Phonetica
Editor: K. Kohler, Kiel
Publishers: S. Karger, Basel
Separatum (Printed in Switzerland)

Phonetica *36:* 373–383 (1979)

Some Relationships between Speech Production and Perception

FREDERICKA BELL-BERTI, LAWRENCE J. RAPHAEL, DAVID B. PISONI and JAMES R. SAWUSCH

Haskins Laboratories, New Haven, Conn.; Herbert H. Lehman College, City University of New York, New York, N.Y.; Indiana University, Bloomington, Ind., and State University of New York at Buffalo, N.Y.

Abstract. EMG studies of the American English vowel pairs /i-ɪ/ and /e-ɛ/ reveal two different production strategies: some speakers appear to differentiate the members of each pair primarily on the basis to tongue height; for others the basis of differentiation appears to be tongue tension. There was no obvious reflection of these differences in the speech waveforms or formant patterns of the two groups. To determine if these differences in production might correspond to differences in perception, two vowel identification tests were given to the EMG subjects. Subjects were asked to label the members of a seven-step vowel continuum, /i/ through /ɪ/. In one condition each item had an equal probability of occurrence. The other condition was an anchoring test; the first stimulus, /i/, was heard four times as often as any other stimulus. Compared with the equal-probability test labelling boundary, the boundary in the anchoring test was displaced toward the more frequently occurring stimulus. The magnitude of the shift of the labelling boundary was greater for subjects using a production strategy based on tongue height than for subjects using tongue tension to differentiate these vowels, suggesting that the stimuli represent adjacent categories in the speakers' phonetic space for the former, but not for the latter, group.

Introduction

It is generally true that studies of human speech communication have been directed to questions of how speech is perceived or how speech is produced, with little direct investigation, and much speculation, about how these two events may be linked. While different schools have posited different cause-effect directions between these events (for example, the motor theory of speech perception and the acoustic theory of speech production), few, if any, suggest that the two are completely independent. For, in whatever direction the relationship proceeds, it is the acoustic signal produced by the human vocal tract and perceived by the human auditory system for which the theories must account [LIBERMAN

et al., 1962, 1967; LADEFOGED *et al.*, 1972; STEVENS and HALLE, 1967; STEVENS and HOUSE, 1972; COOPER, 1972].

Some experimental support for the view that speech production and perception are mediated by a common mechanism may be found in several recent studies by COOPER [1974], COOPER and LAURITSEN [1974], and COOPER and NAGER [1975], dealing with perceptuo-motor adaptation. These studies have shown that immediately after listeners are presented with many repetitions of a voiceless stop consonant, their voice onset time (VOT) values decrease as they produce voiceless stop-plus-vowel sequences. These results were similar to the results of selective adaptation experiments in which repeated listening to an adaptor altered the perception of a test series varying in VOT [EIMAS and CORBIT, 1973]. The perceptuo-motor adaptation studies suggest that the interaction between speech production and perception can be directly demonstrated. This might be accomplished by uncovering already existing (as opposed to experimentally induced) differences between the production strategies of two (or more) populations of speakers for a given class of sounds, and then showing that those differences were isomorphic with the differing perceptual behaviors of the populations for the same class of sounds.

The notion that different groups of speakers employ different articulatory strategies in achieving the same phonological goals finds some support in the literature. Such differences have been observed in both vowel and consonant articulation (even for speakers sharing the same dialect). Differences in articulatory strategies used in vowel production have been reported by LADEFOGED *et al.* [1972], as well as by RAPHAEL and BELL-BERTI [1975]; both of these reports included studies of the English front vowels /i-ɪ-e-ɛ-æ/. Interpreting the LADEFOGED *et al.* [1972] cinéfluorographic data from 6 speakers in light of RAPHAEL and BELL-BERTI's [1975] electromyographic (EMG) data from 3 speakers, it appears that there are at least two strategies for differentiating these vowels. The description of the muscular component of these strategies was derived from studying the activity of the genioglossus muscle, which is primarily responsible for elevating and bunching the tongue [RAPHAEL and BELL-BERTI, 1975; RAPHAEL *et al.*, 1979]. In the first, speakers vary tongue height and the activity of the genioglossus muscle through the series /i-ɪ-e-ɛ-æ/; in the second, speakers vary some other characteristic of vowel articulation so that tongue height and genioglossus muscle activity are progressively lower through the series /i-e-ɪ-ɛ-æ/.

Evidence for differences in articulatory strategies used in consonant articulation have been reported by several authors. For example, BELL-BERTI [1975] has described two mechanisms for controlling pharyngeal cavity expansion in the maintenance of voicing during stop consonant occlusion; and BRONSTEIN [1960] and BORDEN and GAY [1978] have described two tongue-tip positions for the articulation of /s/.

It is possible that the alternative articulatory strategies found among groups of speakers are reflected in alternative perceptual strategies among these groups. The experiments reported here were designed to investigate the possibility of such a perceptuo-productive isomorphism for several members of the class of English vowel sounds.

The members of the front series of English vowels /i-ɪ-e-ɛ/ have been variously described in the phonetics literature either as differing among themselves in both tongue height and duration; or as differing within the pairs /i-ɪ/ and /e-ɛ/ in tongue tension (with consequent differences in duration) and differing between the pairs in tongue height. In an earlier study of the production of these and other vowels [RAPHAEL and BELL-BERTI, 1975], we obtained EMG recordings from the extrinsic tongue muscles of 3 speakers of American English to discover which, if any, of these muscles displayed a difference in overall amount of activity corresponding to the traditional 'tense-lax' distinction between members of the English vowel pairs /i-ɪ/, /e-ɛ/ and /u-ʊ/. The data we gathered provided support for the notion that tension is a necessary, or sufficient, differentia of production for some speakers, but not for others. In the present study we hoped to determine whether differences in vowel production might in some way be related to differences in vowel perception, particularly vowel identification. To this end we collected vowel-identification as well as EMG data from a group of 10 subjects.

The Production Experiment

Method

EMG potentials were recorded with bipolar hooked-wire electrodes inserted percutaneously into the genioglossus muscle. The action of this muscle is to bunch the mass of the tongue and draw it forward in the oral cavity, especially for the high- and mid-front vowels [SMITH, 1970; HARRIS, 1971; RAPHAEL and BELL-BERTI, 1975; KAKITA, 1976; RAPHAEL *et al.*, 1979].

The utterances used in this experiment included the vowels /i/, /ɪ/, /e/, and /ɛ/, produced in a /əpVp/ frame. The utterances were placed in random lists which were read by each subject until 18–30 tokens of each utterance type were recorded.

The EMG potentials and the speech signal were recorded on an FM tape recorder. The onset of voicing of the stressed vowel was identified visually for each syllable from oscillo-

graphic displays. The repetitions of each utterance type were aligned with reference to this point, and the EMG data subsequently computer sampled and averaged.

Results

Figure 1 contains examples of the two patterns of muscular contraction found for the vowels. In one pattern, there is a decreasing order of activity corresponding to the traditional articulatory description of tongue height for the front vowels (fig. 1a).: peak activity decreases through the vowel series /i-ɪ-e-ɛ/. Further, contrasting the EMG curves of the longer vowels /i/ and /e/ between −200 msec and +200 msec with those of the shorter vowels /ɪ/ and /ɛ/ between −200 msec and +100 msec, it is evident that the former are bi-modal, perhaps reflecting this subject's strategy for diphthongizing the 'tense' vowels. In the second pattern of muscular contraction (fig. 1b) there are greater, and almost identical, levels of muscle activity for the two 'tense' vowels /i/ and /e/, and a considerably lesser degree of activity for the two 'lax' vowels, /ɪ/ and /ɛ/; that is, there is a decreasing order of activity through the vowel series /i-e-ɪ-ɛ/.[1] Further, the EMG curves are smooth and unimodal, compared with those of the other pattern, described above, and perhaps reflecting a different strategy for diphthongizing the 'tense' vowels for this speaker.

Although both these patterns of muscular contraction preserve the 'tense-lax' distinction between /i-ɪ/ and /e-ɛ/, they do so in markedly different ways. In the first pattern, corresponding to the traditional picture of tongue height for front vowels, the traditionally 'lax' /ɪ/ is characterized by the data as being more tense than the traditionally 'tense' /e/, and being very close to /i/ with regard to tension – although not with regard to duration. That is, peak EMG activity decreases through the series /i-ɪ-e-ɛ/, suggesting a distinction between vowels which reflects the usual description of a tongue height or tongue bunching continuum. The second pattern, however, does not correspond to the usual description of tongue height, with /i/ and /e/, both traditionally described as 'tense' vowels, showing considerably more genioglossus activity than 'lax' /ɪ/ and /ɛ/. That is, for this pattern EMG activity decreases through the series /i-e-ɪ-ɛ/.

The question posed by these production data for theories of speech perception is whether apparent differences in the necessary differentia

[1] It must be remembered, however, that the differences in absolute microvolt potential and duration of activity separating /i/ from /e/ and /ɪ/ from /ɛ/ are quite small.

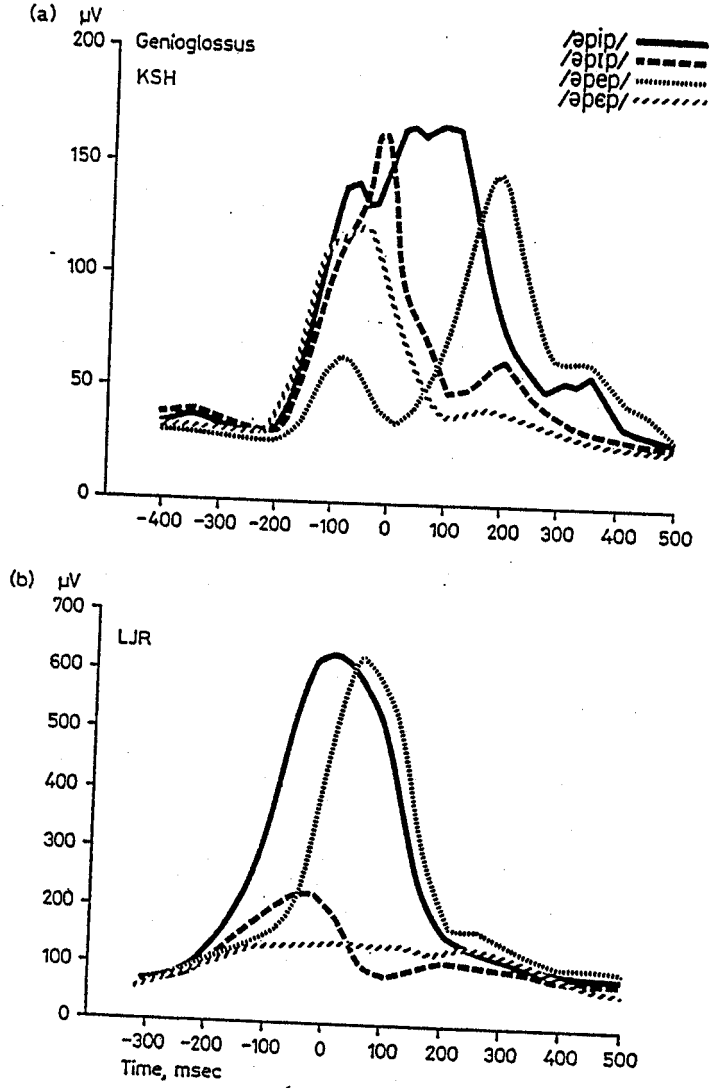

Fig. 1. Averaged EMG activity of the genioglossus muscle for the vowels /i, ɪ, e, ɛ/, for: (a) subject KSH; (b) subject LJR.

for production are reflected in differentia employed in perception. That is, do talkers who rely on different mechanisms for producing vowels also rely on different properties or strategies in perceiving them? In order to answer this question we turned to tests of perception.

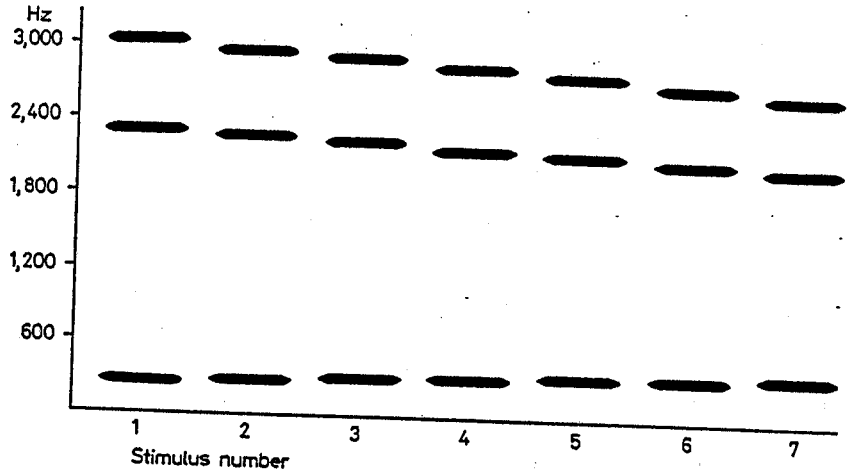

Fig. 2. Schematic spectrograms of the first three formants of the seven synthesized vowel stimuli.

The Perception Experiments

Methods

Two types of vowel perception tests were administered to each of two groups of listeners. The first group of listeners consisted of 137 students at either Indiana University or the State University of New York at Buffalo. The second group consisted of 10 listeners who were students or research associates at Haskins Laboratories.[2] The tests were constructed from a continuum of vowel stimuli ranging from /i/ to /ɪ/ in seven steps. The vowels were synthesized on the vocal tract analogue synthesizer at the Research Laboratory of Electronics at MIT (fig. 2). The frequencies of the first three formants were varied in equal logarithmic steps, while those of the fourth and fifth formants were held constant at 3,500 and 4,500 Hz, respectively, for all seven stimuli. Vowel duration was 300 msec, with rise and decay times of 50 msec. The fundamental frequency fell linearly from 125 to 80 Hz during the duration of each vowel.

These seven vowel stimuli were recorded on magnetic tape as two test series. In the control series, each of the stimuli occurred 10 times; in the anchor series, stimulus 1 (/i/) occurred 40 times and each of the other stimuli occurred 10 times. In both series, stimuli were presented one at a time with a 4-sec pause between successive items. Subjects were asked to identify the stimuli as either /i/ or /ɪ/. All subjects listened to two presentations of the control series followed by the anchor series.

Results

The pooled identification data for the group of 10 subjects, for both control and anchor conditions, are shown in figure 3a. The analogous data for the group of 137 subjects are shown in figure 3b. Each of the

[2] Except for the 3 subjects whose EMG data were reported by RAPHAEL and BELL-BERTI [1975], the subjects were not told the purpose of the perception or production experiments until after both sets of data had been collected.

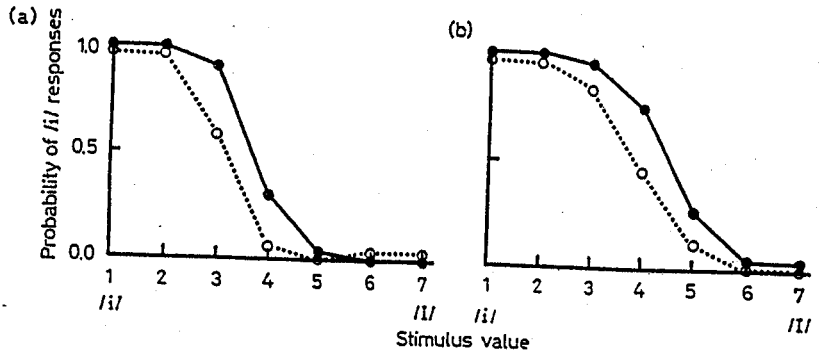

Fig. 3. Vowel identification data in equal-probability (●——●) and anchor-condition (○- - -○) tests. *(a)* Pooled identification functions from the original 10 subjects. *(b)* Pooled identification functions from the additional 137 subjects.

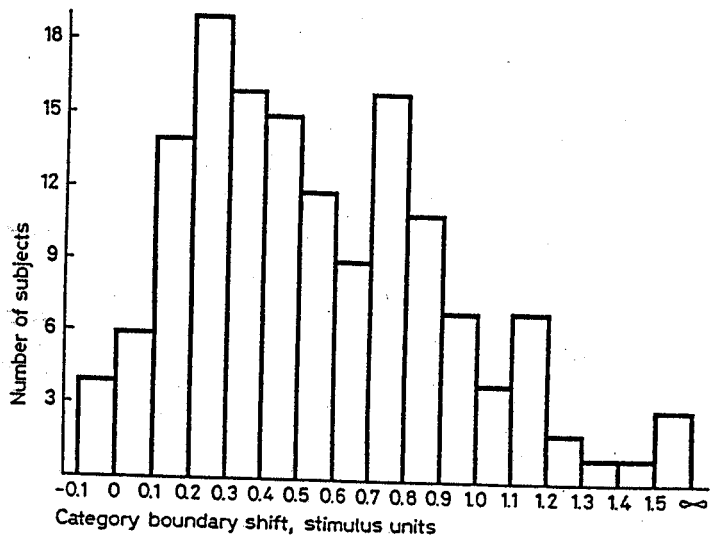

Fig. 4. Distribution of shifts in anchoring condition, for 147 subjects.

group of 10 subjects showed a shift in the /i-ɪ/ category boundary toward /i/ for the anchor series compared with the control series. The group shift, indicated in figure 3a, was highly significant ($t_{(9)} = 4.73$, $p < 0.001$ using a two-tailed, correlated t test). Of the group of 137 subjects, 133 showed the expected shift in the category boundary toward /i/, a highly significant result ($p < 2 \times 10^{-10}$ using a sign test).

Figure 4 shows the distribution of magnitudes of category boundary shifts, expressed in terms of stimulus steps, in the anchoring condition

for all 147 subjects. This distribution displays prominent peaks in the 0.2- to 0.3-unit range and in the 0.7- to 0.8-unit range, implying the existence of real differences among listeners in susceptibility to /i/-anchoring. It should also be noted that the average boundary shift for the 6 EMG subjects who show an /i-e-ɪ-ɛ/ vowel ordering was 0.25 stimulus units, a value coinciding with the first peak in the distribution shown in figure 4. The average boundary shift for the 4 EMG subjects who showed an /i-ɪ-e-ɛ/ vowel ordering was 0.88 stimulus units, a value slightly larger than the second peak in the group distribution, but reasonably close to the 0.7- to 0.8-unit range.

Comparison of the Production and Perception Data

Of the 10 EMG subjects, 4 displayed a pattern of production in which the traditionally delineated order of tongue height for the front vowels is reflected in peak EMG activity for the genioglossus muscle; these 4 subjects demonstrated relatively large anchoring effects in the unequal-probability condition, as measured by the magnitude of shift in the locus of the phoneme boundary brought about by anchoring. The results for these subjects are shown at the right of figure 5. The 6 remaining EMG subjects displayed a pattern of production in which the feature of tongue tension is reflected in the EMG data from the genioglossus muscle; these subjects demonstrated relatively small anchoring effects, and are shown at the left of figure 5.

Assuming that the relationship between EMG patterns and size of anchoring effects demonstrates the existence of an interaction between individual strategies for speech production and perception, there are

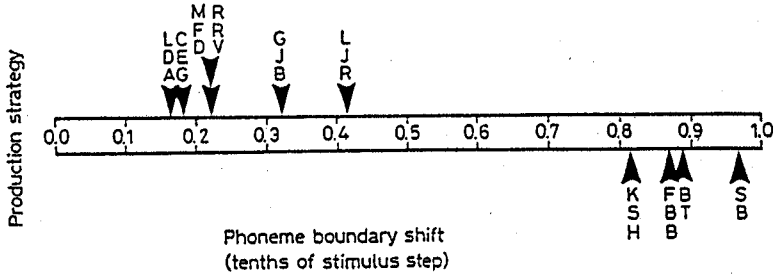

Fig. 5. Distribution of phoneme boundary shifts for 10 subjects in EMG experiment.

several possible explanations for these results. One relatively straightforward explanation might be that there is a correspondence between the acoustic patterns of the stimuli in the anchoring experiment and the particular articulatory strategies of the subjects. For example, we might see the differences in production strategies reflected in the spectral or temporal patterns of the resulting speech. In order to explore this possibility, we subjected the speech samples collected during the EMG experiments to acoustic analysis.

The Acoustic Analysis

Methods

The acoustic analysis was performed using a digital waveform and spectral analysis system. Spectral analysis involved tracing formant patterns from digital spectrograms, using a storage oscilloscope and digitizing tablet, and storing first-, second- and third-formant frequency values for the stressed vowels of the /əpVp/ syllables. Average formant frequencies for 4 subjects, 2 from each group, were computed for a minimum of 15 tokens of each of the four utterance types: /əpip/, /əpɪp/, /əpep/, and /əpɛp/. Durational measurements were made from displays of the digitized waveform. The average durations of the burst-to-closure portions, and of the voiced portions, of the stressed syllables were calculated for all 10 EMG subjects. Again, a minimum of 15 tokens of each utterance was measured and averaged.

Results

The spectral analysis failed to reveal any systematic differences between formant frequency patterns of speakers using differing articulatory strategies. Similarly, no systematic differences were revealed in the durations of the burst-to-closure, or in the voiced, portions of the stressed syllables, expressed either as absolute durations or as ratios of short- to long-vowel syllables.

Discussion

Our data have revealed a high degree of correlation between the speech production strategies used by speakers and their performance on an anchoring task. That is, relatively large anchoring effects were found for the subjects displaying a pattern of muscular activity that parallels the ordering of the vowels on the perceptual test continuum, while relatively small anchoring effects were found for the subjects displaying a pattern of muscular activity that does not correspond to the ordering of the vowels in the test continuum, with genioglossus activity decreasing from /i/ to /e/ to /ɪ/, although the test continuum contained no per-

ceptual category between /i/ and /ɪ/. For these subjects, susceptibility to anchoring effects in vowel identification may be substantially reduced because the test stimuli represent vowels which are not contiguous within each subject's articulatory or phonetic space.

One hypothesis for explaining these data is that differences in perception reflect the different articulatory strategies of tongue height and tongue tension as ways of realizing the vowels in the set /i, ɪ, e, ɛ/. One might anticipate that these two subject groups would reverse their relative positions along the boundary-shift continuum if presented with a set of vowels that varied from /i/ to /e/, without containing intermediate stimuli identified as /ɪ/. This hypothesis, and alternatives not considered here[3], should be extended to other perceptual, articulatory and acoustic dimensions. Although many investigators have argued for the existence of a common mechanism linking speech perception and production, the evidence typically cited in support of these views has been, by necessity, often indirect. We believe that the data of the present experiment, although preliminary, provide a more direct and convincing demonstration of the existence of some common mechanism or process that mediates at least some aspects of the production and perception of vowels. The extent to which these initial findings can be replicated and then extended to other phonetic distinctions obviously awaits the results of additional experiments. Such experiments should be specifically designed to reveal interactions between the dimensions of speech production and perception.

[3] For example, temporal characteristics such as the rate of formant transitions in a CVC syllable may differ between these subject groups.

Acknowledgments

This work was supported by NINCDS grants NS-13617 and NS-05332, NICHHD grant HD-01994, BRSG grant RR-05596 and NIDR grant DE-01774 to the Haskins Laboratories; and by NINCDS grant NS-12179 and NIMH grant MH-24027 to the Department of Psychology, Indiana University; and SUNY Research Foundation Grant-in-Aid to J. R. SAWUSCH.

References

BELL-BERTI, F.: Control of pharyngeal cavity size for English voiced and voiceless stops. J. acoust. Soc. Am. *57:* 456–461 (1975).

BORDEN, G. J. and GAY, T.: On the production of low tongue tip /s/: a case report. J. Commun. Disorders *11:* 425–431 (1978).

BRONSTEIN, A. J.: The pronunciation of American English (Appleton-Century-Crofts, New York 1960).

COOPER, F. S.: How is language conveyed by speech? in KAVANAGH and MATTINGLY Language by ear and by eye (MIT Press, Cambridge 1972).

COOPER, W. E.: Perceptuomotor adaptation to a speech feature. Percept. Psychophys. *16:* 229–234 (1974).

COOPER, W. E. and LAURITSEN, M. R.: Feature processing in the perception and production of speech. Nature, Lond. *252:* 121–123 (1974).

COOPER, W. E. and NAGER, R. M.: Perceptuo-motor adaptation to speech: an analysis of bisyllabic utterances and a neural model. J. acoust. Soc. Am. *58:* 256–265 (1975).

EIMAS, P. D. and CORBIT, J. D.: Selective adaptation of linguistic feature detectors. Cognitive Psychol. *4:* 99–109 (1973).

HARRIS, K. S.: Action of the extrinsic musculature in the control of tongue position: preliminary report. Haskins Lab. Status Rep. Speech Res. *SR-25/26:* 87–96 (1971).

KAKITA, K. M.: Activity of the genioglossus muscle during speech production: an electromyographic study; unpublished DMS diss., University of Tokyo (1976).

LADEFOGED, P.; DECLERK, J. D.; LINDAU, M., and PAPÇUN, G.: An auditory-motor theory of speech production. Univ. Calif., Los Angeles, Working Pap. Phonet. *22:* 48–75 (1972).

LIBERMAN, A. M.; COOPER, F. S.; HARRIS, K. S., and MACNEILAGE, P. F.: A motor theory of speech perception; in Proc. Speech Communication Seminar (Stockholm 1962).

LIBERMAN, A. M.; COOPER, F. S.; SHANKWEILER, D. P., and STUDDERT-KENNEDY, M.: Perception of the speech code. Psychol. Rev. *74:* 431–461 (1967).

RAPHAEL, L. J. and BELL-BERTI, F.: Tongue musculature and the feature of tension in English vowels. Phonetica *32:* 61–73 (1975).

RAPHAEL, L. J.; BELL-BERTI, F.; COLLIER, R., and BAER, T.: Tongue position in rounded and unrounded front vowel pairs. Language Speech *22:* 37–48 (1979).

SMITH, T. S. J.: A phonetic study of the function of the extrinsic tongue muscles; unpublished PhD diss., UCLA (1970).

STEVENS, K. N. and HALLE, M.: Remarks on analysis by synthesis and distinctive features; in WATHEN-DUNN Models for the perception of speech and visual form (MIT Press, Cambridge 1967).

STEVENS, K. N. and HOUSE, A. S.: Speech perception; in TOBIAS Foundations of modern auditory theory (Academic Press, New York 1972).

Received: May 29, 1979; accepted: June 6, 1979.