VIII. LINGUISTICS

Academic and Research Staff

Prof. R. Jakobson
Prof. M. Halle
Prof. N. A. Chomsky
Prof. S. Bromberger
Prof. J. A. Fodor
Prof. M. F. Garrett
Prof. K. L. Hale
Prof. J. W. Harris
Prof. J. J. Katz
Prof. R. P. V. Kiparsky
Prof. W. A. O'Neil
Prof. D. M. Perlmutter
Prof. Krystyna Pomorska
Prof. J. R. Ross
R. Hurtig

Graduate Students

A. E. Ades
A. D. Andrews
M. H. Aronoff
Sarah J. Bell
Ivonne A. Bordelois
R. A. Faraci
R. W. Fiengo, Jr.
J. A. Goldsmith
H. W. Gray
D. Kahn
Ellen J. Kaufman
Mary-Louise Kean
W-C. M. Kim
Judith R. Kornfeld
M. Y. Liberman
Susan E. Martin
G. L. Milsark
R. T. Oehrle
P. R. Platero
A. S. Prince
Dorothy Siegel
A. W. Sparer
J. R. Vergnaud
E. S. Williams
E. F. Woisetschlaeger

A. SOME EFFECTS OF ADAPTATION ON SPEECH PERCEPTION

National Institute of Mental Health (Grant 5 PO1 MH13390-07)

A. E. Ades

1. Introduction

Several experiments have been reported recently that deal with the selective adaptation of "linguistic feature detectors."[1-3] The motivation for this work was to establish the existence of auditory detectors sensitive to the distinctive features of speech, particularly for voice distinction, and for place of articulation.

The method is as follows. First, a continuum of synthetic syllables ranging from, say, the voiced /dae/ to voiceless /tae/ is constructed by cutting back on the onset time of voicing relative to the burst in successive, equal steps.[4] The "phoneme boundary," the point on the continuum that would receive an equal number of "D" and "T" responses is then determined by a two-choice identification test. Next, an adapting sound, say /dae/, is repeatedly presented to the subject, and the phoneme boundary is measured again. In this case the boundary will be shifted by the adaptation toward the "D" end of the continuum. It has been argued[1,2] that this boundary shift is caused by the fatiguing of a "voice detector." Analogous results have been found for place of articulation, as exemplified by a series of syllables from /bae/ to /dae/ to /gae/.[3]

It has been claimed that the effect occurs at a phonetic rather than at an auditory level. Adaptation with either /bae/ or /dae/ can cause a shift in the voiced/voiceless boundary.[1] Similarly, not only /bae/ but also /bi/ and /pae/ have been shown to shift the bae/dae boundary toward "B",[3] presumably because all three sounds, /bae/, /bi/, and /pae/, share the feature labial. On the other hand, these particular cases do not distinguish phonetic similarities from the acoustic similarities that underlie them. We shall report, among other things, some research on critical cases where quite different acoustic structures give rise to the same phonemic percept.

Another important finding from previous work is that the effect is specific to speech stimuli. The first 50 ms of the syllable /dae/, although in some sense containing all the necessary information about the voicing of the consonant, is not a stimulus sufficient to fatigue the hypothetical "voice detector."[2] Since the first 50 ms of /dae/ is generally heard as a nonspeech chirp, one infers that adaptation takes place at a speech-specific level of representation not reached by nonphonetic sounds.

Research that is now in progress in the Department of Psychology, M. I. T., uses a series of 7 three-formant synthetic sounds from /bae/ to /dae/. Three questions are being asked. (i) Can adaptation occur only with speech stimuli as the adaptors? (ii) How abstract is the level at which adaptation takes place? Can a "labial detector" be fatigued by any instance of /b/ or /p/ or /m/, in any environment, in spite of gross acoustic differences between the test stimuli and the adapting stimuli? Can adaptation help with the problem of the variability of the acoustic cues to the same phoneme? (iii) Are there components of the adaptation effect that take place bilaterally, and others that are central, and does this distinction coincide with any distinction between auditory and phonetic levels of analysis?

2. Stimuli and Procedure

The seven test stimuli[5] differ only in their second and third formant frequency transitions. The transitions are 35 ms long and are followed by 190 ms of the steady-state vowel /ae/ (F1 = 742, F2 = 1618, F3 = 2751 Hz; amplitudes of the formants relative to F1 are -6 dB for F2, -17 dB for F3). The distances between the successive stimuli are in the ratio 2, 1, 1, 1, 1, 2, where one unit corresponds to 39 Hz in F2 starting frequency and 87 Hz in F3. Thus the extreme B stimulus had F2 and F3 transitions rise from 1384 Hz and 2422 Hz to the steady state, in a linear fashion, while in the extreme D they fell from 1696 Hz and 3114 Hz. The relative amplitudes of the formant transitions followed the acoustical theory of Fant.[6]

In all experiments the procedure has been as follows. First, the identification (ID) boundary is measured at every session (32 presentations of each of the seven test stimuli). The boundary is then computed as that point on a stimulus scale from 0 to 8 that would receive an equal number of B and D responses. During adaptation every four test stimuli are preceded by one minute of more than 90 presentations of the adapting sound. This cycle is repeated until each test stimulus has been presented 35 times (or 21 times in the backwards tweet and aeb/aed experiments.)

The adapting stimuli are always presented at the same intensity (±0. 5 dB) as the test stimuli or whatever part of the test stimuli they correspond to. The adapting stimuli are always the same as, or derived from, the extreme B or the extreme D on the test series. Stimuli are presented binaurally, unless otherwise stated.

The subjects are paid volunteers from the Cambridge area.

3. Results

a. Syllables, Chirps and Tweets: Speech versus Nonspeech

Adaptation of either the B or D end of the continuum by repetition of /bae/ or /dae/, respectively, is easily obtained, as Cooper[3] also found. Figure VIII-1 is a histogram showing the size of the boundary shift caused by adaptation. There is also a substantial effect of adapting with the chirp portion of the syllable, which is the transitions only. (The first 38 ms of the extreme test stimuli were used; that was the first four glottal pulses.) Although subjects describe chirps as sounding like dripping water, they find them quite easy to identify as b-like or d-like. Another even less speechlike stimulus, the tweet, can be made by removing the F1 portion of the chirp. This leaves only the F2 and F3 transitions, just that part of the sound that distinguishes it from the others in the series. Figure VIII-1 shows that the effect of adapting with tweets is small relative to syllables and chirps, but it is present (Wilcoxon Signed Rank test, $p < 0.005$). The sizes of all of these effects can be compared with the standard deviation of the unadapted boundary, as measured at the start of each session. The least consistent subject had a standard deviation of 0.32 stimulus units and the most consistent had 0.08. These deviations correspond to 13 Hz in the starting frequency of F2 and approximately 30 Hz for F3 in the worst subject; 3 Hz in F2, 7 Hz in F3 for the best.

It must be noted that these chirp stimuli, which give such a large adaptation effect on place of articulation, are substantially the same as those for which Eimas et al.[2] found no effect on a voice continuum. Perhaps the chirp does not convey enough information about voice to affect the hypothetical "voice detector": although the information is present, it seems probable that a voice detector would be more interested in the relative onset time of fairly sustained components, or in the presence of a frequency change before a steady state. Whether or not a nonspeech sound can be made that would adapt the voice detector is not yet known, but there are some discrimination data on nonspeech stimuli that suggest that such sounds do exist.[7,8]

Given that both tweets and chirps have an adapting effect, why are they not of the same magnitude? One possibility is that the small tweet effect is a result of fatiguing some quite low-level energy-detecting mechanisms. Accordingly, adaptation was attempted with tweets played backwards. These stimuli have exactly the same spectrum as tweets but the direction of frequency change is reversed and the phase relationships altered. The results also appear in Fig. VIII-1. There was no effect.

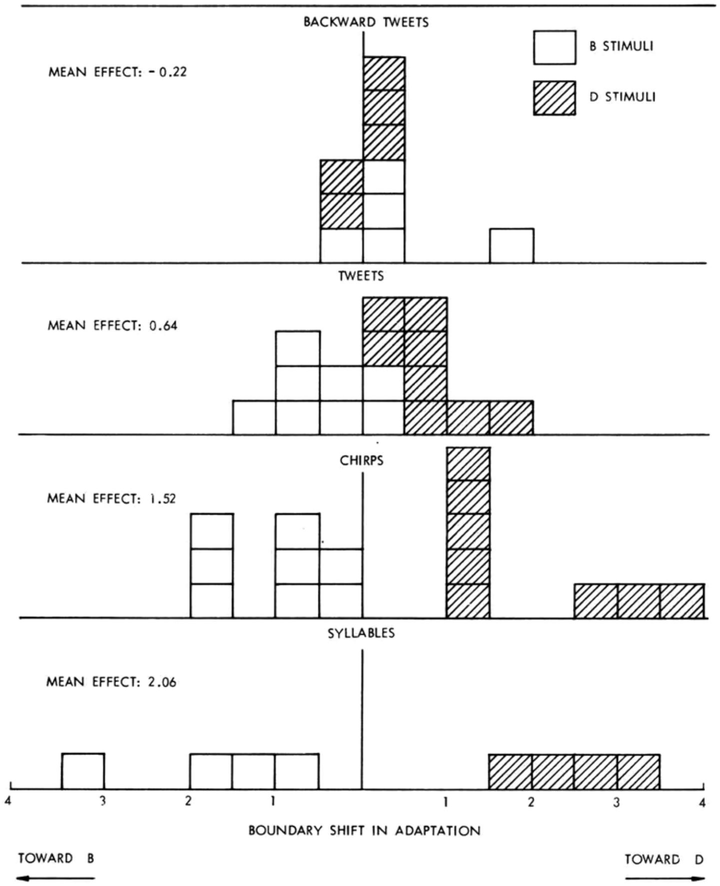

Fig. VIII-1. Comparison of the sizes of adaptation effects obtained with various stimuli. Each block represents the difference between a subject's boundary when adapted with a given stimulus, and his mean unadapted boundary.

Whatever the source of the differences in the size of syllable, chirp, and tweet effects, it appears that the same mechanisms are partly responsible for all of them. The sizes of all three effects are correlated with each other, across subjects (syllables and chirps, $r = 0.85$, $p < 0.005$; syllables and tweets, $r = 0.70$, $p < 0.05$; chirps and tweets, $r = 0.68$, $p < 0.005$).

Further research is being conducted to try to determine what makes a stimulus sufficiently speechlike, in some nonphenomenological sense, to cause adaptation, and to understand the differences in the tweet, chirp, and syllable effects.

b. Variability of Acoustic Cues to the Same Phoneme

If adaptation were taking place at a truly phonemic level, then repetition of any sound containing a /b/ should shift the ID boundary toward B. Attempts were made to adapt with /aeb/ and /aed/. These sounds were very similar to the bae/dae sounds used so far, except that the transitions were at the end, rather than at the beginning of the syllable. The transitions were slightly curved, which made them sound much more natural, but of the same overall excursion. The results of trying to adapt the bae/dae test continuum with /aeb/ and /aed/ appear in Table VIII-1. Subjects performed so differently under the same conditions that tests were run twice on /aeb/ and twice on /aed/: consistency within each subject is reassuringly good. The boundary shifts are measured relative to the unadapted boundary of that session; a positive shift on /aeb/ adaptation means a shift toward the B end of the continuum. There is no evidence here that /aeb/ or /aed/ are adapting the same mechanisms as /bae/ and /dae/ do. If anything, /aeb/ and /aed/ have the <u>same</u> effects on bae/dae perception, rather than opposite ones. Possibly the initial portion of the /aeb/ and /aed/ syllables, i.e., the initial steady state, is responsible for whatever effect there is.

Table VIII-1. Adapting the bae/dae continuum with /aeb/ or /aed/.

Subject	Boundary Shifts	
	Adapting with /aeb/	Adapting with /aed/
PC	0.18	-0.28
	0.375	-0.40
TS	-0.46	0.01
	-0.18	-0.005
NT	-1.35	1.20
	-0.51	0.91
BO	0.03	0.54
	0.03	0.48
Mean Effect	-0.24	0.31

Note: A positive boundary shift for /aeb/ indicates a shift toward the B end of the test continuum.

A follow-up experiment, now in progress, shows that /aeb/ and /aed/ can be used successfully to shift the boundary of an aeb/aed test continuum. It can be concluded that adaptation does not take place at a truly phonemic level. This is a disappointment for those of us who had looked forward to a powerful method of evaluating the perceptual significance of quite abstract linguistic feature systems, which are now motivated by purely linguistic and/or articulatory considerations.

Another case of acoustic variability underlying cues to the same phoneme arises with the syllable /de/. (This is one example from a class of similar ones.) A syllable perceived as /de/ can be synthesized with a slightly rising F2 transition (1875 Hz to a steady state at 2048 Hz) and a flat F3. By contrast, the d-transitions before the vowel /ae/, or any other vowel that is not $\begin{bmatrix}+\text{front}\\ -\text{low}\end{bmatrix}$, are falling. Preliminary results where /de/ is presented as an adapting stimulus suggest, however, that /de/ can fatigue the same mechanisms responsible for /dae/, in spite of the fact that acoustically /de/ is more like /bae/ than /dae/. Evidently adaptation can occur at quite an abstract level, yet not a level sufficiently abstract to generalize between /bae/ and /aeb/.

There is an interesting possibility that the tweet or chirp components of the same /de/, if such sounds do indeed have their adapting effect at a prephonetic auditory level, would fatigue the B end of the bae/dae continuum.

c. Dichotic Adaptation and Transfer of Adaptation

In these experiments the procedure differs, in that the unadapted boundaries for the left and right ears are measured separately. During adaptation the left and right ears are also tested separately. Various factors were counterbalanced across experimental conditions: orientation of the subjects' headsets, order of testing, and certain factors to do with the equipment.

In the dichotic experiment /bae/ is presented to one ear, and /dae/ to the other simultaneously. A single ID boundary in the adapted state is taken for each subject from two sessions (14 presentations of each stimulus per ear per session). The boundary shifts given in Table VIII-2, along with the standard deviation of the unadapted boundary, are relative to the mean unadapted boundary for a given subject's ear (N = 8). The experiment was performed with /bae/ left, /dae/ right, and the converse (two sessions each). Evidently it is possible to shift the boundary one way in one ear, and the other way in the

Table VIII-2. Dichotic syllable adaptation.

Subject	Unadapted Boundary Standard Deviation (N = 8)		Boundary Shifts /bae/ L, /dae/ R		Boundary Shifts /dae/ L, /bae/ R	
	L	R	L	R	L	R
FB	0.68	0.93	0.86	1.40	-0.78	-0.69
JJ	0.48	0.18	0.87	0.69	0.61	0.56
TT	0.37	0.42	0.86	1.03	1.12	0.23
BK	0.26	0.48	0.46	0.34	1.11	1.54
Mean Effect			0.76	0.865	0.515	0.41

Note: Negative entries indicate boundary shifts away from the predicted direction.

other ear. The two negative boundary shifts were produced by a subject who spoke Italian exclusively until she was seven, and whose unadapted boundaries were rather unstable. Nonetheless the effect is highly significant (Wilcoxon Signed Rank test, $p < 0.005$). We must conclude that some component of the adaptation effect is bilateral, one "detector" on the left, and one on the right.[9]

The transfer experiment simply involves adapting one ear with a syllable, and testing both ears separately. The results appear in Table VIII-3. The shifts are computed from the same baseline for each subject that was used for the dichotic boundary shifts. The adapted boundary was measured from a single session (14 presentations of each stimulus per ear). The percent transfer is the ratio of the shift in the unadapted ear to the shift in the adapted ear. For these stimuli the overall percent transfer was ~55%. Eimas, Cooper, and Corbit[2] found that transfer for their voiced/voiceless continuum was ~100%.

Table VIII-3. Transfer of syllable adaptation.

Subject	Boundary Shifts							
	Transfer Left → Right				Transfer Right → Left			
	/bae/ Left		/dae/ Left		/bae/ Right		/dae/ Right	
	L	R	L	R	L	R	L	R
FB	3.16	1.45	3.08	1.09	3.49	>4.31	-0.99	0.64
Percent Transfer	46		39		>81		? ?	
JJ	1.36	0.09	3.07	0.97	0.65	2.31	1.19	1.21
Percent Transfer	9		32		28		98	
TT	0.91	0.64	2.45	0.93	0.70	0.64	1.09	1.36
Percent Transfer	70		38		109		80	
BK	1.64	1.64	3.11	2.27	1.49	2.51	2.61	>3.16
Percent Transfer	100		73		59		>83	
Mean Effect	1.77	0.955	2.93	1.315	1.42	2.19	0.975	1.59
Percent Transfer	54		45		65		61	

Note: The symbol ">" indicates that the adaptation was so large that the boundary was shifted beyond the ends of the stimulus continuum.

A transfer effect is generally taken as an indication of adaptation at a central site. Alternative arguments do exist, and will be explored, but provisionally we shall interpret the dichotic and transfer results as pointing to both central and bilateral components of adaptation. It should be stressed that the bilateral component should not be considered as "peripheral." The backwards tweet data rule out any simple energy-detecting mechanisms.

The question arises whether the bilateral effect is at a nonphonetic level, while the central effect reflects a much more abstract phonetic analysis. One could hypothesize that the bilateral effect is due to the tweet part of the syllable. (The dichotic and tweet effects are the same size, interestingly enough.) If this is correct, then there should be no transfer of adaptation for tweets, and there should be a dichotic effect of the same size as the dichotic syllable effect. On the other hand, there should be no dichotic effect with /be/ and /de/, since these sounds are quite different from the bae/dae test sounds at the auditory level. But there should be transfer with /de/ and /be/ because at the phonetic level they would be analyzed the same as /bae/ and /dae/. This experiment awaits more tests of the bae/dae boundary after adaptation with /be/, /de/ and their associated chirps and tweets.

4. Summary

Identification of stimuli on a bae/dae test continuum after adaptation with various stimuli has revealed several things.

1. Appropriate nonspeech sounds can fatigue the "detectors" responsible for identification of the bae/dae test sounds, but are not as effective as the syllables themselves.
2. Mechanisms that simply detect the energy at certain frequencies cannot explain any of the adaptation results.
3. Some adaptation takes place at a phonetic level sufficiently abstract to generalize between /de/ and /dae/ in spite of there being no obvious acoustic similarity between these sounds, but it is not so abstract that it treats initial and final consonants alike.
4. There are certainly bilateral, and almost certainly central, components of adaptation.

Further research will concentrate on defining an auditory/phonetic distinction which reflects the bilateral/central distinction.

Footnotes and References

1. P. D. Eimas and J. D. Corbit, "Selective Adaptation of Linguistic Feature Detectors," Cognitive Psychol. 4, 99-109 (1973).
2. P. D. Eimas, W. E. Cooper, and J. D. Corbit, "Some Properties of Linguistic Feature Detectors," Perception Psychophys. 13, 247-252 (1973).
3. W. E. Cooper, "Adaptation of Linguistic Feature Detectors for Place of Articulation," Master's Thesis, Brown University, 1973.

4. L. Lisker and A. S. Abramson, "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements," Word **20**, 384-422 (1964).
5. All stimuli were generated by a program written by Dr. D. H. Klatt on the PDP-9 computer facility of the Speech Communication Group of the Research Laboratory of Electronics. The program, which was essentially a simulation of a parallel resonance synthesizer, also controlled the output of the stimuli onto recording tape, in a format suitable for the identification tests. Adapting stimuli were output in long sequences which were later made into tape loops.
6. C. G. M. Fant, "On the Predictability of Formant Levels and Spectrum Envelopes from Formant Frequencies," in Ilse Lehiste (Ed.), *Readings in Acoustic Phonetics* (The M. I. T. Press, Cambridge, Mass., 1967).
7. I. Nabelek and I. J. Hirsh, "On the Discrimination of Frequency Transitions," J. Acoust. Soc. Am. **45**, 1510-1519 (1969).
8. A. E. Ades, "Categorical Perception and the Speech Mode" (unpublished paper).
9. Under dichotic syllable stimulation subjects were completely unaware that the sound in the left ear is different from that in the right: a single, fused percept is heard, even by non-naive listeners. This situation, not characteristic of most dichotic syllable stimulation, occurs because the left and right inputs are identical for the vowel. The precision of the temporal alignment achieved with the program AUDITS, written by Dr. A. W. F. Huggins, is also a factor.