The effect of selective adaptation on the identification of speech sounds

RANDY L. DIEHL

University of Minnesota, Minneapolis, Minnesota 55455

An experiment was performed to determine the effect of selective adaptation on the identification of synthetic speech sounds which varied along the phonetic dimension *place of articulation*. Adaptation with a stimulus of a particular place value led to a reduction in the number of test stimuli identified as having that place value. An identification shift was obtained even when the acoustic information specifying place value for the adapting stimulus had virtually nothing in common with the information specifying place value for any of the test stimuli. Removing the vowel portion of an adapting stimulus eliminated identification shift only when the resulting stimulus was no longer perceived as speech-like. The results indicate that at least part of the adaptation effect occurs at a site of phonetic, not merely acoustic, feature analysis.

Eimas and Corbit (1973) used a selective adaptation procedure to investigate the possible involvement of linguistic feature detectors in speech perception. They reasoned that if such detectors mediate the perception of voiced and voiceless stops, then repeated presentation of stimuli having, e.g., the feature *voiced* should cause the detector of that feature to become fatigued and thus rendered less sensitive. Subjects first identified a series of synthetic consonant-vowel (CV) syllables which varied only along the acoustic dimension of voice onset time (VOT). Variations in VOT, the interval between release burst and onset of laryngeal pulsing, signal the distinction between voiced stops (b, d, g) and voiceless stops (p, t, k). Following the initial identification test, subjects listened to the repeated presentation of either a voiced or a voiceless stop and then identified the series again. It was found that if the repeated stimulus was voiced, a greater number of identification responses were now of the unadapted, voiceless category, whereas if a voiceless adapting stimulus was used, there was an increase in identification responses of the voiced category. For example, if the identification series was (ba-pa), adaptation with /ba/ or /da/ led to an increase in the number of stimuli identified as /pa/. These results tend to support the hypothesis that there exist linguistic feature detectors which may be reduced in sensitivity through selective adaptation.

Eimas, Cooper, and Corbit (1973) obtained little or no shift in stimulus identification along the VOT

dimension when the steady-state vowel portion of the adapting stimulus was removed (leaving a stimulus which retained the VOT information but which was usually perceived as nonspeech). It appeared that a necessary condition for an effective shift in identification was that the adapting stimulus be perceived as speech. According to Eimas et al., this result suggests that the VOT detectors are part of the speech processing system, i.e., that they have a specialized linguistic function beyond mere auditory processing.

Cooper (1974) sought to determine whether the identification shifts result from adaptation of detectors which are sensitive to relatively invariant acoustic parameter values or whether they result instead from adaptation of detectors sensitive to phonetic information which may not be acoustically invariant. He had subjects identify a series of synthetic CV syllables which varied in starting frequency and direction of second-formant (F-2) and third-formant (F-3) transitions. Such variations signal phonetic distinctions along the *place of articulation* dimension; in this case, the distinctions were among the syllables /bae/, /dae/, and /gae/, which have the place values *bilabial*, *alveolar*, and *velar*, respectively. After the identification test, an adapting stimulus was presented repeatedly, and the subjects were retested on the identification series. In general, test stimuli lying near a phonetic boundary, which were identified before adaptation as having the same place value as the adapting stimulus, were identified after adaptation as having a different place value. For example, adaptation with either /bae/, /pae/, or /bi/, all of which share the place feature *bilabial*, led to a decrease in the number of test stimuli identified as /bae/. The effectiveness of /bi/ as an adapting stimulus was of particular interest. Because

This research was supported in part by a National Science Foundation graduate fellowship to the author and by grants to the University of Minnesota, Center for Research in Human Learning, from the National Science Foundation (GB-35703X), the National Institute of Child Health and Human Development (HD-01136), and from the Graduate School of the University of Minnesota. The author is indebted to James Jenkins for his many helpful comments and criticisms.

the formant transitions of this stimulus differed considerably from those of the test stimuli, Cooper suggested that adaptation for the place feature occurred mainly at a site of phonetic, rather than acoustic, feature detection.

In a further study of the place dimension, Ades (1974) obtained evidence both for and against this Phonetic hypothesis. On the one hand, he obtained a shift in the identification of /dae/ after repeated presentation of /de/, even though the adapting /d/ was acoustically different from the test /d/. This result seems to favor the phonetic hypothesis. On the other hand, Ades found that repeated presentation of a CV syllable produced no identification shift on a series of VC syllables; nor did a VC adapting stimulus affect the identification of a CV series. Phonetic equivalence between the consonant of the adapting stimulus and the consonant of some of the test stimuli was not sufficient to produce an adaptation effect, a result which poses a problem for any strong version of the phonetic hypothesis.

One might argue that those results of Cooper (1974) and Ades (1974) which lend *prima facie* support to the phonetic hypothesis appear on closer analysis to be somewhat equivocal. Though identification shifts were obtained using adapting stimuli which differed acoustically from any of the test stimuli, the magnitude of that acoustic difference may have been overstated. Varying the vowel context of an initial consonant does indeed alter the absolute direction of the formant transitions, but this does not imply that nothing remains invariant across such changes in context. DeLattre, Liberman, and Cooper (1955) suggested that each consonant has a characteristic frequency position, or locus, to which the F-2 transition "points." This abstract locus appears to remain relatively fixed for each of the consonants /b/ and /d/, despite changes in vowel context. Thus, the results of Cooper and Ades may well be consistent with the hypothesis that adaptation mainly affects detectors of invariant, though highly abstract, acoustic features.

A critical test of the phonetic hypothesis requires an adapting stimulus which shares a phonetic feature with some of the test stimuli but which has virtually nothing in common with them acoustically (not even an abstract frequency locus). If such an adapting stimulus produces an identification shift, the phonetic hypothesis is supported. If no identification shift is produced, the phonetic hypothesis may probably be rejected. The present experiment was designed to provide this kind of critical test. Specifically, the plan was: (1) to demonstrate shifts in identification along the place dimension using either voiced or voiceless adapting stimuli of a particular place value, and (2) to see whether shifts occur even when the acoustic information which specifies place for the adapting stimulus is totally different from that which specifies place for the test stimuli.

A second question relevant to the phonetic hypothesis was whether removal of the vowel context of an adapting stimulus would eliminate identification shift along the place dimension. Inasmuch as the adapting stimulus loses its speech-like quality when the vowel portion is removed, it was predicted that such a stimulus would produce little or no identification shift.

METHOD

Stimuli

The test stimuli were a series of 11 synthetic CV syllables prepared by means of a computer-controlled parallel resonance synthesizer (Glance-Holmes) at the University of Minnesota. The stimuli, which were perceived as either the bilabial sound /b/ or the alveolar sound /d/, plus the vowel /ε/, varied in starting frequency and direction of the F-2 and F-3 transitions. These starting frequencies, together with the terminal steady-state frequencies, are displayed in Table 1. The formant transitions lasted for a period of 40 msec, which was followed by a 250-msec period of steady-state formants corresponding to the vowel. The voiced quality of the stimuli was obtained by making the temporal onset of the F-1 transition equal to the onset of the F-2 and F-3 transitions.

Ten different adapting stimuli were synthesized. Two of these were copies of the first and last stimuli in the test series, which were perceived as /be/ and /de/. Two other stimuli were acoustically identical to the above two, except that the onset of F-1 was delayed 60 msec relative to the onset of F-2 and F-3, and F-2 and F-3 were excited by a noise source rather than a periodic source when F-1 was absent. In this manner, the /be/ and /de/ were transformed into their voiceless cognates /pe/ and /te/, respectively, with place value remaining unchanged. These will be referred to as *transition-cued* /pe/ and /te/ to distinguish them from *burst-cued* /pe/ and /te/ which were synthesized as follows: A 15-msec burst of noise at a particular frequency range was followed by 15 msec of silence, which was in turn followed by a 245-msec period of steady-state formants (whose frequency values were identical to the steady-state formants of the test stimuli). The noise burst of the /pe/ stimulus had a bandwidth of 80 Hz with a center frequency at 310 Hz; the burst of the /te/ stimulus had a 110-Hz bandwidth with a center frequency at 2,665 Hz. It should be noted that, with the exception of the steady-state formant values, the burst-cued /pe/ and /te/ had virtually nothing in common acoustically with any of the test stimuli. Nevertheless, /pe/ shared the phonetic place feature *bilabial* with those test stimuli which were perceived as /be/, while /te/ shared the feature *alveolar* with those test stimuli perceived as /de/.

The four remaining adapting stimuli were produced by eliminating the steady-state (vowel) portions from the transition-cued /pe/ and /te/ and from the burst-cued /pe/ and /te/. That is, they consisted of either a 40-msec period of F-2 and F-3 transitions or a 15-msec noise burst.

Procedure

Eleven experiment sessions, each lasting about 25 min, were run on consecutive weekdays. In the first session, subjects identified the test stimuli and then, after a 15-min interval with no stimulus presentation, identified the test stimuli a second time. In each of the remaining sessions, subjects first identified the test stimuli and then listened to one of the adapting stimuli presented repeatedly, after which they again identified the test stimuli. A Crown CX822 tape recorder was used to present the test stimuli, and the adapting stimuli were presented using a Revox A77 tape recorder. Both sets of stimuli were channeled through a mixer and heard through sets of Koss Pro 600AA earphones at a comfortable level.

The identification functions for the subjects in the unadapted state were obtained as follows: The test stimuli were presented in 10 blocks, each consisting of a different randomization of the 11

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

Stimulus	Starting Frequency (in Hz)	
	F-2	F-3
1	1257	2113
2	1317	2174
3	1377	2235
4	1437	2297
5	1497	2358
6	1557	2419
7	1617	2481
8	1677	2542
9	1737	2603
10	1797	2665
11	1857	2726

Note—The terminal steady state frequencies were centered at 529 (F-1), 1824 (F-2), and 2481 (F-3).

stimuli. The interval separating the stimuli was 2 sec within a block, and 4 sec between blocks. Subjects were instructed to identify each stimulus by writing "B" or "D" on an answer sheet provided.

Next, identification functions were obtained after selective adaptation. Subjects listened to an adapting stimulus for periods of 1 min and, after each such period, identified one block of the 11 test stimuli. This continued until all of the original 10 blocks were identified. A 365-msec interval separated each presentation of an adapting stimulus, and there was a 4-sec interval between the end of an adaptation period and the onset of an initial test stimulus. At the end of a session, each subject was asked to describe his perception of the adapting stimulus. To prevent subjects from memorizing the order of the test stimuli, four different orders of stimulus presentation were alternately used in the 11 experimental sessions. In a given session, subjects always heard the same order of presentation both before and after selective adaptation. The adapting stimuli were assigned to the last 10 experimental sessions in the following order: transition-cued /pε/, burst-cued /tε/, transitions alone of /tε/, /bε/, burst alone of /pε/, transition-cued /tε/, transitions alone of /pε/, burst alone of /tε/, burst-cued /pε/, /dε/.

Subjects

Six students at the University of Minnesota served as subjects. They received \$2 per session and were not informed of the nature of the experiment beforehand.

RESULTS

Table 2 displays the net change in the number of stimuli identified as /b/ between the first and second identification tests, for each of the 10 adapting stimuli and for the control condition in which no adapting stimulus was presented. For this control condition, the subjects showed a very stable performance. From the first test to the second, there was no significant shift in the number of stimuli identified as /b/. In contrast, adapting with /bε/ produced a highly reliable decrease in the number of /b/ identifications ($p < .005$),¹ while adapting with /dε/ led to a reliable increase in /b/ identifications ($p < .025$).

The transition-cued /pε/ was also an effective adapting stimulus, producing a significant decrease in the number of /b/ identifications ($p < .005$). The results obtained using the transition-cued /tε/ are less straightforward. With this adapting stimulus, the six subjects as a group did not show a significant shift in a given direction. When asked to describe their perception of this stimulus, one subject reported hearing /tε/, another reported hearing /hε/, and four reported hearing /pε/. The subject who perceived /tε/ showed a substantial increase in /b/ identifications, while those subjects who perceived /pε/ showed a reliable decrease in /b/ identifications ($p < .05$).

A similar division among the subjects occurred when the adapting stimulus was the burst-cued /pε/. For the group as a whole, this stimulus did not produce a significant identification shift in a given direction. However, only four of the subjects reported hearing the stimulus as /pε/ (the other two subjects reported hearing varieties of nonspeech sounds), and for these four there was a significant decrease in /b/ identifications ($p < .05$). Adapting with the burst-cued /tε/, on the other hand, led to a reliable increase in the number of test stimuli identified as /b/ ($p < .025$). (It should be pointed out that all of the adapting stimuli discussed so far in this section, with the exception of the transition-cued /tε/ and the

Table 2
Net Change in Number of Stimuli Identified as /b/ Between First and Second Identification Tests

Adapting Stimulus	Subjects						Mean
	1	2	3	4	5	6	
No Adaptation	2	-2	3	-1	1	0	.50
/bε/	-9	-17	-16	-8	-25	-10	-14.17
/dε/	8	10	8	-3	17	4	7.33
/pε/ (Transition Cued)	-3	-13	-12	-8	-10	-13	-9.83
/tε/ (Transition Cued)	-3*	-2**	-3**	-10**	-1**	7	-2.00
/pε/ (Burst Cued)	4†	5†	-3	-3	-6	-6	-1.50
/tε/ (Burst Cued)	2	-2	6	3	8	8	4.17
/pε/ (Transitions Alone)	5	1	-1	-8	-3	-10	-2.67
/tε/ (Transitions Alone)	-3	-2	1	3	-2	-3	-1.00
/pε/ (Burst Alone)	0	3	-3	5	-1	-2	.33
/tε/ (Burst Alone)	3	1	2	2	8	5	3.50

Note—A positive number indicates an increase, from the first test to the second, in the number of stimuli identified as /b/; a negative number indicates a decrease.

*Heard /hε/

**Heard /pε/, not /tε/

†Did not hear /pε/

Figure 1. The percentages of /b/ identification responses of Subject 3 for four different adapting stimuli.

burst-cued /pε/, were perceived in the expected manner by all the subjects.)

No significant identification shifts were obtained when the transition portion alone of /pε/ or /tε/ (transition-cued) was used as the adapting stimulus. Nor did a significant shift occur when only the noise burst of the burst-cued /pε/ was used. As expected, none of these three stimuli was perceived as speech. But, surprisingly, the noise burst alone of the burst-cued /tε/ was heard as /t/-like by all the subjects. Moreover, this adapting stimulus produced a small, but reliable, increment in the number of /b/ identifications ($p < .025$).

The shift in identification, when it occurred, was most pronounced for stimuli lying near the middle of the test series, i.e., near the phonetic boundary. Stimuli lying near either end of the series were almost always identified consistently. This fact is illustrated in Figure 1, which shows the identification functions of a typical subject for four different adapting stimuli.

DISCUSSION

By replicating some of the earlier results of Cooper (1974), the present experiment confirmed the notion that feature detectors mediate the perception of phonetic distinctions along the place of articulation dimension. Selective adaptation with a stimulus of a particular place value led to a reduction in the number of test stimuli identified as having that place

value. This may be explained by the hypothesis of Eimas and Corbit (1973) that repeated presentation of a feature to which a detector is sensitive fatigues the detector and thus reduces its sensitivity. A shift in stimulus identification along the place dimension was obtained even when the adapting stimulus was voiceless and the test stimuli were voiced, suggesting that adaptation affects feature-specific, rather than phoneme-specific, detector mechanisms.

In addition, the present experiment provided considerable evidence that at least part of the adaptation effect occurs at a site of phonetic, rather than merely acoustic, feature detection. This evidence may be summarized as follows: (1) Reliable identification shifts were obtained when the acoustic information specifying place value for the adapting stimulus had virtually nothing in common with the information specifying place value for the test stimuli. Adaptation with a burst-cued /pε/, provided it was perceived as a bilabial speech sound, reduced the number of test stimuli identified as bilabial, even though the place value of the test stimuli was cued by formant transitions. Similarly, adapting with a burst-cued /tε/ reduced the number of test stimuli identified as alveolar. (2) Those subjects who perceived the transition-cued /tε/ adapting stimulus as the bilabial sound /pε/ showed a reliable decrease in the number of test stimuli identified as bilabial. This result is particularly interesting in view of the fact that the transition-cued /tε/ has exactly the same acoustic information for place value (i.e., the same starting frequency and direction of F-2 and F-3 transitions) as the last stimulus of the test series, which was uniformly identified as alveolar. Thus the perceived phonetic character of the adapting stimulus appears to have been more important than its actual acoustic character in determining the direction of the identification shift. (3) Of the adapting stimuli which lacked steady-state (vowel) portions, only one, the noise burst alone of the burst-cued /tε/, produced an identification shift. This stimulus was also the only one which was perceived as speech-like. It appears, therefore, that phonetic processing may be necessary in order to produce a reliable adaptation effect.

In short, the results of the present experiment strongly indicate that at least part of the adaptation effect occurs at a level of phonetic analysis.

REFERENCES

- ADES, A. E. How phonetic is selective adaptation? Experiments on syllable position and vowel environment. *Perception & Psychophysics*, 1974, **16**, 61-66.
- COOPER, W. E. Adaptation of phonetic feature analyzers for place of articulation. *Journal of the Acoustical Society of America*, 1974, **56**, 617-627.
- DELATTRE, P. C., LIBERMAN, A. M., & COOPER, F. S. Acoustic loci and transitional cues of consonants. *Journal of the Acoustical Society of America*, 1955, **27**, 769-773.
- EIMAS, P. D., COOPER, W. E., & CORBIT, J. D. Some properties of linguistic feature detectors. *Perception & Psycho-*

physics, 1973, **13**, 247-252.

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

shifts was determined by one-tailed t tests for correlated measures.

NOTE

1. In all instances, the statistical significance of identification

(Received for publication June 12, 1974;
revision received August 9, 1974.)