Perception & Psychophysics
1974, Vol. 16 (2), 229-234

Perceptuomotor adaptation to a speech feature*

WILLIAM E. COOPER
Massachusetts Institute of Technology, E10-044, Cambridge, Massachusetts 02139

Selective adaptation experiments were conducted to test for the presence of a mechanism that mediates an aspect of both speech perception and speech production. Ss were instructed to utter /pi/ or /bi/ after listening to repetitions of either of these syllables or to repetitions of the vowel /i/. Analysis of the utterances showed that a timing relation which distinguishes /pi/ from /bi/, namely the latency in onset of voicing relative to the release burst of the consonant, varied systematically for the /pi/ utterances but not for the /bi/ utterances as a function of the speech input. The effect for the /pi/ utterances was shown not to be attributable to factors such as compensation for distorted perception of the /pi/ adapting stimulus or voluntary mimicry of this stimulus.

A number of writers, including both philosophers and scientists, have proposed that some aspects of speech perception and speech production are subserved by a common mechanism (cf. de Cordemoy, 1668; von Humboldt, 1836; Lashley, 1951; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Stevens, 1972). Convincing evidence in support of this proposal has, however, been lacking (for a recent discussion, see Bailey & Haggard, 1973).

The psychophysical technique of selective adaptation has recently been employed to study the processing operations that occur during speech perception (cf. Eimas & Corbit, 1973; Eimas, Cooper, & Corbit, 1973; Bailey, 1973; Ades, 1974a, b; Cooper, 1974a, b; Cooper, 1975; Cooper[1]; Cooper & Blumstein, 1974). Using a variation of this technique, it has been possible in this study to provide a fairly direct test for the existence of a mechanism that mediates one aspect of both speech perception and articulation.

In the present experiments, Ss were required to *listen* to repetitions of an adapting stimulus immediately prior to *uttering* a selected syllable. The utterances were then analyzed to determine whether a feature of the waveforms systematically varied as a function of perceptual adaptation. Attention was focused on a single phonetic feature, namely the feature of *voicing* in initial consonants. This feature was chosen for three reasons: (a) it has an acoustic correlate that can be measured from real speech waveforms with a relatively high degree of speed and accuracy, (b) numerous data on perceptual adaptation have already been obtained for this feature (cf. Eimas & Corbit, 1973; Eimas et al, 1973; Cooper, 1974a; Cooper[1]), and (c) the feature is virtually universal in that it serves to mark phonemic distinctions in most, if not all, natural languages (cf. Lisker & Abramson, 1964).

For English, *voicing* serves to minimally distinguish the voiced from the voiceless stop consonants (i.e., /b/ from /p/; /d/ from /t/; /g/ from /k/). During speech production, *voicing* distinctions for word-initial stops can be signaled by the cue of voice onset time (VOT), a temporal relation between the onset of vocal cord vibration and the release of oral closure. Voiceless stops are normally produced with an onset of vocal cord vibration that lags the release of closure by more than 30 msec, whereas voiced stops are produced with an earlier onset of voicing (cf. Lisker & Abramson, 1964). Acoustically, VOT is specified as the onset of first formant energy in the spectrum relative to the onset of higher formant energy, with the higher formants being excited by a noise source rather than a periodic source during the period when the first formant is absent (cf. Lisker & Abramson, 1970). In the present study, the question of perceptuomotor control was tested by analyzing the VOT values for a large sample of stop consonant + vowel utterances, produced under differing conditions of perceptual adaptation.

METHOD

Subjects

Sixteen M.I.T. students, nine males and seven females, participated in this experiment. All were native speakers of American English. None had prior phonetic training or any known hearing or speech defects.

Stimuli

The adapting stimuli consisted of three synthetically generated speech syllables, /i/, /pi/, and /bi/. The syllables were five-formant patterns generated on a terminal analog speech synthesizer at the M.I.T. Research Laboratory of Electronics (cf. Klatt[2]). The acoustic form of the steady state vowel was identical for the three syllables and simulated the output of a male vocal tract. The two CV syllables, /pi/ and /bi/, differed from each other only in VOT and the onset of the formant transitions after voicing onset (cf. Stevens & Klatt, 1974); for /pi/, the VOT value was +80 msec (voicing onset lagged the release burst of the consonant by 80 msec), whereas for /bi/, the VOT value was 0 msec (voicing onset was simultaneous with the release burst of the consonant). The peak amplitudes of the three syllables were matched, and the overall duration of each syllable was 255 msec.

*Supported by NIH Training Grant NIH-5-T01-GM-01064-12 and by the Research Laboratory of Electronics, M.I.T. I thank R. M. Held, K. N. Stevens, A. W. F. Huggins, D. H. Klatt, A. M. Liberman, J. A. Fodor, M. F. Garrett, W. F. Ganong, A. E. Ades, T. B. Felton, and H.-L. Teuber for their suggestions and encouragement during this work.

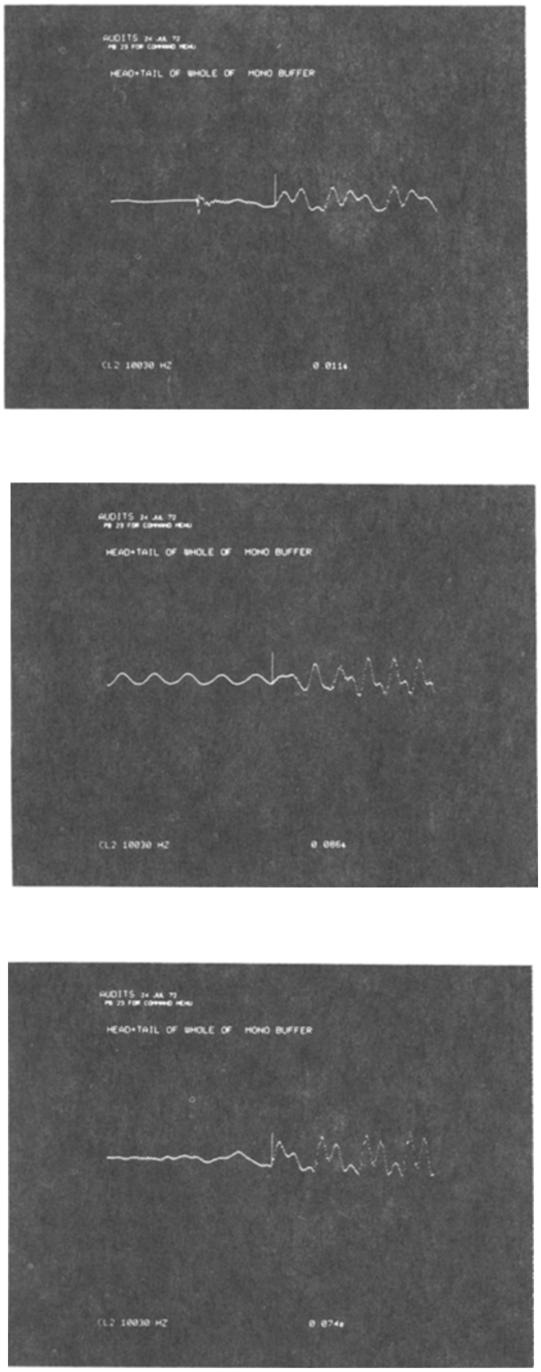

Fig. 1. Sample oscillographic displays of the test utterances, illustrating the method used to analyze VOTs. The top display shows the initial segment of a /bi/ utterance containing a short-lag VOT (voicing onset occurs shortly after the release burst of the consonant). The pointer "T" indicates the onset of the release burst, the vertical line to the right of "T" marks the onset of voicing, and the time difference between the "T" pointer and the vertical line is displayed at the bottom right of the oscilloscope screen. For this utterance, the VOT is measured to be 11.6 msec. The middle display shows a segment of a prevoiced /bi/ syllable (voicing onset precedes the release burst of the consonant). For prevoiced /bi/s, the pointer "T" (not shown in the display segment here) is set at the onset of voicing, while the vertical line cursor is aligned with the onset of the release burst. The time difference between the two markers is taken as the measure of VOT, in this case −86.6 msec (negative VOT values signify prevoicing). The bottom display shows a segment of aspiration and the onset of voicing for a /pi/ utterance. The "T" pointer is set at the onset of the release burst (not shown in this display segment), and the vertical line cursor is moved to the position of voicing onset as in the case of short-lag VOT /bi/ utterances. The actual screen size is 14 x 13 in., and the vertical and horizontal display scales can be magnified to facilitate the marking procedure.

Procedure

The Ss were tested individually in a soundproofed room. Each S was presented 40 adaptation trials during a single, hour-long session. The trials were arranged in blocks of 10, with an approximately 5-min rest period between blocks. Within each block, Ss were presented a single adapting stimulus type (/i/, /pi/, or /bi/). On each individual trial, Ss first listened to 70 repetitions of the adapting stimulus, with an interrepetition interval of 350 msec. The adapting stimulus was played via an Ampex PR-10 tape recorder while the Ss listened binaurally over KLH-Z61 headphones. The Ss were instructed to hold their tongues firmly between their teeth and lips while listening and were told to minimize subvocalization. After 70 repetitions of the adapting stimulus were presented, Ss released their tongues immediately and uttered a single CV syllable (either /pi/ or /bi/ for a given block) in a natural voice as soon as possible. Ss were capable of noticing the cessation of the adapting r‹ ɔetitions on each trial in a virtually automatic manner (mean ›.esponse latency < 1.5 sec), and for this reason no external signal was employed at the end of the reptitions to prompt a response. Five seconds lapsed between trials, and Ss were told to regard each new trial as a separate task. All utterances were recorded onto tape via a Neumann U87 microphone and a Revox A77 tape recorder.

The four blocks of 10 trials consisted of two major groups; one group of 20 trials required the verbal response /pi/, the other required the verbal response /bi/. Eight Ss were presented the group of 20 /pi/-response trials first; the other eight Ss were presented the two groups of trials in reverse order. Within each group of 20 trials, the first block of 10 trials always involved listening to repetitions of the isolated vowel /i/ immediately prior to each utterance (the control condition[3]), while the second block involved listening to repetitions of the syllable belonging to the same CV type as the required utterance. In the second block of trials, Ss thus listened to repetitions of /pi/ immediately prior to uttering /pi/ responses and listened to repetitions of /bi/ immediately prior to uttering /bi/ responses.

RESULTS AND DISCUSSION

Each of the 640 utterances was analyzed for VOT oscillographically, aided by the AUDITS computer program written by Huggins (1969). A cursor was manipulated to mark the onset of the consonant release burst and the onset of voicing for each utterance. The time difference between the two markers was displayed on the oscilloscope screen to the nearest 100 microsec. The accuracy of each VOT measurement was estimated to be within ±1 msec, except in the case of those /bi/ utterances having VOTs near 0 msec, where the accuracy was reduced to about ±3 msec. Examples of the oscilloscope displays are shown in Fig. 1.

The results for each S are shown in Table 1, where the mean VOT value is displayed for each test condition. Figures 2 and 3 show the frequency distribution of the VOT values for the entire group of Ss. For the /pi/ utterances, a significant decrease in VOT values was obtained after perceptual adaptation to /pi/, as compared with the VOT values obtained in the control condition (adaptation with the vowel /i/) $(p < .05)$.[4] Thirteen of the 16 Ss showed this decline in VOT as a function of perceptual adaptation. For the /bi/ utterances, no systematic shift in VOT was obtained after perceptual adaptation to /bi/. Some Ss did show marked shifts after /bi/ adaptation (cf. Table 1), but these shifts were not systematic in direction across Ss.

The results for the /pi/ utterances indicate that perceptual adaptation can indeed exert a systematic influence on the speech production values of VOT. The effect was obtained despite a fairly wide range of individual VOT values (cf. Table 1 and Fig. 2). In addition, the effect occurred for the male and female Ss to an approximately equal extent, despite the fact that the adapting stimuli simulated the output of a male vocal tract.

Since subvocalization was minimized in the experiment, the effect for the /pi/ utterances is probably central in origin and does not operate at the level of peripheral motor control. A consideration of the direction of the adaptation shift for /pi/ and the range of individual VOT values adds support to this claim and provides further information about the nature of the effect. From these considerations, I will argue below that the adaptation effect represents the fatiguing of a single mechanism utilized during both speech perception and speech production. This account of the effect will be contrasted with explanations based on (a) compensation for the effects of perceptual distortion produced during repetitive listening to the /pi/ adapting stimulus, and (b) voluntary mimicry of the perceived adapting stimulus.

Table 1
Mean VOT Values (in Milliseconds of VOT) for Each Subject in Each Test Condition

	/pi/ Utterances Adapt With		/bi/ Utterances Adapt With	
S	/i/	/pi/	/i/	/bi/
B.B.	68.2	50.8	−11.7	5.7
D.B.	66.8	85.0	6.4	−41.4
S.B.	25.9	20.8	−19.7	−9.6
A.C.	94.9	75.9	−23.9	−61.1
J.C.	43.3	37.9	11.3	5.5
L.C.	33.1	29.4	2.3	3.6
L.Ch.	61.0	47.6	8.4	6.3
L.Co.	77.1	71.4	−51.9	−63.8
G.D.	53.3	61.2	14.0	12.2
S.G.	91.0	81.3	6.7	0.5
H.H.	72.5	59.9	5.7	7.5
M.H.	48.4	46.2	7.0	8.5
P.H.	38.8	44.0	−22.8	−41.5
J.I.	88.3	74.4	11.1	16.2
C.L.	93.1	90.1	−10.3	−3.7
H.S.	67.6	58.4	10.7	11.4
Grand Mean	64.0	58.4	−3.5	−9.0

Note–Negative VOT values indicate that voicing onset occurred prior to the release burst of the consonant; positive values indicate that voicing onset occurred after the release burst (see text).

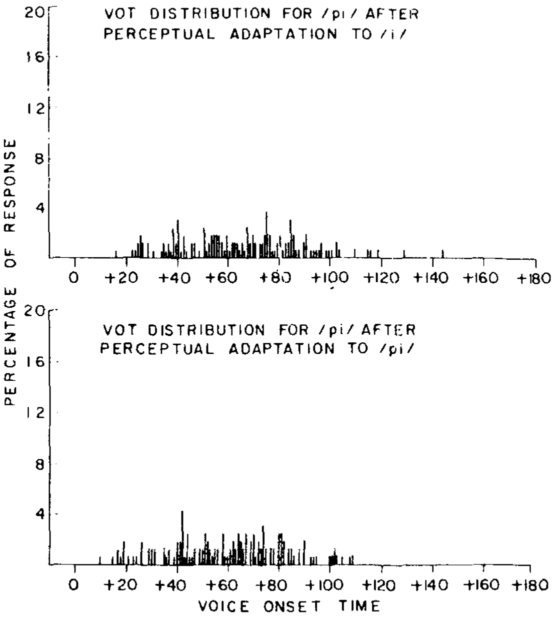

Fig. 2. Frequency distribution of the /pi/ utterances for the group of 16 Ss.

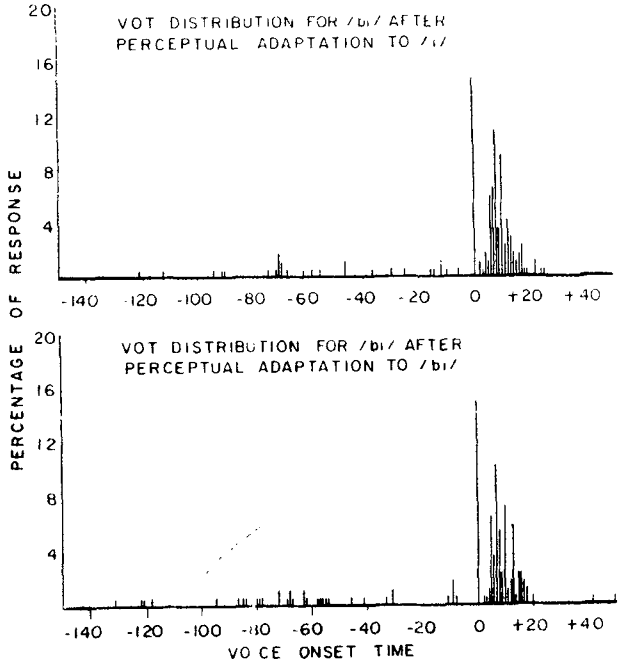

Fig. 3. Frequency distribution of the /bi/ utterances for the group of 16 Ss.

To provide evidence that the effect stems from the fatiguing of a perceptuomotor component of the speech system, we must, at the very least, establish that the present effect works in the same direction as the shifts obtained during perceptual adaptation for the *voicing* feature. The perceptual adaptation studies of Eimas and Corbit (1973) and Eimas et al (1973) show that after adaptation to a voiceless stop, some VOT stimuli identified as voiceless in the unadapted state were identified as voiced after adaptation. *We can infer from this finding that a given VOT stimulus was perceived as having a shortened VOT after adaptation to a voiceless stop.*[5] If the shifts for speech production represented the same adaptation effect, the articulated VOT values for /pi/ should have decreased after perceptual adaptation to /pi/, and our results confirm this prediction.

It should be clear that the present results cannot be accounted for by an effect of perceptuomotor compensation (cf. Held & Freedman, 1963). According to the compensation hypothesis, the articulated VOT values for /pi/ should have become *longer* in order to compensate for the perceived shortening of the VOT value of the voiceless adapting stimulus, contrary to fact. Since Ss in this experiment were instructed to make their responses as quickly and automatically as possible on each trial, it is not surprising that compensation (at least in voluntary form) did not govern the results.

Another possibility—that Ss simply mimicked the perceived VOT value of the /pi/ adapting stimulus—can also be ruled out on the basis of the present data. The mimicry interpretation cannot account for the wide range of individual VOT values obtained after adaptation to /pi/. Some Ss showing a decline in VOT after adaptation to /pi/ produced shifts in VOT directed away from the VOT value of the adapting stimulus (+80 msec), whereas other Ss showing the shortening effect produced VOT shifts in the direction toward the adapting stimulus value. In effect, the direction of shift was not systematic in relation to the absolute VOT value of the adapting stimulus; the VOT values for nine Ss shifted away from the +80 value of the adapting stimulus, while the VOTs for seven Ss shifted toward it. This finding, plus the generally wide range of VOT values for the /pi/ utterances, indicates that voluntary mimicry of the perceived adapting stimulus cannot account for the present effect.

Having provided some evidence that the shortening effect for the /pi/ utterances represents the fatiguing of a perceptuomotor aspect of the speech system and neither compensation nor mimicry, we now turn to the problematic question, "Why was a systematic effect of adaptation observed for the /pi/ utterances but not for the /bi/ utterances?" Two important differences between the processing of these syllables may lead to an explanation. With regard to speech production, an examination of Figs. 2 and 3 shows that the distribution of VOT values for the /pi/ and /bi/ utterances differed greatly in their distributional type. Whereas the VOT values for /pi/ were distributed in an approximately Gaussian fashion in both test conditions, the VOT values for /bi/ provided a poor fit to the Gaussian distribution, and in addition showed a strong clustering of responses within the relatively narrow range between 0 and 20 msec VOT.[6] The basic difference in distributional pattern of the /pi/ and /bi/ VOT values occurred for the individual Ss as well as for the group as a whole. This same distributional difference was also found by Lisker and Abramson (1964) in their original study of VOT production for the syllables /pa/ and /ba/. The presence of a strong response clustering between 0 and 20 msec VOT indicates that Ss have a well-defined target region of VOT when uttering the syllable /bi/, unlike the case for /pi/ utterances. One might speculate on this basis that /bi/ utterances would be less susceptible to the effect of perceptuomotor adaptation.

Perceptual evidence from the studies of Eimas and Corbit (1973) and Eimas et al (1973) is in accord with the notion that the voiced stops are less susceptible to adaptation. In these perceptual studies, it was found in a variety of test conditions that the voiceless stops were more affected by selective adaptation than were their voiced counterparts.

Considering the correlated differences in perception and production between the adaptation effects for voiced vs voiceless stop consonants, it seemed reasonable to ask a further question, namely whether the processing of voiced vs voiceless stops is carried out independently of each other. This question, of particular importance for evaluating models of the adaptation effects (see below), was taken up in a second perceptuomotor experiment.

This experiment was designed to test whether perceptual adaptation to /pi/ would alter the VOT values of /bi/ utterances, and vice versa. Eight Ss, six of whom had served in the main experiment, listened to repetitions of the syllables /i/, /pi/, and /bi/ as before, only in this case the repetitions of /pi/ immediately preceded /bi/ utterances, and vice versa. The 320 utterances were analyzed for VOT in the same manner used in the main experiment.

No systematic effect of adaptation was found for either the /pi/ or the /bi/ utterances (see Table 2).[7] The results of the experiment are of interest because they (a) rule out any remaining explanations of the original adaptation effect for /pi/ based on notions other than an effect that is dependent on the *voicing* property of the adapting syllable, and (b) suggest that the mechanisms for processing voiced vs voiceless stop consonants operate independently of each other. The claim for independent processing is supported by the additional fact that the VOT distributions for voiced vs voiceless consonants show virtually no overlap for individual speakers, contextual factors being equated (cf. Lisker & Abramson, 1964, and present data).

The claim for independent processing of voiced and voiceless stops makes it difficult to account for the perceptual adaptation effects obtained by Eimas and Corbit (1973) as well as for the perceptuomotor effect

Table 2
Mean VOT Values (in Milliseconds of VOT) for Each Subject in Each Test Condition of Experiment 2

	/pi/ Utterances Adapt With		/bi/ Utterances Adapt With	
S	/i/	/bi/	/i/	/pi/
B.B.	61.9	62.5	−34.3	−10.3
D.B.	71.5	70.2	7.8	−7.3
R.B.	32.4	39.5	5.3	5.0
M.C.	61.1	57.8	−0.9	−7.9
L.Ch.	64.4	58.7	10.8	8.6
L.Co.	87.9	90.8	−61.2	−58.3
S.G.	62.6	58.1	4.3	2.5
H.S.	58.0	59.3	14.8	11.9
Grand Mean	62.5	62.1	−6.7	−7.0

Note—Negative VOT values indicate that voicing onset occurred prior to the release burst of the consonant; positive values indicate that voicing onset occurred after the release burst (see text).

observed in the main experiment here. Eimas and Corbit accounted for the perceptual adaptation shifts by proposing the existence of two detector mechanisms, each selectively sensitive to a range of VOT values. They proposed further that the range of VOT values to which each detector is sensitive partially overlaps the range of sensitivity of the other detector. The assumption of overlapping, or alternatively, of inhibition of one detector by the other, forms an essential aspect of the proposed explanation of the adaptation effects, since fatiguing one detector should result in a *directional shift* in VOT perception only if the fatigued detector operates in an opponent-process fashion with the unadapted detector. The effects observed here for speech production appear inconsistent with the notion that the voiced and voiceless stops are processed in such a conjoint manner. Although multistage models of the adaptation process can be designed to avoid this apparent inconsistency, none of these models has sufficient justification on independent grounds to warrant consideration here.

The important question of processing independence notwithstanding, the effect observed in the main experiment of the present study provides initial support for the existence of a mechanism that mediates both the perception and articulation of speech. Research with the perceptuomotor paradigm has been extended in a number of ways since the completion of this study to determine both the reliability and generality of the presently reported effect.[8] The overall findings of our additional work have substantiated each of the major trends observed here.

REFERENCES

Ades, A. E. How phonetic is selective adaptation? Experiments on syllable position and vowel environment. Perception & Psychophysics, 1974, 16, 57-62.

Ades, A. E. A bilateral component in speech perception. Journal of the Acoustical Society of America, 1974b, in press.

Bailey, P. Perceptual adaptation for acoustical features in speech. Speech perception: Report on research in progress in the Department of Psychology, The Queen's University of Belfast, Northern Ireland, 2.2, 29-34, 1973.

Bailey, P. J., & Haggard, M. P. Perception and production: Some correlations on voicing of an initial stop. Language & Speech, 1973, 16, 189-195.

Cooper, W. E. Contingent feature analysis in speech perception. Perception & Psychophysics, 1974a, 16, 201-204.

Cooper, W. E. Adaptation of phonetic feature analyzers for place of articulation. Journal of the Acoustical Society of America, 1974b, in press.

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

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

de Cordemoy, G. *A philosophical discourse concerning speech.* London: Martin, 1668.

Eimas, P. D., Cooper, W. E., & Corbit, J. D. Some properties of linguistic feature detectors. Perception & Psychophysics, 1973, 13, 247-252.

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

Held, R., & Freedman, S. Plasticity in human sensorimotor control. Science, 1963, 142, 455-462.

Huggins, A. W. F. A facility for studying perception of timing in natural speech. Quarterly Progress Report of the M.I.T. Research Laboratory of Electronics, 1969, 95, 81-83.

Lashley, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior; the Hixon Symposium.* New York: Wiley, 1951.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-461.

Lisker, L., & Abramson, A. S. A cross-language study of voicing in initial stops: Acoustic measurements. Word, 1964, 20, 384-422.

Lisker, L., & Abramson, A. S. The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967.* Prague: Academia, 1970. Pp. 563-567.

Stevens, K. N. Segments, features, and analysis by synthesis. In J. F. Kavanaugh and I. G. Mattingly (Eds.), *Language by eye and ear: The relationships between speech and reading.* Cambridge: M.I.T. Press, 1972. Pp. 47-52.

Stevens, K. N., & Klatt, D. H. Role of formant transitions in the voiced-voiceless distinction for stops. Journal of the Acoustical Society of America, 1974, 55, 653-659.

von Humboldt, W. *Über die Vershiedenheit des menschlichen Sprachbaues.* (Facsimile ed. of 1st German ed. of 1836). Bonn: Ferdinand Dümmlers, 1960.

NOTES

1. Cooper, W. E. Selective adaptation for acoustic cues of voicing in initial stops. Forthcoming.

2. Klatt, D. H. An acoustic theory of terminal analog speech synthesis and a control strategy for the replication of a natural utterance. Forthcoming.

3. Cf. Cooper (1974b) for perceptual data showing a close relation between identification of consonants in the unadapted state and after adaptation with an isolated vowel.

4. Two-tailed t test for correlated observations.

5. It is important to note that one should consider the perceptual data from the standpoint of the shift in perception for a given VOT stimulus, not from the standpoint of the shift in physical VOT to yield a constant voiceless response. The proper comparison between the perception and production data concerns shifts in perceived and produced VOT, not shifts in the voiced-voiceless response, since the latter is variable in the perceptual studies but is phonetically constant in the production task here.

6. Experiments conducted since the completion of this study[8] indicate that the obtained 16% responses at 0 msec VOT (see Fig. 3) is too high a percentage, resulting from the failure to pick up very low-amplitude /b/ bursts for some utterances. It should be noted also that Lisker and Abramson (1964) reported 75% /ba/ utterances having 0 msec VOT. Lisker and Abramson's measurements were based on an analysis of wide-band spectrograms, and they rounded off each measurement to the

nearest 5 msec of VOT. It seems certain that the highly elevated response peak obtained in their study as well as in the present study was due to a failure to notice very low-amplitude onset bursts. As for the present research, it is important to point out that the generally strong clustering of responses between 0 and 20 msec VOT continues to show up consistently in additional work with improved measuring techniques.

7. Although no systematic shift was obtained for the group data /bi/ utterances, there existed a clear trend for the VOT values to decline afte;perceptual adaptation to /pi/ for those Ss having a nonnegative mean VOT in the control condition. There was no highly systematic relation in turn, however, between the performance of such Ss after adaptation in this experiment and their earlier performance in the main experiment.

8. Cooper, W. E., & Lauritsen, M. R. Feature processing in the perception and production of speech. Forthcoming.

(Received for publication January 15, 1974;
revision received May 2, 1974.)