ABX sequences in which they judged which of the first two sounds was identical to the third; the continuum endpoints were used here to insure clarity of the instructions. The actual test, begun when all agreed that they understood the instructions, consisted of the 36 trial discrimination sequence played twice.

Results and Discussion

Three subjects were dropped because they either failed to follow instructions (two subjects declined to judge difficult items) or responded at chance on the identifications (one subject). Results for identification and discrimination appear in Fig. 3a. Each point is the mean of 130 observations in the identification test, and 104 observations in the discrimination test. These functions are reasonably consistent overall with the criteria for categorical perception (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970), in that a peak in discriminability occurs at the crossover

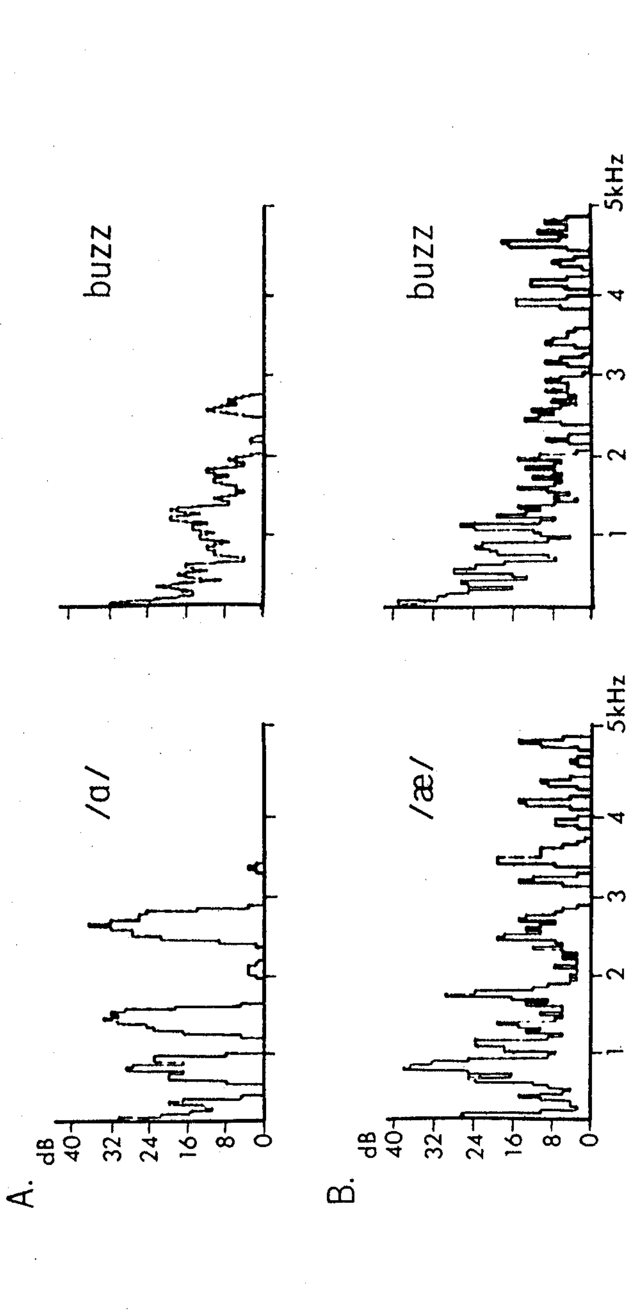

FIG. 2. Spectral sections through the endpoints of the Playback synthesized continuum (a), and the software synthesized continuum (b).

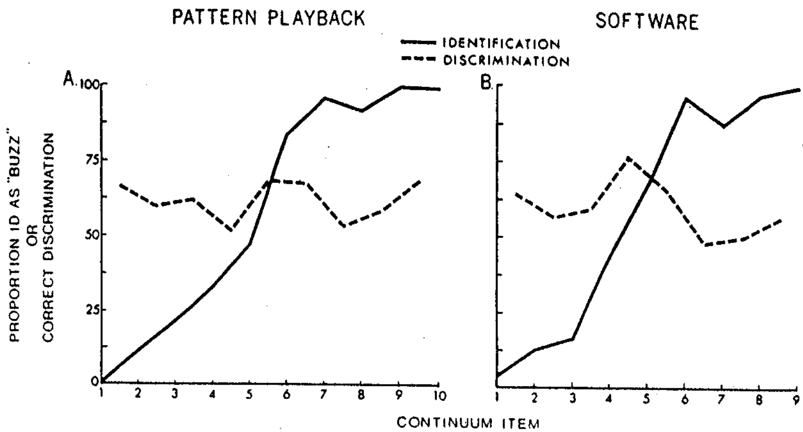

FIG. 3. Identification and discrimination plots for Playback (a) and software (b) continua.

between the identification categories. The term "categorical perception" describes a situation in which the judged difference between two entities is contingent on their identities rather than on the physical differences between them.

One anomaly in the discrimination function should be addressed, namely, the troughs of the function. Given two categories, speech and buzz, the peaks should number one, not three as shown here; in the function of Fig. 3a, the discrimination peaks between Items 1 and 2, and between 9 and 10, are as prominent as the category boundary peak. Examination of the wave-forms of these tokens revealed amplitude differences between two of the 12 pitch periods in each item of the pairs. No such difference could be discovered between items in the other pairs in the set. One possible account, then, of the spurious peaks in the discrimination function is that they result from judgments of amplitude rather than spectrum differences. If this reasoning is correct, the peaks can be discounted in any challenge to categoricity, because the manipulation of interest is spectral, and this particular discrimination is made on a nonspectral basis. We may also note that in an additional check of categoricity, each subject's identification data was used to calculate a predicted discrimination function (Liberman et al., 1957). From these predictions, a series of chi-square tests was used to evaluate the goodness of fit of the predicted and obtained discrimination performance for each individual subject; six subjects' performance departed significantly from the predictions ($p < .02$). With the understanding, then, that slightly fewer than half the subjects did not show categorical perception, but that the presence of an amplitude cue means that this failure may be potentially discounted, the test of perceptual adaptation was performed.

EXPERIMENT Ib

In this part of the experiment, the synthetic /ɑ/-buzz continuum was used in the adaptation paradigm of speech research. Given the finding of Experiment Ia, of qualified support for the claim of categorical perception of the continuum, there was some justification for expecting that any adaptation effects would be due to perceptual rather than response-criterial alterations. Further, because the continuum does not change from one proper speech sound to another, a shift in the category boundary due to adaptation would be difficult to explain by the notion that adaptation is a matter of selective fatigue in a set of detectors tuned to phonetic features.

Methods

Subjects. Eight University of Connecticut undergraduates were paid to listen in this part of the study. They had all participated in Experiment Ia (five from the original group could not attend the listening sessions for scheduling reasons).

Stimuli. The 10 tokens from Experiment Ia were used. An adaptation sequence consisted of an initial 100 repetitions of the adapting item, one of the continuum endpoints, at 1-sec intervals. After a 10-sec pause, which cued the listeners that the identification trials were coming up, six items from the continuum were presented in random order for identification, as either speech or buzz (S or B). At the conclusion of the block of six, there was a 10-sec pause followed by 50 repetitions of the adapting item, another block of six, and so on for the remainder of the test.

Each of the 10 sounds drawn from the continuum was presented for identification 12 times, with the exception of the four most extreme, the two on each end, which were presented six times each. This preserves sensitivity in the midrange of the continuum and shortens the test by two blocks of identifications, to the relief of the listener. With 96 trials (6 twelves and 4 sixes) there were 19 blocks of six trials each. The random order for these items was the same in both the speech and the buzz adaptation sequences.

Procedure. An identification sequence was used to determine a baseline identification function in each of the two sessions. This was used for comparison with the identification function obtained after adaptation. All subjects took part in both conditions; half took the /ɑ/ adaptation test first, half took the buzz adaptation test first. Several days separated the test sessions. The equipment and test conditions were in all other respects the same as in Experiment Ia.

Results

Each subject contributed two sets of judgments per session, a baseline pretest set and an adaptation set. To each of these sets of data a standard ogive was fitted, after Woodworth (1938). Thus, two scores were available for each subject per test, one mean of the fitted ogive, measured in continuum units, for the pretest, and one for the adaptation test.

The curves for the grouped data for each session appear in Fig. 4. Each pretest plot represents the means of 80 trials per continuum item; in the adaptation plot, the two extreme points on either end, Items 1, 2, 9, and 10, are the means of 48 observations each; the remaining six medial points are the means of 96 judgments each. The change in the ogive mean due to adaptation with the vowel was 1.32 continuum units, whereas that due to adaptation with buzz was .638 continuum units in the opposite direction. Table 1 summarizes the values of the ogive means for pretest and adaptation conditions for each subject.

A two-factor repeated measures analysis of variance was performed on the ogive means, with two levels of Adaptor (SPEECH or BUZZ) and two of Condition (PRETEST or ADAPTATION). There were no significant main effects, indicating that the means, when collapsed across conditions or across adaptor, did not depart from the grand mean. However, the interaction of Adaptor-by-Condition was significant [$F(1, 7) = 12.771, p < .01$), reflecting the different, and opposing effects of each adaptor, relative to the pretest mean, on the adaptation test mean.

Discussion

The shifts of the category boundary between /ɑ/ and buzz after adaptation in this experiment are damaging to the proposal that the units mediating these effects in speech perception are isomorphic with the primitives of phonetic feature analysis. Miller (1975), for example, has proposed that feature analyzing devices are arranged so that the fatigue of one leads to

TABLE 1

PRETEST AND ADAPTATION TEST OGIVE MEANS BY SUBJECT FOR EXPERIMENTS Ib AND IIb

Experiment Ib	/ɑ/ Adaptor		Buzz adaptor	
Subject	Baseline	Adaptation	Baseline	Adaptation
1	4.292	0.855*	4.537	5.318*
2	4.225	4.859	3.410	3.368
3	4.210	2.748*	3.468	5.179*
4	4.159	1.581*	3.400	4.205*
5	5.486	3.825*	5.486	5.278
6	4.969	3.484*	4.733	5.227*
7	5.337	5.304*	4.713	6.169*
8	4.684	4.115*	4.336	4.667*

* Change in mean in the expected direction.

Experiment IIb	/æ/ Adaptor		Buzz adaptor	
Subject	Baseline	Adaptation	Baseline	Adaptation
1	3.578	2.990**	4.158	4.586**
2	3.836	2.732**	4.336	5.199**
3	3.799	1.612**	3.798	4.534**
4	3.529	1.693**	3.431	3.663**
5	4.200	2.198**	3.282	4.323**
6	3.984	3.404**	3.984	4.529**
7	4.293	4.249**	4.707	5.187**
8	1.705	1.691**	2.108	3.327**

** Change in mean in the expected direction.

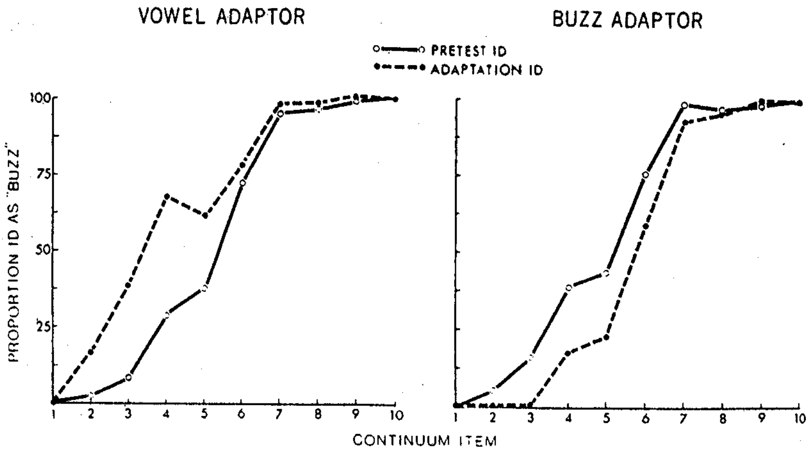

FIG. 4. Pretest and adaptation test curves for the Playback continuum.

the relatively enhanced strength of its inverse or opponent. But, in the case of /ɑ/ fatigue, the resulting change requires going outside the feature set used in phonetic descriptions to capture the distinction between speech and nonspeech. In other words, a processing device arranged along the lines of phonetic features could neither predict nor explain this case of adaptation.[4]

Must we then have recourse to a "lower level" account of the results? Other authors have found higher order phonetic descriptions unwarranted by the adaptation effects (Pisoni & Tash, 1975; Bailey, Note 1; Ades, 1977), and since higher order phenomena can be described in lower order physical terms, they have ascribed their effects to alterations in sensitivity at a lower level within the auditory system. For example, a receptive unit with a "best" stimulus may tend to show, under fatigue, a decreased sensitivity to that stimulus; its decrease leads to the release from inhibition of adjacent, similar receptors, and consequently, to the relative enhancement of near misses from the best value. Thus, the auditory system,

[4] Morse, Kass, and Turkienicz (1976) found that an /ι/ or /ε/ adaptor, but not /ɪ/, changed both boundaries on an /ι/-/ɪ/-/ε/ continuum: they concluded that the binary distinctions "tense" and "high" of Chomsky and Halle (1968), which would specifically rule out such a finding, had been empirically falsified. In replacement, they offered that, because their results had shown continuity rather than discreteness, the feature system underlying the result was necessarily continuous, many-valued rather than binary; by extrapolation, so was the detector. Their approach in this matter was like that of Cooper and Blumstein (1974), who were the first to use adaptation to ferret out perceptual interactions and thereby to define the phonetic features perceptually rather than acoustically or articulatorily. Nevertheless, because no language makes phonetic use of the distinction [+speech, −speech], it is safe to say that /ɑ/-buzz adaptation does not require that we posit a new feature; rather, it properly undermines the interpretation of speech adaptation effects in terms of phonetic features or of phonetic feature detectors.

early on in the course of an analysis, can yield a mistransformed description which the unsuspecting analyzers in the next step are helpless to reverse. [The actual neurophysiology of this is still open to question. One current topic of investigation is whether the reciprocal inhibition demonstrable at the VIII nerve nucleus arises cochlearly or in the nucleus itself (Mountcastle, 1974)]. By this mechanism, then, a unit or units sensitive to a portion of the frequency range, when fatigued, will be less sensitive to the absolute values of stimulation, and will, via disinhibition, effectively amplify departures from the original fatigued values. To take an example: a receptor which mediates a rising second formant value over the course of 35 msec (which specifies a voiced bilabial stop consonant in some circumstances) will be less sensitive, when fatigued, to values which exactly conform to the pattern of fatigue. Disinhibition of neighboring receptors would, in effect, boost receptor responses to second formant transitions which depart by small amounts from the fatigued values. The essential point of this type of explanation is to insist that the auditory transcription which the phonetic system is given to work with has been irretrievably altered in this manner.

However, if this reasoning is applied to the adaptation by /ɑ/ and buzz, a curious situation arises. Fatigue caused by buzz should decrease sensitivity throughout the range of frequencies used; no frequency-specific effects should occur. Indeed, the auditory view of adaptation would predict *no adaptation* at all. On the other hand, fatigue caused by /ɑ/ should reduce the sensitivity of the "neural spectogram" at 600Hz, 1200Hz, and 2400Hz. If a listener were then presented with the buzz, he should hear a pattern the inverse of /ɑ/, with formants at 300Hz, 900Hz, and 1800Hz.[5] In this case, if the listener judges the sound on the basis of the acoustic feature of presence or absence of formant structure, then the auditory point of view predicts that the boundary should move toward the buzz, since the fatigued spectral receptors, even in this extreme case, might be expected to retain a pattern showing acoustic maxima and minima. However, precisely the reverse boundary movement was actually observed in these data, suggesting that a lower level account of the phenomenon is unacceptable.

The present experiment therefore produces a situation unique in the adaptation literature. While the conventional approach has been to suspect auditory processes by default whenever a phonetic account of adaptation fails, this is obviously not possible here. The listener must be judging the sounds on other than the simple acoustic basis of presence or absence of peaks in the neural spectrogram. The perception of the novel distinction between /ɑ/ and buzz (a speech vs nonspeech opposition) indicates that a fixed perceptual capability, one which cannot adopt parame-

[5] This pattern, when rendered by the Playback, does not sound like a speech sound.

ters to suit the demands of a particular situation, must be abandoned. Neither phonetic feature nor acoustic property detectors can be reconciled with this type of perceptual versatility, since, in addition to their inflexibility, both accounts make incorrect predictions here.

Based on the foregoing results, there are several motives for extending this line of investigation. First, the anomalous discrimination peaks mildly threaten the claim of categoricity for this perceptual distinction. By implication, the explanation that the adaptation effect is judgmental rather than perceptual cannot be confidently rejected. Second, the artificiality of the speech synthesized on the Pattern Playback may be a factor to consider. Because the Pattern Playback, however phonetically identifiable its message may be, has a voice quality unlike that of any person, its use in an experiment of this kind may produce synthesizer artifacts. As a precaution, then, it would be valuable to try this procedure with a more natural sounding synthesizer. Finally, the possibility that this effect is restricted to /ɑ/, that /ɑ/ might intrinsically be more nonspeechlike than other vowels, could be assessed by using a different vowel in the same adaptation paradigm. On these accounts, Experiment II was performed.

EXPERIMENT IIa

Methods

Subjects. Eight University of Connecticut undergraduates, not those of Experiment I, were paid for their participation. All were naive with respect to synthetic speech.

Stimuli. The software synthesizer of Fisher and Engebretson (1975), modified to permit variable parameterization of all five formants, was used to make the acoustic tokens. This program calculates a digital waveform from user-determined parameters of source frequency, formant frequency and bandwidth, and overall duration and amplitude. The digital waveform is then converted to audio via a digital-to-analog converter. These programs, implemented by Joe Kupin and Hal Tzeutschler, run on the University of Connecticut "Language and Psychology" Data General NOVA 2 computer.

A nine-step continuum from /æ/ to buzz was made by successive 50Hz increments in the formant bandwidths starting from an initial bandwidth of 100Hz for each formant. Duration was 140msec; overall amplitude was 45dB; fundamental frequency was 120Hz; the frequencies of the first five formants for the vowel were 750, 1650, 2460, 3500, and 4500Hz, respectively. The audio output was transferred to the Haskins PCM system via Ampex tape recording, to permit algorithmic envelope shaping. Each item was 16 pitch periods (140msec) long, with ramp on and off of three periods (25msec); overall amplitudes were equated. Spectral sections through the endpoints appear in Fig. 2b.

The identification test consisted of 10 judgments of each of the nine items in random order. The discrimination test consisted of eight judgments of each of the eight one-step comparisons, in random order.

Procedure and apparatus. The outline of Experiment Ia was followed.

Results and Discussion

Figure 3b displays the functions for identification and discrimination. The identification plot displays the means of 80 trials per point, the discrimination plot the means of 64 trials per point. Inspection of the figure will reveal that, relative to Experiment Ia, the peak at the category boundary remains a property of the discrimination function, while the peaks at the extremes of the continuum have disappeared at the buzz end, and all but disappeared at the speech end. Of the eight participants, seven produced discrimination data consistent with the prediction calculated from individual identification performance. It is reasonable, then, to conclude that this more carefully controlled stimulus continuum elicited a more convincing demonstration of categorical perception.

EXPERIMENT IIb

Methods

Subjects. The eight listeners from Experiment IIa took part in this section of the study. They were paid for their time.

Stimuli. The nine-item stimulus continuum from Experiment IIa was used to make the adaptation sequences. These tests differed from Ib only in the consequences of using a nine-, as opposed to a 10-, step continuum. Here, the four most extreme items were presented six times each for identification during adaptation, and the remaining five medial items 12 times each. With 84 trials overall (4 sixes and 5 twelves) there were 14 blocks of six trials each, which alternated with the repeating adaptation item, either /æ/ or buzz. The random order of identifications during adaptation was the same in the speech and buzz adaptor conditions.

Procedure. As in the previous experiment, each testing day began with the presentation of the identification sequence to obtain baseline data for comparison with the adapted identification. All subjects took part in both adaptation conditions, counter-balanced for order. Test days were consecutive.

Results

The ogive fitting method was again used on the two tests per day contributed by each subject. Individual subject means are presented in Table 1. Averaged functions for both adaptation conditions appear in Fig. 5. Pretest plots show the means of 80 trials per continuum item, adaptation plots show the means of 48 trials for Items 1, 2, 8, and 9 and 96 trials for Items 3 through 7. The change in the ogive mean due to adaptation with the speech endpoint is .883 continuum units, whereas adaptation with the buzz endpoint was .692 in the opposite direction.

An analysis of variance was performed on the ogive means, with two levels of each factor, Adaptor (SPEECH/BUZZ) and Condition (PRE/POST). The interaction of Adaptor-by-Condition is the term of interest here [$F(1, 7) = 32.842$, $p < .001$]. The main effect of adaptor was also significant [$F(1, 7) = 65.858$, $p < .001$]. The statistical significance of the adaptor term was due to the close correspondence of the two pretest means, which, when averaged with the adaptation means, clearly reflect the differential effects of adaptation. (Experiment Ib failed to show such significance for this term in the analysis because the pretest means varied in opposition to the adaptation means, thus cancelling the effect of adaptor upon averaging.)

Discussion

This study with software synthesized sounds strengthens the original argument made from the Playback data. The results show that it is neither

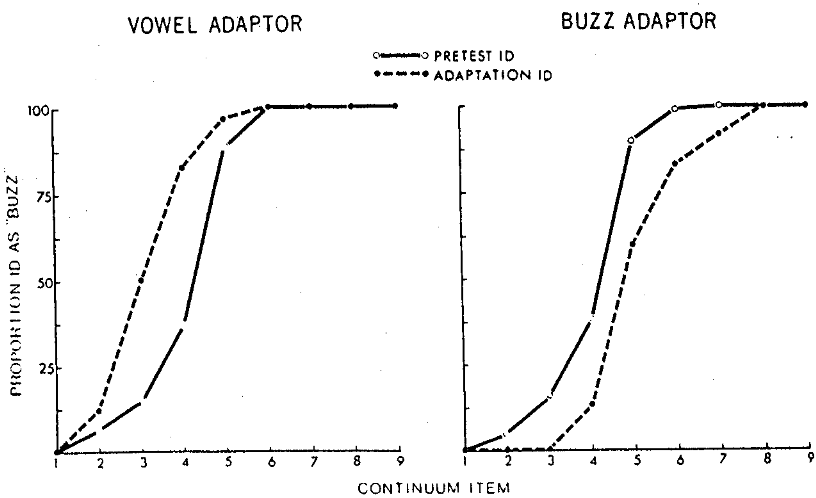

FIG. 5. Pretest and adaptation test curves for the software continuum.

the artificiality of the speech synthesizer nor the particular vowel sound involved which enables listeners to treat the present pair of continua as they do continua of proper speech sounds. This comparability suggests, at least initially, that some kind of detector may be responsible here, a detector just like those which have been assumed to underlie phonetic adaptation (e.g., Cooper, 1975). However, a simple auditory feature account is ruled out by the argument presented earlier. For several reasons, even an intermediate kind of auditory detector, one sensitive perhaps to "stimulus bandwidth," seems also to be an unlikely candidate for explaining the effects here.

First, bandwidth is defined as the frequency difference between the points of the spectrum envelope -3dB on either side of the resonant peak (Fant, 1956). Because its estimation starts from the frequency of the peak, the characterization of a resonance is two-dimensional, and the characterization of a bandwidth peak-frequency dependent; it may, therefore, be nonsensical to speak of "bandwidth" as an independent dimension of auditory coding, though the auditory pattern must certainly preserve this aspect of the stimulus in some form. Second, the assumption implied by the complex auditory detector notion, that adaptation causes a rebiasing of the mechanism which processes bandwidth, when made explicit does not provide a satisfactory account of the phenomena in question. In the present experiment, Continuum Item 1 has five resonances, the most prominent of which is F1, with its center frequency at 750Hz. Continuum Item 9, however, would be described as having an amplitude peak at 120Hz (see Fig. 2b).[6] In view of some of the highly specific effects of

[6] This synthesizer creates a source function and then modifies it with a filter function; when the formant bandwidth values (of the filter function) are increased substantially, the

adaptation demonstrated by Ganong (1976) for amplitude, and Ades (1977) for frequency, the suggestion that resonances with peaks at great frequency disparity should naturally interact must be received with caution; further study may reveal whether the limits of perceptual adaptation for resonance frequencies are similar to those for fundamental frequency. Third, because the vocal tract is an acoustic resonator (Stevens & House, 1955) it belongs to a class of natural and artifactual objects which produce resonant maxima and minima when excited. Thus, the distinction between speech sounds and nonspeech sounds requires an analysis of the stimulus which goes beyond merely detecting the presence of peaks in the spectrum. A complex auditory detector tuned to bandwidth might then serve to distinguish sharp resonators from broad, but not speech from nonspeech. The explanatory inadequacy of both simple and complex auditory detectors thus leaves two alternatives to consider, (1) a phonetic-type detector and (2) a detectorless appreciation of adaptation phenomena.

A phonetic detector explanation here would require the extension of the detector inventory, since a speech/nonspeech distinctive feature is not found in linguistic analysis. The distinction, in fact, is not even truly linguistic, in the sense of distinctive feature theory, but it certainly *is* a feature of human perceptual sensitivity, and on that basis might seem to be a candidate for detectorhood. But the existence of perceptual sensitivity should not be the only criterion for postulating a feature detector. The very advantage of this approach to problems of pattern recognition is that it makes infinite use of finite means; if a new detector is to be added to the set at every new discovery, then the contradiction of an indefinitely expandable finite means reduces the attractiveness of the original model.[7] The device required by these data can preserve its economy only if it has a

filter function is smoothed, and the resulting ouput spectrum approximates the source function, in general. The particular source spectrum used in this experiment had a fundamental frequency of 120Hz and a series of harmonics which decreased in amplitude at a rate of 10–12dB per octave; the large bandwidth values used to produce the buzz endpoint with the synthesizer resulted in an output spectrum which resembled the source spectrum, with an amplitude peak at the fundamental and a gradual rolloff.

When the formant bandwidth values used in synthesis are small, however, sharp peaks in the filter function are produced, creating the type of output spectrum shown for the vowel in Fig. 2b. In this case, the first formant (set at 750Hz) filters (i.e., subtracts from) the energy at and around the fundamental frequency, thereby producing a spectrum with a peak amplitude at the value of the first formant. This is observable in the section of the /æ/ endpoint in Fig. 2b. Notice that there is discernable but decreased energy at the fundamental, relative to the buzz spectrum.

[7] In arguing against feature detector proliferation, Weisstein (1973) proposed that a stimulus analyzing mechanism may increase its versatility by resorting to an overlord rule structure; the actual perceptual processing would take place at this more powerful rule level, making use of the detector-array outputs as food for heuristics. Thus, the essential problem of speech perception, of constancy in face of change and novelty, by this extension would

small group of detectors, tuned to speech signals, set in opposition to a small group of detectors tuned to nonspeech signal properties. On this account, the search for independent confirmation of this organizational plan finds the neurophysiology not encouraging. Although there have been discussions of single cell mediation of all perception, along the lines of innate taxa (Stent, 1975), as well as descriptions of arrays of phonetic single units (Miller, 1975), no proposal has yet been made to oppose speech neurons and nonspeech neurons. In fact, some claims for uniqueness of the speech neurology imply that the speech processor, whatever it may be, is separate from the nonspeech processor (Milner, 1962). Speech, in this view, is a functional mode, like vision or audition, and, by analogy, interacts with other modes but is independent of them. In short, a vast opponent process system for speech/nonspeech is not to be endorsed on the basis of any current view, and it may be presumed, in addition, that such a system is unlikely to exist given what is already known about cortical function.

Finally, although the reality of the *distinctive features* in linguistic analysis seems unquestionable, the only direct evidence for *feature detectors* in speech, as opposed to the invitation to such a conceptualization offered by neurophysiological metaphor, is the selective adaptation work. Boundary shifts occasioned by adaptation are precisely the effects which would permit the perceptual correlates of phonetic feature manipulations to be recast as the products of hypothetical detectors. However, though the hypothesis is reasonable when the endpoints differ by a single feature, it is difficult to imagine that a vowel and a buzz are also distinguished by but a single feature, speech/nonspeech. The adaptation technique, the only currently available test for the presence of feature detectors, is, ironically, not a demonstration of feature detectors at all. The adaptation test simply reveals that certain perceptual contrasts undergo alteration following saturation. In studying adaptation using speech sounds, research has examined the mechanics of fatigue, not the process of speech perception.

In summary, this study of vowel-buzz adaptation suggests that selective adaptation of speech does not depend on the existence of feature detectors. An auditory detector account must be rejected because a simple auditory detector makes incorrect predictions, and a complex auditory detector can neither predict the results of adaptation obtained here, nor can it serve to distinguish what is speech from what is not. A set of abstract, phonetic detectors is incapable of handling the adaptation of

---

take place after the detectors have done their work. Whatever functions the detectors might perform would appear to be fairly remote from the accomplishment of ultimate (phonetic) perception. In this case, the attempt to avoid expansion of the detector set *eliminates* the attractive simplicity of the model. See Turvey (1974), as well, for a critique of constructive models of this genre.

a speech-nonspeech boundary because that set, as defined by phonetic analysis, contains no features of nonspeech sounds. And, the type of global opponent-process organization which might account for the result is physiologically unlikely. In conclusion, the basis for adaptation, and perhaps speech perception as well, might be understood as the sensitivity to the higher order values inherent in acoustic pressure fluctuations, without decomposition into features. If so, then the description of such a process, not mere verification of analytic features, is the goal toward which further research might well proceed.

REFERENCES

Ades, A. E. How phonetic is selective adaptation? Experiments on syllable position and vowel environment. *Perception & Psychophysics*, 1974, **16**, 61–66.

Ades, A. E. Source assignment and feature extraction in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, **3**, 673–685.

Capranica, R. R., Frishkopf, L. S., & Nevo, E. Encoding of geographic dialects in the auditory system of the cricket frog. *Science*, 1973, **182**, 1272–1275.

Chomsky, N. *Aspects of a theory of syntax*. Cambridge: MIT Press, 1965.

Chomsky, N. *Language and mind*. New York: Harcourt, Brace, & World, 1968.

Chomsky, N., & Halle, M. *Sound pattern of English*. New York: Harper & Row, 1968.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, 1952, **24**, 597–608.

Cooper, F. S., Liberman, A. M., & Borst, J. M. The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proceedings of the National Academy of Sciences*, 1951, **37**, 318–325.

Cooper, F. S., & Mattingly, I. G. Computer controlled PCM system for investigation of dichotic speech perception. *Journal of the Acoustical Society of America*, 1969, **46**, 115(A).

Cooper, W. E. Selective adaptation to speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory* (vol. 1). Hillsdale, NJ: Erlbaum, 1975. Pp. 23–54.

Cooper, W., & Blumstein, S. A "labial" feature analyzer in speech perception. *Perception & Psychophysics*, 1974, **15**, 591–600.

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

Diehl, R. Feature analyzers for the phonetic dimension stop vs continuant. *Perception & Psychophysics*, 1976, **19**, 267–272.

Eimas, P. D., & Corbit, J. D. Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 1973, **4**, 99–109.

Eimas, P. D., Siqueland, E. P., Jusczyk, P., & Vigorito, J. Speech perception in infants. *Science*, 1971, **171**, 303–306.

Fant, C. G. M. On the predictability of formant levels and spectrum envelopes from formant frequencies. In M. Halle, H. Lunt, & H. MacLean (Eds.), *For Roman Jakobson*. The Hague: Mouton, 1956. Pp. 109–120.

Fisher, W. M., & Engebretson, A. M. Simple digital speech synthesis. *American Journal of Computational Linguistics*, 1975, Microfiche 16.

Fodor, J. A. How to learn to talk: Some simple ways. In F. Smith & G. A. Miller (Eds.), *The genesis of language*. Cambridge: MIT Press, 1966. Pp. 105–122.

Fromkin, V. The non-anomalous nature of anomalous utterances. *Language*, 1971, **47**, 27–52.

Ganong, W. F. Amplitude contingent selective adaptation to speech. *Journal of the Acousti-*

cal Society of America, 1976, **59**, S26(A).
Geschwind, N., & Levitsky, W. Human brain: Left-right asymmetries in temporal speech region. *Science*, 1968, **161**, 186–187.
Haggard, M. Figural after-effect phenomena for speech and non-speech stimuli. *Haskins Laboratories Status Report on Speech Research*, 1967, SR-9, 61–62.
Halle, M., & Stevens, K. Speech recognition: A model and a program for research. *IRE Transactions of the Professional Group on Information Theory*, 1962, IT-8(2), 155–159.
Halwes, T., & Jenkins, J. J. Problem of serial order in behavior is not resolved by context sensitive associative memory models. *Psychological Review*, 1971, **78**, 122–129.
Hoy, R. R., & Paul, R. C. Genetic control of song specificity in crickets. *Science*, 1973, **180**, 82–83.
Hubel, D. H., & Wiesel, T. N. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 1965, **28**, 229–289.
Jakobson, R., Fant, G., & Halle, M. *Preliminaries to speech analysis*. Cambridge: MIT Press, 1963.
Lenneberg, E. H. Language, evolution, and purposive behavior. In S. Diamond (Ed.), *Culture in history: Essays in honor of Paul Radin*. New York: Columbia University Press (published for Brandeis University), 1960. Pp. 869–893.
Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. What the frog's eye tells the frog's brain. *Proceedings of the IRE*, 1959, **47**, 1040–1059.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 421–461.
Liberman, A. M., Harris, K., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, **54**, 358–368.
Liberman, A. M., & Studdert-Kennedy, M. Phonetic perception. In R. Held, H. Liebowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology* (Vol. VIII). *Perception*. Heidelberg: Springer Verlag, 1978. Pp. 143–178.
Lieberman, P. *On the origins of language*. New York: Macmillan, 1975.
McCulloch, C. Color adaptation of edge-detectors in the human visual system. *Science*, 1965, **149**, 1115–1116.
Miller, G. A., & Nicely, P. E. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 1955, **27**, 338–352.
Miller, J. L. Properties of feature detectors for speech: Evidence from the effects of selective adaptation on dichotic listening. *Perception & Psychophysics*, 1975, **18**, 389–397.
Milner, B. Laterality effects in audition. In V. B. Mountcastle (Ed.), *Interhemispheric relations and cerebral dominance*. Baltimore: Johns Hopkins Press, 1962. Pp. 177–195.
Morse, P. A., Kass, J. E., & Turkienicz, R. Selective adaptation of vowels. *Perception & Psychophysics*, 1976, **19**, 137–143.
Mountcastle, V. B. Central neural mechanisms in hearing. In V. B. Mountcastle (Ed.), *Medical physiology* (vol. I). St. Louis: C. V. Mosby, 1974. Pp. 412–439.
Neisser, U. *Cognitive psychology*. Englewood Cliffs, NJ: Prentice–Hall, 1967.
Penfield, W., & Roberts, L. *Speech and brain mechanisms*. Princeton: Princeton University Press, 1959.
Pisoni, D. B. On the nature of categorical perception of speech sounds. *Supplement to Status Report on Speech Research*, SR-27, Haskins Laboratories, New Haven, 1971.
Pisoni, D. B., & Tash, J. Auditory property detectors and processing place features in stop consonants. *Perception & Psychophysics*, 1975, **18**, 401–408.
Pribram, K. H. *Languages of the brain*. Englewood Cliffs, NJ: Prentice–Hall, 1971.
Sawusch, J. R., Pisoni, D. B., & Cutting, J. E. Category boundaries for linguistic and

nonlinguistic dimensions of the same stimuli. *Journal of the Acoustical Society of America*, 1974, **55**, S55.

Stent, G. S. Limits to the scientific understanding of man. *Science*, 1975, **187**, 1052–1057.

Stevens, K. N., & House, A. S. Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America*, 1955, **27**, 484–493.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 1970, **77**, 234–249.

Turvey, M. T. Perspectives in vision: Conception or perception? In D. D. Duane & M. B. Rawson (Eds.), *Reading, perception and language*. Baltimore: York Press, 1975. Pp. 131–194.

Weisstein, N. What the frog's eye tells the human brain: Single cell analyzers in the human visual system. *Psychological Bulletin*, 1969, **72**, 157–176.

Weisstein, N. Beyond the yellow-Volkswagen detector and the grandmother cell: A general strategy for the exploration of operations in human pattern recognition. In R. L. Solso (Ed.), *Contemporary issues in cognitive psychology: The Loyola symposium*. Washington, DC: V. H. Winston, 1973. Pp. 17–51.

Whitfield, I. C., & Evans, E. F. Responses of auditory cortical neurones to stimuli of changing frequency. *Journal of Neurophysiology*, 1965, **28**, 655–672.

Winter, P., & Funkenstein, H. H. The effect of species-specific vocalization on the discharge of auditory cortical cells in the awake squirrel monkey (Saimiri sciureus). *Experimental Brain Research*, 1973, **18**, 489–504.

Witelson, S., & Pallie, A. Left hemisphere specialization for language in newborn. *Brain*, 1973, **96**, 641–646.

Woodworth, R. S. *Experimental psychology*. New York: Holt, 1938.

Woolsey, C. N., & Walzl, E. M. Topical projection of nerve fibers from local regions of the cochlea to the cerebral cortex of the cat. *Bulletin of the Johns Hopkins Hospital*, 1942, **71**, 315–344.

REFERENCE NOTES

1. Bailey, P. J. *Perceptual adaptation in speech*. Unpublished dissertation, Cambridge University, 1975.
2. Remez, R. E., Cutting, J. E., & Studdert-Kennedy, M. *Acoustic similarity or phonetic identity: A cross adaptation study employing song and string*. Manuscript in preparation.

(Accepted July 11, 1978)