383

LANGUAGE AND SPEECH, Vol. 25, Part 2, 1982

DYNAMICS OF VOWEL ARTICULATION*

PETER J. ALFONSO
University of Connecticut
and
Haskins Laboratories
and
THOMAS BAER
Haskins Laboratories

The dynamics of vowel articulation was studied by simultaneously monitoring articulator muscle activity, movement, and acoustic output while a single speaker produced multiple repetitions of 10 different vowels in a /əpVp/ environment. Analysis of the data from the three measurement levels provided a complete physiological description of the subject's vowel space and showed good agreement among the different levels of description. Intra-articulator timing measurements, e.g., the relationship between posterior genioglossus EMG activity and tongue movements, and inter-articulator timing measurements, e.g., the relationship between tongue and lip movements, indicated a differentiation in the timing of tongue movement components. Vowel-related vertical tongue movements always began at about the moments of implosion for the initial stop. Horizontal movements associated with front vowels began at the same time as vertical movements, whereas horizontal movements associated with back vowels began earlier, at the onset of the syllable. For certain vowel environments, tongue and other articulator anticipatory gestures were found to be linguistically significant since listeners were able to correctly label the vowel when presented with only the schwa segment of the syllable.

INTRODUCTION

There are many studies in the phonetic literature, based on various combinations of electromyographic (EMG), cinefluorographic, and acoustic data, that describe the positioning of various articulators, most notably the tongue, during the production of vowels. However, with the exception of a few experiments carried out at Haskins Laboratories and at the Research Institute of Logopedics and Phoniatrics at the University of Tokyo (e.g., Gay, Ushijima, Hirose and Cooper, 1974; Borden and Gay, 1978; and Kiritani, Sekimoto, Imagawa, Itoh, Ushijima and Hirose, 1976), none of these studies have incorporated simultaneous recording of all three types of measurement. The paucity of studies incorporating simultaneous measurements is most likely due to the inherent technical

---

* *Versions of this paper were presented at the 99th Meeting of the Acoustical Society of America, Atlanta, Georgia, April, 1980 and the Annual Convention of the American Speech-Language and Hearing Association, Detroit, Michigan, November, 1980. We wish to acknowledge the assistance of Linda D'Antonio in processing the X-ray data and Katherine S. Harris for her comments on earlier versions of this manuscript. This research was supported by NIH grant NS 13617 to Haskins Laboratories.*

difficulties of the methodology, since the information gained from simultaneous monitoring of the different levels of speech articulation, namely neuromuscular, articulator movement, and acoustic, would contribute significantly to our understanding of dynamic speech production.

With respect to vowel articulation, it would be worthwhile to establish the agreement among muscle activity underlying tongue movement, positioning of the tongue, and the resultant acoustic output during the production of various vowels for the same speaker. Simultaneous data such as these may shed some light on the controversy that still exists whether acoustic or articulatory parameters better represent phonetic feature specification for vowels. For example, Nearey (1977) argues that one reason that vowel phonetic features are better represented by acoustic parameters is that articulatory data are characterized by unsystematic intersubject variability that may not be phoneme dependent. Wood (1979), on the other hand, argues in favor of articulatory representation and claims that the observed articulatory inconsistencies are the result of measuring inappropriate articulatory parameters. Unfortunately, a problem that further complicates issues such as this is that often EMG, movement, and acoustic data collected from different experiments, which usually use different talkers, are used to make comparisons and assumptions about each measurement level. Similarly, simultaneous measurements would be useful for investigating other physiological bases of linguistic classification – for example the "tense-lax" vowel distinction. It would seem that the testing and formulation of models of vowel articulation would be significantly enhanced if they were based on a complete physiological description provided only by simultaneous measures.

For analyzing the dynamics of vowel production, simultaneous measurement at the three levels is more useful than any combination of two. That is, simultaneous measures allow for both intra-articulator timing measurements and inter-articulator timing measurements. Intra-articulator measurements may be used to explore the mechanisms for producing articulatory movements and their acoustic consequences by relating muscle activity to relevant articulator movement or by relating movement to acoustics. Inter-articulator measurements, such as tongue versus jaw relationships, are used to explore the organization of speech production as a linguistically conditioned motor act.

Perception represents still another level at which speech dynamics can be analyzed. It is well known that EMG, movement, and acoustic records associated with given speech segments often vary with the identity of neighboring segments. Presumably, the perceptual system is specialized to take these context effects into account on the basis of the listener's knowledge of the phonetic code (Liberman *et al.*, 1967, Liberman and Studdert-Kennedy, 1978). Thus, the phonetic code is the link between speech production and perception. Perhaps the nature of the phonetic code could be better understood if the signals giving rise to specific perceptual results are better defined at the three levels of speech production, and if the results of speech production measurements are submitted to perceptual analysis.

The purpose of this study was to investigate the dynamics of vowel articulation by simultaneously monitoring muscle activity (using electromyography), articulatory movements (using lateral cinefluorography), and acoustic output. Dynamics of ten

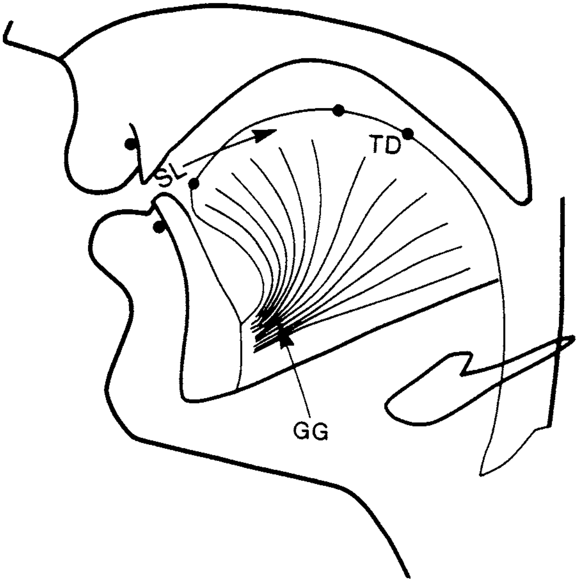


Fig. 1. Schematic representation of lead pellets attached to the tongue tip, blade, and dorsum, and to the upper and lower incisors. The arrows indicate the paths of insertion of the hooked wire electrodes for the genioglossus and superior longitudinal muscles.

different vowels were analyzed in syllables of the form /əpVp/ produced in isolation by a single speaker of American English with a New York City dialect. We wanted to: (1) "Map" the subject's vowel space in each of the three physiological measurement domains. (2) With respect to intra-articulator measurements, to further investigate the role of the genioglossus muscle in producing tongue movements. (3) With respect to inter-articulator measurements, to investigate the timing of tongue versus lip and jaw movements associated with different vowels in /əpVp/ context. (4) Finally, to investigate the perceptual significance of the acoustic consequences of vowel-related anticipatory tongue movements.

PHYSIOLOGICAL MEASUREMENTS

Methods

Cinefluorographic films were made at a rate of 60 frames per second. For these films, pellets were glued to the tongue tip, blade, and dorsum and to the upper and lower incisors, as indicated in Figure 1. A contrast medium was used to enhance the outline of the lips. In addition, a gold chain was laid on the floor of the nasal tract for monitoring velar movements. Velar movements will be reported in a separate paper. Jaw movements, indicated by the pellet on the lower incisors, were negligible for this speaker and therefore will not be addressed. Pellets on the tongue tip and blade could not be reliably tracked in all environments. Therefore, we will consider here only movements of the tongue dorsum.

EMG signals were recorded from the orbicularis oris muscle, the anterior belly of the digastric muscle, and from two muscles of the tongue, the superior longitudinal and the posterior part of the genioglossus. The paths of insertion of the hooked wire electrodes for the tongue muscles are also indicated in Figure 1. Low-intensity signals were recorded from the anterior belly of the digastric, consistent with the lack of jaw movement for this speaker. Signals from the other muscles showed patterns of activity that varied across vowels. Since we will restrict our discussion of movement data to the tongue dorsum, we will report EMG data from only the genioglossus muscle. Good quality acoustic recordings were made by a close-talking directional microphone.

During the X-ray filming, the subject read a randomized 20-word list producing two tokens each of the 10 vowels. We will refer to this portion of the experiment as the "X-ray run." He then continued without X-ray filming, producing an additional 20 tokens of each vowel to extend the base of the acoustic and electromyographic data. This portion of the experiment will be referred to as the "EMG-only run." The complete set of the subject's utterances were later presented to a panel of listeners in an identification task, and all utterances were unambiguously perceived as intended by the talker.

Measurements of pellet movements with respect to the reference pellet (upper incisor) were made on a frame-by-frame basis with the aid of a digitizing tablet. Electromyographic and acoustic data were processed using standard methods at Haskins Laboratories. More specifically, the EMG signals were rectified and integrated over a 5 msec window before sampling. The sampled signals were then smoothed using a 35 msec triangular function, and were then ensemble-averaged across repetitions. Formant frequencies were traced by hand from digital dot-density spectrographic displays with the aid of a digitizing tablet.

Electromyographic and acoustic data obtained during the X-ray run and from the EMG-only run were processed separately. For reasons discussed below, data from the EMG-only run were considered more appropriate for articulator timing measurements.

Results

The next three figures illustrate acoustic, movement, and electromyographic properties of the 10 vowels. Data shown in these figures were collected from the X-ray run.

Figure 2 shows an F_1 versus F_2 plot of the 10 vowels used in this study. Formant frequency trajectories were measured from each of the two tokens of the 10 utterances. The formant trajectories were then averaged for each utterance. Peak values for the averaged first and second formant frequencies are shown in Figure 2. The back vowels, with the exception of /a/, were all relatively high and were tightly grouped. However, the front vowels were spread out approximately along a diagonal, with the tense vowels /i/ and /e/ higher and more forward than their lax counterparts, /ɪ/ and /ɛ/.

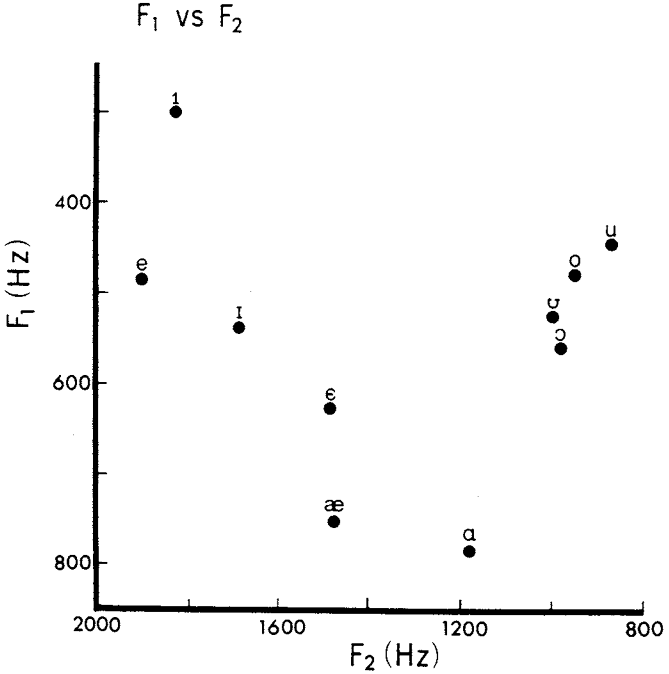


Fig. 2. Peak values in Hz of the first and second formants for the 10 vowels used in this study. Each data point represents the average of the two tokens of each vowel produced during the X-ray run.

Figure 3 shows the smoothed movement trajectories of the tongue dorsum pellet for each vowel during the interval from its voice onset until lip closure for the final consonant (that is, the vocalic period). Since analysis of the movement trajectories for the lower incisor pellet indicated no jaw movement during vowel production, tongue dorsum trajectories can be attributed to movements of the tongue alone. Movements along all of these trajectories, except the one for /ɔ/, are in an ascending direction and away from the center. The pattern of locations of the endpoints of these trajectories grossly resembles the vowel pattern in the acoustic domain shown in Figure 2, although it may be noted that the diphthongized vowels /e/ and /o/, as might be expected, do not fit this pattern as well as the remaining vowels.

Figure 4 shows the pattern of peak EMG activity for the posterior part of the genioglossus muscle for each of the ten vowels. Note that maximum EMG activity occurs during production of high tense vowels. Greatest activity is noted for /i/ and /e/ and somewhat less for /u/ and /o/. Among the remaining vowels, there is somewhat more activity for the front than for the back. Thus, comparison of Figures 2–4 demonstrates good agreement among the three types of measures made in this study.

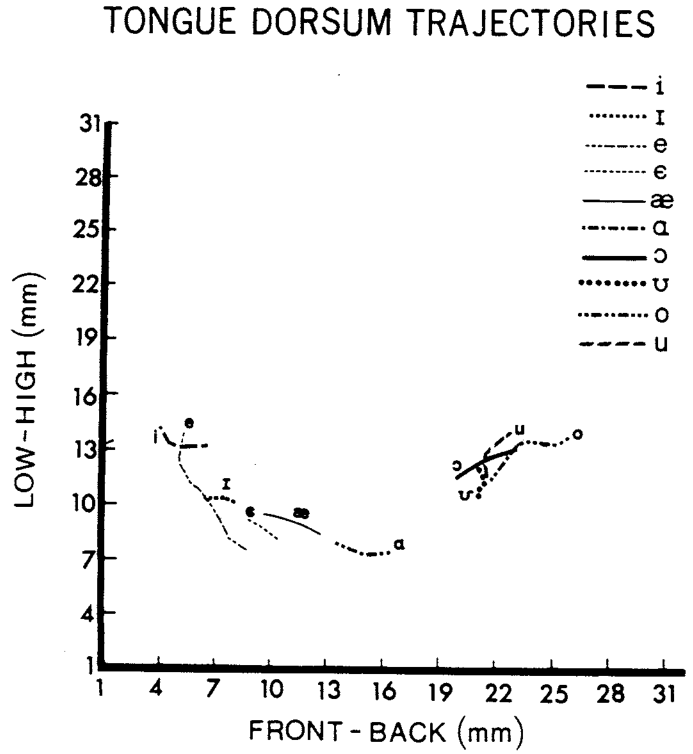


Fig. 3. Movement trajectories of the tongue dorsum pellet during the interval from the voice onset for the vowel to the lip closure for the final consonant. With the exception of /ɔ/, movements along the trajectories are in an ascending direction and away from the center. Each trajectory represents the average movement of two tokens.

In order to further characterize the subject's vowel space, data from the three measurement domains were used to describe the differences between the pairs /i–ɪ/, /e–ɛ/, and /u–ʊ/ produced by this speaker. These pairs have been referred to as "tense-lax" cognates within some linguistic classification systems. In the following discussion we will refer to the vowels /i, e, u/ as "tense" and to the vowels /ɪ, ɛ, ʊ/ as "lax." Considering the acoustic domain first, tense vowels were found to be of longer duration when measured from the release of the initial stop to the closure for the final stop. Fundamental frequency did not vary systematically with tongue height. Similar results for the same speaker were reported by Honda and Baer (1981) and Honda (1981).[1]

[1] *Using cinefluorography, Honda and Baer examined the relationships between movements of the tongue and the hyoid-larynx complex and observed that horizontal movement of the hyoid bone is strongly correlated with fundamental frequency change. They found relatively little variation in hyoid bone position and fundamental frequency for this speaker.*

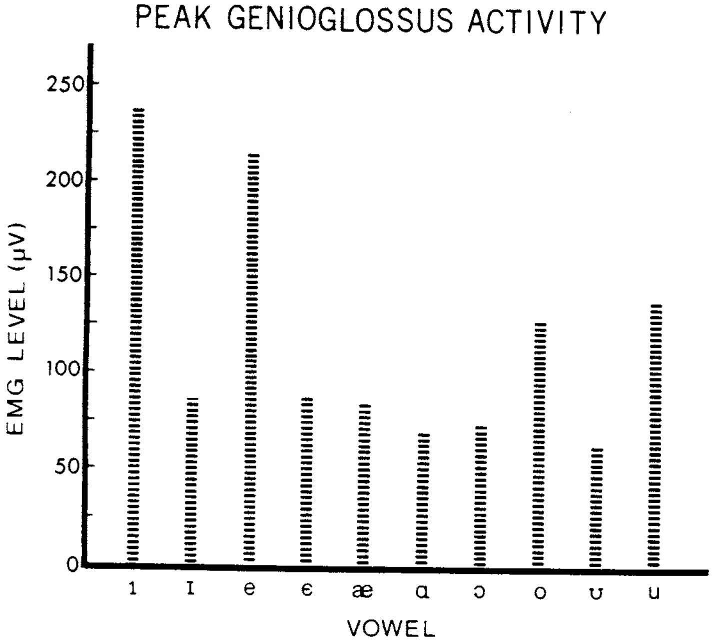


Fig. 4. Peak genioglossus EMG activity for each of the 10 vowels. Each data point represents the average of two tokens produced during the X-ray run.

The lax cognates had greater amplitude than their tense counterparts. The acoustic vowel space shown in Figure 2 demonstrates that the lax vowels were more central than the tense vowels. The acoustic bases for the tense-lax distinction are summarized in Table 1. In the articulatory domain, Figure 3 demonstrates that tongue positions for lax vowels are more central than for their tense cognates. These relationships were also demonstrated in the acoustic domain shown in Figure 2. For all tense-lax pairs, most remarkably in the /e–ɛ/ pair, the tense vowel is diphthongized . It shows more movement over a greater amount of time than its lax cognate. Finally, in the EMG domain, Figure 4 demonstrates that there is greater genioglossus muscle activity for the tense vowels /i/, /e/, and /u/ than for the lax cognates /ɪ/, /ɛ/, and /ʊ/.

Because of the stochastic nature of single motor unit activity, EMG signals averaged from an ensemble of many tokens of an utterance are more appropriate than those representing few tokens in analyzing the dynamics of speech production (Kewley-Port, 1977; Harris, 1981; Baer, Bell-Berti and Tuller, 1979).[2] Since the X-ray run consisted of only two tokens of each utterance, data from the EMG-only run, which consisted of 20

[2] *At low levels of activity, few motor units within the recording field of the hooked-wire electrodes may be recruited. Thus, the population of firings for estimating the activity of the muscle as a whole may be too small, unless a large number of repetitions is used.*

TABLE 1.

Acoustic values of the vowels produced during the X-ray run.
Shown are frequencies in Hz, duration in msec, and relative amplitude in dB.

Vowel	F_0	F_1	F_2	Duration	Amplitude
i	161	299	1830	178	-3.7
ɪ	171	539	1689	165	0
e	161	484	1905	215	-2.1
ɛ	158	624	1484	188	-0.2
æ	164	751	1477	248	-0.9
ɑ	158	783	1182	220	-0.4
ɔ	159	558	986	240	-2.3
o	166	479	956	215	-1.1
ʊ	163	525	1005	148	-1.3
u	163	444	877	200	-2.2

tokens, were used to study the dynamics of these productions. The validity of comparisons between these data and the movement data from the X-ray run depends on the subject's ability to produce similar articulations in both runs. Therefore, vowels produced during the X-ray and EMG-only runs were compared on the basis of both acoustic and electromyographic characteristics. Acoustic similarity was quantified by correlating the peak F_1 and F_2 values of corresponding vowels produced in the two runs. These correlations were fairly high, namely 0.93 for F_1 and 0.98 for F_2, suggesting a high degree of similarity between vowel productions. Similarly, comparison of peak genioglossus EMG levels across conditions yielded a correlation of 0.98, again suggesting that the subject produced similar gestures during both runs. However, correlations between the EMG activity patterns, as time functions, for the individual vowels produced variable results. For /i/ and /e/, the vowels with the highest levels of genioglossus activity, the correlations were high, 0.98 and 0.87 respectively. Vowels with lower levels of activity, however, showed substantially lower correlations. This result is predictable, for the reasons stated above. Thus, these analyses indicate that the utterances were produced in a similar manner in the two runs, and that data from the EMG-only run are more appropriate for dynamic measurements.

With these comments in mind, we turn our attention to timing measurements. As an example of intra-articulator measures, Figures 5 and 6 demonstrate the relationship

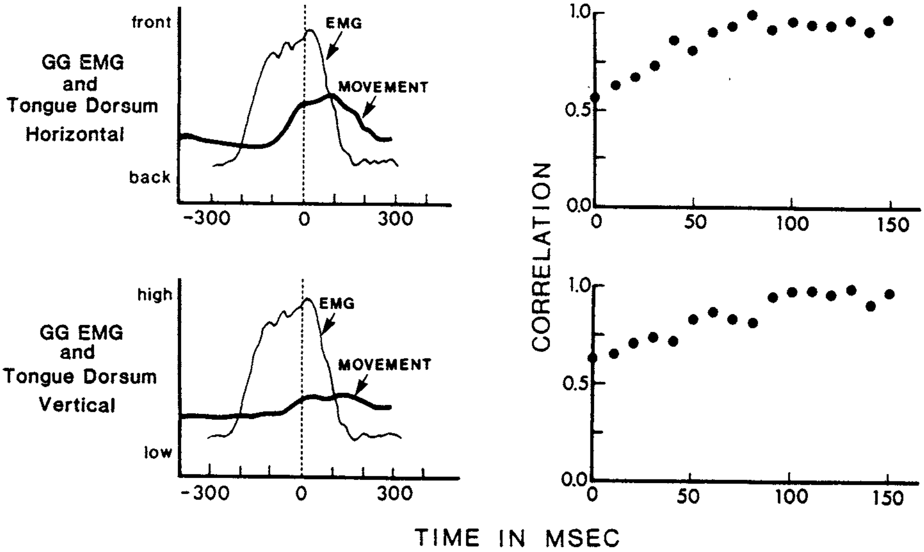


Fig. 5. Genioglossus EMG activity with tongue dorsum horizontal movement (top left) and with tongue dorsum vertical movement (bottom left) during /i/. Correlation functions between EMG and respective movement curves are shown on the right.

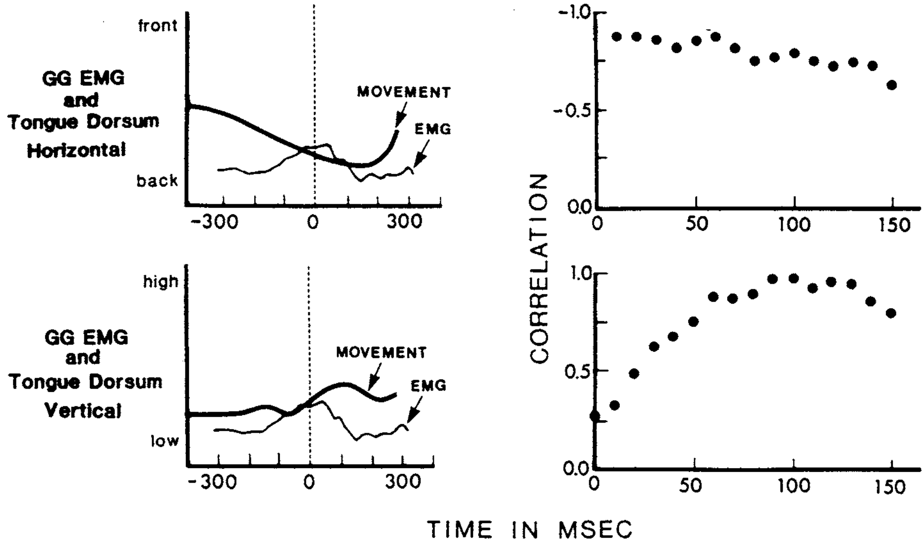


Fig. 6. Genioglossus EMG activity with tongue dorsum horizontal movement (top left) and with tongue dorsum vertical movement (bottom left) during /u/. Correlation functions between EMG and respective movement curves are shown on the right.

between genioglossus EMG activity and tongue movements. The EMG data shown in these figures were taken from the EMG-only run. The line-up point, zero time on the abscissa, represents the voice onset of the vocalic segment. Correlation functions between genioglossus EMG activity and tongue movement are shown on the right. Correlation functions were calculated by shifting the movement waveform with respect to the EMG. For each value of the time-shift variable, a Pearson Product-Moment correlation coefficient was calculated over the interval from the onset to the major peak of the EMG curve. This procedure was adopted because it restricts the range of the calculations to the rising portion of the EMG curve. Figure 5 shows that correlation functions between patterns of genioglossus EMG activity with tongue horizontal and tongue vertical movements for the vowel /i/ nearly reach unity at latencies of about 110 msec. This latency seems to be a reasonable value for the mechanical response time of this muscle-articulator system. Figure 6 demonstrates the relationship between genioglossus EMG activity with tongue horizontal and vertical movements for the vowel /u/. There is less EMG activity for the vowel /u/ than for /i/. Tongue vertical movements for /u/ and /i/ begin at about the same time, but horizontal tongue movements for /u/ begin much earlier. This observation is supported by a comparison of the correlation functions between /i/ and /u/. The correlation functions for vertical and horizontal movements for /i/ and vertical movements for /u/ all appear roughly similar, showing a peak in the vicinity of 100 msec, while the correlation function for horizontal movements for /u/ has its peak at or before 0 msec and has the opposite sign. These results show that the posterior part of the genioglossus muscle contributes to fronting and bunching movements for these vowels, but not to the backing movements for /u/.

Peak genioglossus EMG activity for the vowels /e/ and /o/ was shown in Figure 4 to be similar to that for /i/ and /u/ respectively. While it is true that /e/ and /o/ are much more diphthongized than /i/ and /u/, the relationship between horizontal and vertical movements for /e/ should be similar to that for /i/ since they are both front vowels, and the relationship between movement components for /o/ should resemble that for /u/ since they are both back vowels. Figures 7 and 8 show that this is the case. Figure 7 shows that vertical and horizontal movements for /e/ and /i/ begin at the same time, about 100 msec before the voice onset for the vowel. Figure 8 demonstrates that vertical movements for /u/ and /o/ begin at about the same time as horizontal and vertical movements for /i/ and /e/, but horizontal movements for both /u/ and /o/ begin much earlier. Thus, data from the second set of front-back vowels, /e/ and /o/, support the data from the first set, /i/ and /u/. Taken together, the results show that the posterior part of the genioglossus muscle participates in both fronting and bunching movements for front vowels and in bunching but not backing movements for back vowels. Thus, fronting and bunching for front vowels appear to be time-locked, but backing movements can occur independently of bunching movements for back vowels.[3]

---

[3] *We do not imply that the posterior part of the genioglossus muscle is solely responsible for fronting movements. However, as we discuss later, fronting movements seem to require coordinated activity of this and other muscles, including the anterior part of the genioglossus.*

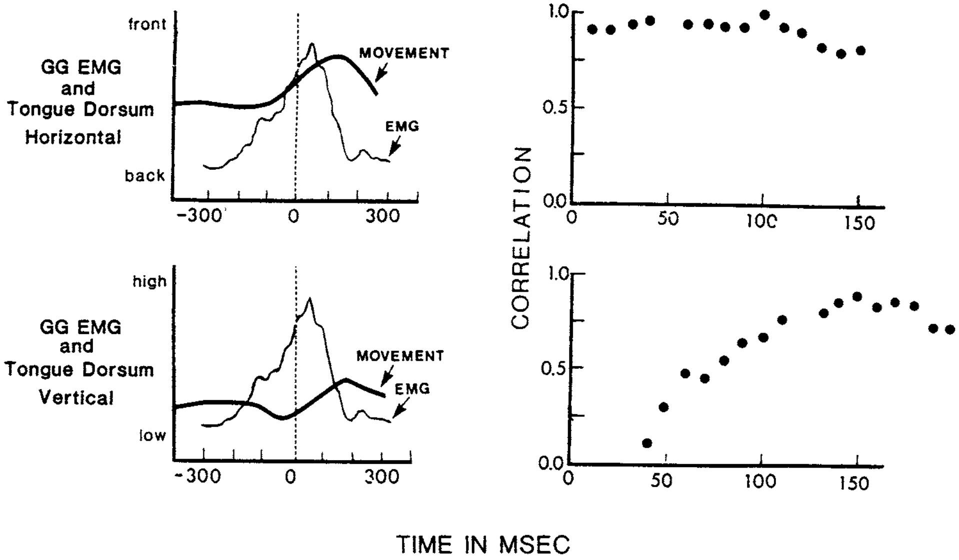


Fig. 7. Genioglossus EMG activity with tongue dorsum horizontal movement (top left) and with tongue dorsum vertical movement (bottom left) during /e/. Correlation functions between EMG and respective movement curves are shown on the right.

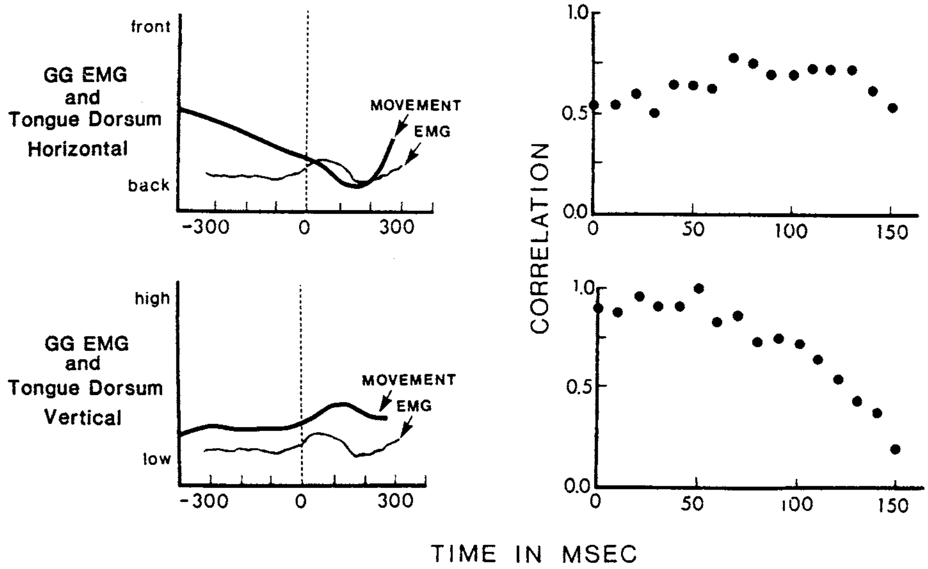


Fig. 8. Genioglossus EMG activity with tongue dorsum horizontal movement (top left) and with tongue dorsum vertical movement (bottom left) during /o/. Correlation functions between EMG and respective movement curves are shown on the right.

While the relationships between posterior genioglossus EMG activity and tongue movements for the /e/ and /o/ vowel pair resemble those for the /i/ and /u/ pair, the correlation functions between the vowel pairs appear to be quite different. For /i/ and /u/, peak correlation values occurred at latencies of about 100 msec with the exception of horizontal movements for /u/, where the peak (negative) correlation value occurred at 0 msec. The correlations for /e/ and /o/ show the same general pattern, with respect to sign, but the functions do not peak uniformly near 100 msec. This is partly an artifact of the method. The correlation function falls off above 100 msec as expected but remains high at shorter delays, because both curves are monotonically increasing in the interval over which the correlations are calculated. In addition, tongue movements for the diphthongized vowels /e/ and /o/ may be more affected by forces other than those exerted by the genioglossus muscle than tongue movements for /i/ and /u/. In particular, the vertical movements for /e/ during the closure period (the interval just before the line-up point) are apparently not due to the genioglossus, but this interval has a large effect on the correlation function between movement and genioglossus EMG at short delays. It is also possible that the mechanical response characteristics of the muscle differ in different vowel environments. Thus, we cannot determine without additional data whether the falloff above 50 msec in the correlation function between the genioglossus and vertical movements for /o/ is due to the effects of other forces, or whether it represents a faster genioglossus response time in this phonetic environment.

Similar patterns of genioglossus activity were reported by Raphael and Bell-Berti (1975) for the same talker producing six of these vowels in a similar frame. The Raphael and Bell-Berti study, in addition, reports data from other lingual muscles. Their data, as well as our own, demonstrate that the onset of genioglossus activity never preceded the onset of voicing for any vowel by more than 250 msec. For back vowels, however, styloglossus muscle activity begins at least 500 msec before the onset of voicing. This muscle is thought to participate in tongue backing. Thus, EMG data suggest a timing difference for backing and fronting maneuvers for this subject.

With these comments in mind, we turn our attention to inter-articulator timing measurements. Figure 9 shows sagittal plane trajectories for the tongue-dorsum pellet for four vowels. The number of vowels shown has been limited and the trajectories have been highly smoothed to simplify the figure. Data from the remaining six vowels are qualitatively similar to those shown in Figure 9. The time interval for these plots begins at the voice onset of the schwa and ends at lip contact for the final consonant. Lines have been superimposed on the trajectories in Figure 9 to indicate three different time intervals. The trajectories during the production of the schwa are enclosed by the inner line. The trajectories during the production of the bilabial closure are enclosed by the outer line. With the exception of /a/, trajectories after the consonant release appear outside the region enclosed by the lines.

Considering tongue positioning during the schwa, we note that the region is long and flat. Anticipatory movement for the back vowel /u/ occurs primarily in the horizontal direction but very little in the vertical direction. The front vowels cluster near the left end of this region, and demonstrate only small movements before the period of consonantal closure. Within the /p/ closure region, the trajectories continue to spread

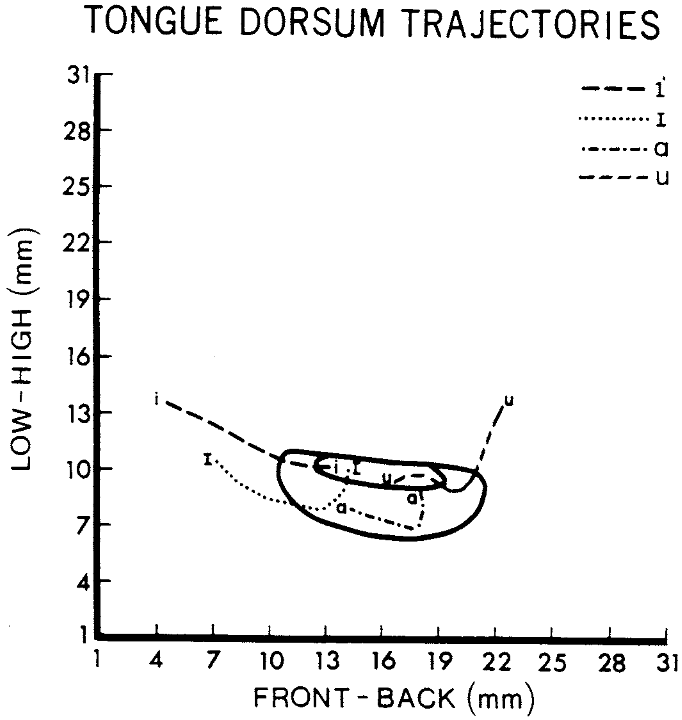


Fig. 9. Smoothed movement trajectories of the tongue dorsum pellet during the interval beginning with voice onset of the schwa, including the initial consonant and the vowel, and ending with the lip contact for the final consonant. Trajectories during the production of the schwa are enclosed by the inner black line, during the production of the initial bilabial closure are enclosed by the outer black line, and during the interval from the release of the initial consonant to the lip closure for the final consonant appear outside the black lines.

horizontally and also lower. Lowering movements during bilabial stops have been noted previously (Houde, 1967). It is unclear whether this movement is produced by active or passive forces. In either case, there is a movement apparently related to the consonant. Finally, the trajectories move upward and out toward the extremes of the space.

The next two figures show the time course of tongue-dorsum movements for all ten vowels. First, we consider the vertical dimension, shown in Figure 10. In this plot, the line-up point – zero time – was the onset of voicing for the vowel. Implosion for the consonant occurred at different times depending on vowel type, and ranges from about 120 to 160 msec. Vertical tongue-position curves for all 10 vowels begin to diverge from each other at this point. Therefore, the onset of vertical vowel-related movements appears to be time-locked to the consonant.

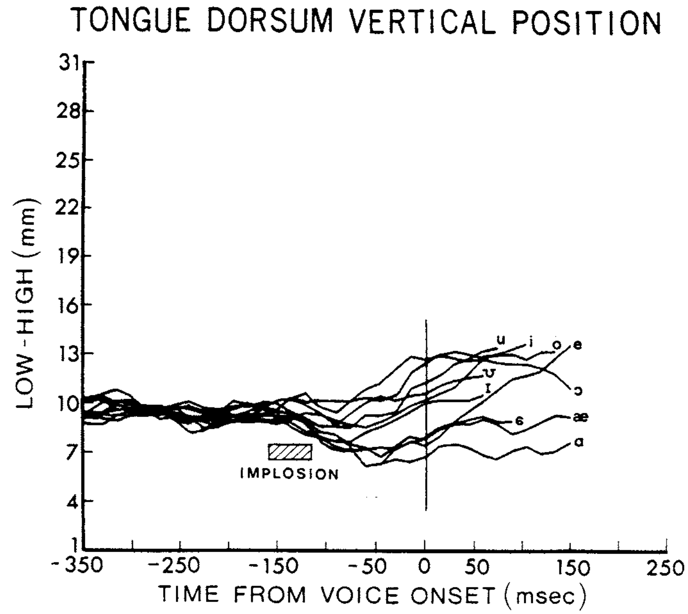


Fig. 10. Tongue dorsum vertical movements. Zero time represents the onset of voicing for the vowel. Implosion of the initial consonant ranged from –120 to –160 msec depending on vowel type, and is shown by the rectangle.

Horizontal movements shown in Figure 11 are different. These curves are separate even at the earliest time measured, 350 msec before voice onset for the vowel. More significantly, the curves for back vowels and high front vowels begin to diverge from each other almost immediately. Notice that while backing movements for the back vowels begin much earlier than their vertical movements, the fronting movements for front vowels begin only at about the same time as their vertical movements – that is, at about the moments of implosion.

We can perhaps explain the difference between fronting and backing on physiological grounds. At least for the high front vowels, a single muscle – namely the genioglossus – may be primarily responsible for moving the tongue both forward and upward. On the other hand, tongue backing is achieved by muscles other than the genioglossus – for example, the styloglossus. Thus, backing movements could occur independently of vertical movements in high back vowels.

Why vertical and horizontal movements should be controlled independently, however, cannot be determined from the above data alone. Several explanations are possible. On physiological grounds, it may be that backing movements must begin earlier because they are intrinsically slower than raising and fronting movements. On perceptual grounds, anticipatory vertical and horizontal movements may be necessary in that they spread

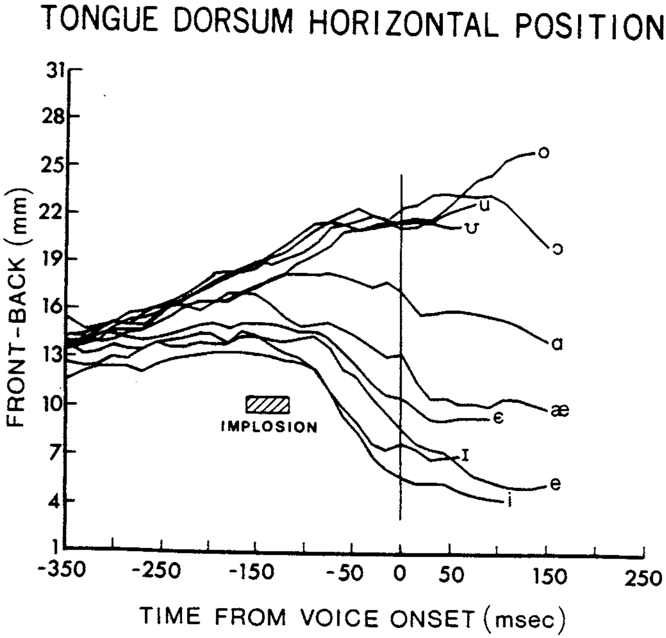


Fig. 11. Tongue dorsum horizontal movements. Zero time represents the onset of voicing for the vowel. Implosion of the initial consonant ranged from –120 to –160 msec depending on vowel type, and is shown by the rectangle.

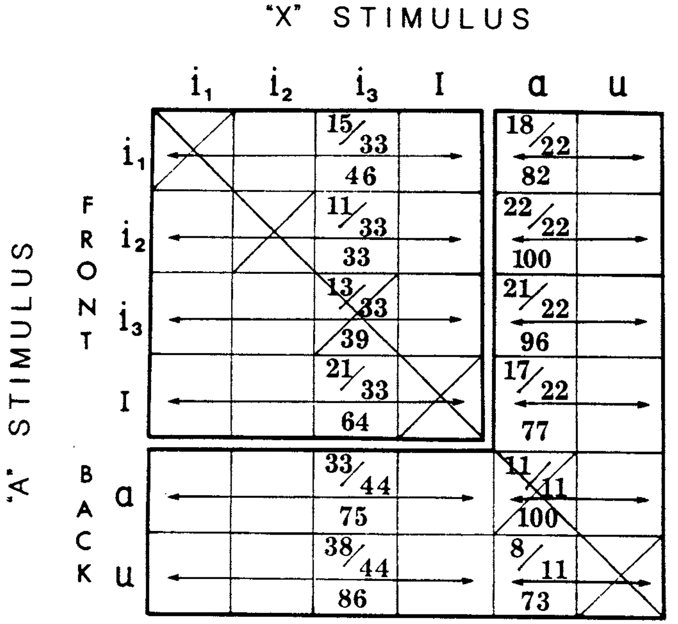
phonetic information across neighboring segments. However, in this context, there may be physiological constraints that restrict anticipatory vertical tongue movements. Other explanations might rest on acoustic/aerodynamic grounds. In any case, tongue-dorsum movement data for this speaker suggest that front-back information about the vowel is available before high-low information.

To test the notion that the anticipatory horizontal tongue movements during the production of the schwa were perceptually significant, AX discrimination and phoneme labeling tests were conducted. Specifically, we wanted to know if listeners could discriminate between schwas produced with front versus back tongue positions.

PERCEPTUAL ANALYSIS

Methods

Schwa segments from three tokens of /əpip/, and from a single token of /əpɪp/, /əpup/, and /əpap/ produced during the EMG-only run were excised by computer waveform editing. Each of the six stimuli consisted of about three pitch periods and was about 25 msec in duration. AX discrimination and labeling tests were prepared using



Comparison	Correct/Total	% Correct	X^2 Value
Front vs Front	60/132	46%	3.9
Back vs Back	19/22	86%	6.64 **
Between Groups	149/176	85%	44.9 **
Total Correct	520/660	79%	146.4 **

** $p < 0.01$

Fig. 12. Results of AX discrimination testing. The ordinate represents the A stimulus and the abscissa represents the X stimulus of all possible AX pairs. Data are collapsed across a front group consisting of three "/i/ schwas" and one "/ɪ/ schwa," and across a back group consisting of a single "/a/ and /u/ schwa." Pooled data for all combinations of front and back pairs are shown below.

these six stimuli and presented to 12 subjects. The subjects had not previously taken part in speech perception experiments.

Results

Discrimination. The results of the discrimination test are shown in Figure 12. The ordinate represents the A stimulus and the abscissa represents the X stimulus of all possible AX discrimination pairs. The data are collapsed across the front group, which

RESPONSE

STIMULUS	i	ɪ	u	ɑ
i 240	72 30	71 30	57 24	40 17
ɪ 240	54 23	93 39	33 14	59 25
u 240	43 18	27 11	126 53	48 20
ɑ 240	26 11	63 26	35 15	113 47

Correct/Total	% Correct	X^2 Value
404/960	42%	139.8 ***

***p<0.001

Fig. 13. Results of the labeling tests. Each row represents the distribution of the responses for 240 presentations of a stimulus. In each cell, the upper left score represents the frequency of that response, and the bottom right score represents percent occurrence.

consisted of the three schwas taken from three different productions of /əpip/ (hereafter referred to as the /i/ schwas) and one schwa taken from /əpɪp/ (hereafter the /ɪ/ schwa), and a back group that consisted of one schwa taken from single productions of /əpap/ and /əpup/ (the /a/ and /u/ schwas, respectively). For each row, the upper score represents the ratio of correct responses to total presentations and the lower score represents percentage correct.. For example, the first row shows that when the first token of one of the three /i/ schwas, i_1, was paired with front group schwas, i_2, i_3, and /ɪ/ schwas, discrimination performance was at chance level, 46% correct. However, when the i_1 schwa was paired with back group schwas (the /a/ and /u/ schwas), discrimination performance improved to 82% correct. The summary data shown at the bottom of the figure demonstrate that subjects were successfully able to discriminate between schwas produced with front versus back tongue positions. Note that "between group" performance, which included all combinations of front versus back schwas, was significantly greater than chance level at 85% correct ($\chi^2 = 44.9, p < 0.001$). The pooled front versus front group discrimination score of 46% correct shows that subjects were not able to discriminate between AX pairs consisting of schwas produced only with front tongue positions ($\chi^2 = 3.9, p > 0.1$). However, it appears that discrimination was better for **ɪ–i** pairs than any other combination of front schwas. That is, the fourth row demonstrates that **ɪ–i**$_{1,2,3}$ pairs yielded 64% correct performance, while the first three rows show that any /i/ schwa paired with the remaining two /i/ schwas and the /ɪ/ schwa produced 39% correct performance. In addition, back versus back discrimination performance was above chance level at 86% correct ($\chi^2 = 6.6, p < 0.01$). Finally, note that overall performance, shown as "total correct" which includes "same" (for example $\mathbf{i}_1$ versus $\mathbf{i}_1$) and "different" AX pairs, was significantly greater than chance level performance at 79% correct ($\chi^2 = 146.4, p < 0.0001$). These data led us to conclude that listeners were able to discriminate between the front and back group schwas produced by the same speaker. However, discrimination was probably aided by the acoustic consequences of articulatory movements other than tongue fronting and backing, since discrimination performance between /ɪ/ and /i/, between the back group schwas, and overall discrimination performance, was very high.

Labeling. On the basis of the results of the discrimination test, we decided to test further the perceptual significance of the anticipatory horizontal movement and perhaps other differentiating articulatory gestures occurring during the production of the schwa by asking our subjects to label the stimuli as /i/, /ɪ/, /u/, or /a/. The same stimuli used in the discrimination test were used in the labeling tests, except that only one /i/ schwa was used. The labeling test included five repetitions of each of the four stimuli. There were no training presentations, and subjects were not given feedback during the experiment. The same 12 subjects who participated in the discrimination test participated in the labeling test. The results are shown in Figure 13. Here, each row represents the distribution of responses for 240 presentations of a stimulus. In each cell, the upper left score represents the frequency of that response, and the bottom right score represents percentage of occurrence. Scores on the main diagonal show that /i, ɪ, u, a/ schwas were labeled as such 72, 93, 126, and 113 times, respectively. Overall performance was significantly above chance at 42% correct ($\chi^2 = 139.8, p < 0.001$). Even though the schwa stimuli are only about 25 msec long, and represent reduced vocal tract shapes as plotted in both the movement and acoustic space, they appear to have a distinguishable vowel-like quality that results in the surprisingly accurate labeling.

Finally, notice that the subjects appeared to have more difficulty labeling the front schwas than the back. Although it is quite probable that other vocal tract parameters, such as lip rounding for the back vowels, contributed to the increased accuracy in which schwas were labeled, it seems that the anticipatory backing gesture observed in the movement data during schwa production preceding back vowels is perceptually significant and contributes to the increased accuracy with which the back schwas were labeled.

To determine the extent to which the schwas produced in different vowel environments were spectrally distinct from each other and spectrally similar to the vocalic portions of the syllables in which they occurred, an F_1 versus F_2 schwa space, similar

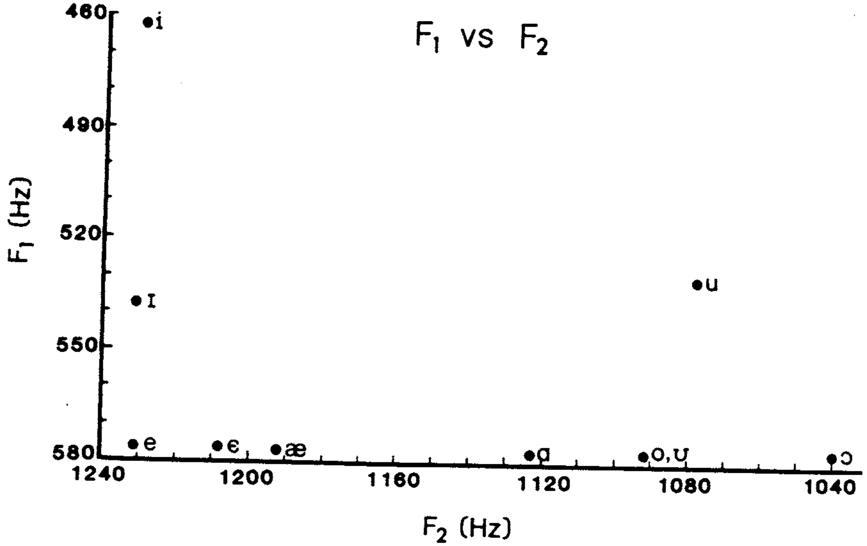


Fig. 14. First and second formant values in Hz for the schwas produced in each of the 10 vowel environments used in this study. For example, the data point labeled /i/ represents F_1 and F_2 values for the schwa segment of a single production of the syllable /əpip/. See text for description of analysis techniques.

to the vowel space shown in Figure 2, was plotted. For each of the 10 vowel environments, the schwa segment of one of the 20 tokens of the syllables was excised by computer waveform editing. The criterion for the selection of a schwa was the appearance of two clearly defined pitch pulses. Since the durations of the schwa segments were not long enough to allow for spectral analysis, it was necessary to create 200 msec signals that consisted of multiple repetitions of the original stimulus. Spectral information from the 200 msec versions of the schwas was obtained using standard procedures at Haskins Laboratories. The results are shown in Figure 14. Note that the schwas occurring in the vowel environments used in the discrimination and labeling tests, the /i, ɪ, u, a/ schwas, are spectrally distinct. Also note that the schwa acoustic space is somewhat similar to the vowel acoustic space shown in Figure 2. Front and back schwas map into distinct regions, and some vowel-height information also appears. Figure 14 clearly demonstrates that the second-formant frequency of the schwa segment of the syllable provides anticipatory front-back information about the following vowel. To a lesser degree, the first-formant frequency of the schwa makes anticipatory high-low and lip spread-round information about the vowel available to the listener.

Thus, EMG, movement, and acoustic data support the perceptual effect noted in this experiment: that information about the vowel is available at the onset of the syllable and apparently can be utilized by the listener.

DISCUSSION

Our results show that the placement of the 10 vowels in the acoustic and movement domains was strikingly similar. Information from the tongue dorsum pellet alone corresponded surprisingly well with the traditional F_1-F_2 plot, suggesting that the acoustic effects of such other articulations as jaw opening and lip rounding varied directly with tongue dorsum position for this speaker. In the EMG domain, data were reported for one tongue muscle, the posterior genioglossus, which is associated with fronting and bunching in the movement domain. Considering these limitations, the EMG data corresponded well with those aspects of the movement and acoustic measurements associated with fronting and bunching.

The vowel space for this speaker, as plotted in the three measurement domains, appears different in some respects from other vowel spaces plotted in the acoustic domain (e.g., Peterson and Barney, 1952) and the movement domain (e.g., Nearey, 1977). The major difference between this and other studies is associated with this speaker's production of the vowel pairs /i–ɪ/, /e–ɛ/, and /u–ʊ/. In a previous study based on electromyographic and acoustic data (Raphael and Bell-Berti, 1975; Bell-Berti *et al.*, 1979), it was concluded that speakers differentiate into two classes with respect to their production of the so-called "tense-lax" vowels. It was found that for some speakers, including the subject in this experiment, tongue height for the front vowels decreased in the order /i, e, ɪ, ɛ/, and for other speakers they decreased in the order /i, ɪ, e, ɛ/. Our acoustic and EMG data agree with those previously reported, and our movement data furthermore support the validity of inferring tongue height from EMG and acoustic data for this speaker.

The second purpose of this study was to further examine the role of the genioglossus muscle in tongue horizontal and vertical movements. Correlation analysis of our EMG and movement data shows that this muscle contributes to tongue fronting for the vowel /i/ and bunching for both /i/ and /u/. The anatomy of the genioglossus muscle is shown schematically in Figure 1. Its fibers arise from the symphasis menti of the mandible. The posterior fibers radiate toward the tongue root and toward the posterior parts of the tongue dorsum at the midline. The most anterior fibers course upward and forward, also near the midline. Miyawaki *et al.* (1975) and Smith (1971) have shown that posterior fibers contribute to tongue bunching for the front and back high vowels /i/ and /u/. This finding is supported by modeling studies (Perkell, 1974, Kiritani *et al.*, 1975). Anterior fibers have been shown to be active during front but not back vowel production (Miyawaki *et al.*, 1975). The path of insertion used in the present study, illustrated in Figure 1, was near the origin of the muscle. In this region, a shallow insertion will sample relatively posterior fibers, while a deeper insertion will record from more anterior fibers. Basing ourselves on the Miyawaki *et al.* data, we confirm that the genioglossus insertion in the present study was into relatively posterior fibers of this muscle. We conclude that the anterior and posterior parts of the genioglossus muscle act together in raising and fronting for high front vowels such as /i/, while the posterior part of the genioglossus and other extrinsic tongue muscles, most probably the styloglossus, act together to raise and back the tongue for /u/. For the high front vowel, the genioglossus is thus involved in both the vertical and horizontal components of movement. For the high back vowel, the genioglossus contributes mostly to raising while the styloglossus accounts for the backing movements. Thus, there is independent control of vertical and horizontal movements for high back vowels but not for high front vowels.

The third purpose of this study was to investigate the timing of tongue versus lip and jaw movements for the different vowel environments. Analysis of tongue dorsum movements for the 10 vowels shows that vertical movements begin to differentiate at about the time of lip contact (Fig. 10). Horizontal movements, on the other hand, begin to differentiate well before the onset of the schwa (Fig. 11), since movements for back vowels begin earlier than those for front vowels. Since jaw movements were negligible, the vertical and horizontal components of the trajectories can be attributed to tongue movements alone. This result suggests that there are different constraints on the timing of vertical and horizontal tongue movements for these syllables and that studies of coarticulation should take this difference into account.

Lastly, we wanted to determine whether anticipatory tongue movements were observable in the appropriate acoustic parameters, and if so, to determine whether the resultant acoustic information would lead to perceptual judgments about the upcoming vowel. Consequently, it was necessary to conduct perceptual analyses of the anticipatory tongue movements observed during the schwa portions of the syllables. On the basis of the physiological data, we expected to find differential perception of schwas preceding front and back vowels. Perceptual results indicated that the front and back group schwas were indeed discriminable and furthermore carried front-back labeling information about the upcoming vowel. Unexpectedly, both discrimination and labeling results showed that the schwa carried high-low information as well. Acoustic analysis of the schwa segments revealed the basis for the perceptual results: Relative F_1–F_2 values of schwas preceding different vowels were similar to those of the vowels themselves. Our EMG and movement data show no basis for this latter result, suggesting that the perception of anticipatory fronting and raising must be due to movements of articulators other than the tongue dorsum, most likely the lip or larynx.

In summary, this multiple-level study of vowel production leads to the following conclusions: (1) Intra-articulator timing measurements, taken on simultaneous genioglossus EMG and tongue dorsum movement data, show that the genioglossus is responsible for both raising and fronting. When raising and fronting occur together, they are caused primarily by the same muscle and are therefore time-locked. (2) Inter-articulator timing measurements, taken on simultaneous lip and tongue movements, show that vertical tongue movements have different time constraints than horizontal movements when different muscle groups are responsible for the two components of movement. Thus, backing has a large temporal extent, encompassing the first two segments of the syllables used in this study, while fronting and raising have a shorter temporal extent that is usually less than the duration of one segment. (3) For the vowel segment, comparisons of tongue dorsum movement data with acoustic data show agreement between tongue horizontal position and F_2 values and between tongue vertical position and F_1 values. However, the same comparisons for the schwa segments do not totally agree. We find the expected relationship between tongue dorsum horizontal position and schwa F_2 values, but not between tongue dorsum vertical position and schwa F_1 values. That is, the acoustic data suggest vertical differences among schwas that are not found in the tongue dorsum vertical movement data. (4) The results of perceptual tests show that the schwa F_1 values were just as significant in signaling tongue height information about the upcoming vowel as were the schwa F_2 values in signaling tongue front-back vowel information. (5) The data related to the schwa segment, when viewed on all four measurement levels, suggest that articulators other than the tongue can supply information about the vowel that is normally associated with vowel height. This implies that the speech production system is organized to transmit anticipatory information about vowel identity in a number of ways. For instance, when a motor constraint restricts movement of a primary articulator, for example anticipatory tongue vertical movement in a high-front vowel environment, secondary articulators are enlisted to transmit vowel-height information that is usually associated with tongue height. This result lends support to the notion that the motor system is organized to transmit a phonetic code in which information about a segment needs to overlap with information about its neighbors. This need is so strong that the same phonetic information can be carried by different articulators according to the phonetic environment.

REFERENCES

BAER, T., BELL-BERTI, F. and TULLER, B. (1979). On determining EMG onset time. In J.J. Wolf and D.H. Klatt (eds.), *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America* (New York), pp. 75-78.

BELL-BERTI, F., RAPHAEL, L.J., PISONI, D.B. and SAWUSCH, J.R. (1979). Some relationships between speech production and perception. *Phonetica*, **36**, 373-383.

BORDEN, G.B. and GAY, T.S. (1978). On the production of low tongue tip /s/: A case report. *Journal of Communication Disorders*, **11**, 425-431.

GAY, T.J., USHIJIMA, T., HIROSE, H. and COOPER, F.S. (1974). Effect of speaking rate on labial consonant-vowel articulation. *Journal of Phonetics*, **2**, 47-63.

HARRIS, K.S. (1981). Electromyography as a technique in laryngeal research. *Haskins Laboratories Status Report on Speech Research*, **SR-66**, 1-33.

HONDA, K. (1981). Relationship between pitch control and vowel articulation. Paper presented at the Conference on Vocal Fold Physiology, Madison, Wisconsin.

HONDA, K. and BAER, T. (1981). Interaction between pitch control and vowel articulation. *Journal of the Acoustical Society of America*, **70**, S12 (abstract).

HOUDE, R.A. (1967). A study of tongue body motion during selected speech sounds. Ph.D. thesis, University of Michigan.

KEWLEY-PORT, D. (1977). EMG processing for speech research. *Haskins Laboratories Status Report on Speech Research*, **SR-50**, 123-146.

KIRITANI, S., MIYAWAKI, K., FUJIMURA, O. and MILLER, J.E. (1975). Computational model of the tongue. *Journal of the Acoustical Society of America*, **57**, S3 (abstract).

KIRITANI, S., SEKIMOTO, S., IMAGAWA, H., ITOH, K., USHIJIMA, T. and HIROSE, H. (1976). Simultaneous recording of EMG with pellet tracking by use of X-ray microbeam. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo*, **10**, 13-14.

LIBERMAN, A.M., COOPER, F.S. SHANKWEILER, D.P. and STUDDERT-KENNEDY, M. (1967). The perception of the speech code. *Psychological Review*, **74**, 431-461.

LIBERMAN, A.M. and STUDDERT-KENNEDY, M. (1978). Phonetic perception. In R. Held, H. Leibowitz and H.L. Tueber (eds.), *Handbook of Sensory Physiology* (Vol. VIII) (Heidelberg).

MIYAWAKI, K., HIROSE, H., USHIJIMA, T. and SAWASHIMA, M. (1975). A preliminary report on the electromyographic study of the activity of the lingual muscles. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, **9**, 91-106.

NEAREY, T.M. (1977). Phonetic feature systems for vowels. Ph.D. thesis, University of Connecticut (reproduced by the Indiana University Linguistics Club), 1978.

PERKELL, J.S. (1974). A physiologically-oriented model of tongue activity in speech production. Ph.D. thesis, Massachusetts Institute of Technology.

PETERSON, G. and BARNEY, H. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America*, **24**, 175-184.

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

SMITH, T.S. (1971). A phonetic study of extrinsic tongue muscles. *UCLA Working Papers in Phonetics*, **18**.

WOOD, S. (1979). A radiographic analysis of constriction location for vowels. *Journal of Phonetics*, **7**, 25-43.