AN ELECTROMYOGRAPHIC-CINEFLUOROGRAPHIC-ACOUSTIC STUDY OF DYNAMIC VOWEL PRODUCTION*

Peter J. Alfonso+ and Thomas Baer

There are many studies in the phonetics 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, & Cooper, 1974; Borden & Gay, 1978; and Kiritani, Sekimoto, Imagawa, Itoh, Ushijima, & 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 difficulties of the methodology, since the information gained from simultaneous monitoring of the different levels of speech production, 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. For instance, Wood (1979) has pointed out that the controversy that still exists over the more appropriate level of vowel description, acoustic or articulatory, is related to the inconsistencies among different X-ray studies, and to the poor agreement between these studies and other acoustic studies. This seems to be the source of a recurring problem; often EMG, movement, and acoustic data collected from different experiments that usually use different talkers are used to make comparisons and assumptions about each measurement level. Certainly, the testing and formulation of models of vowel articulation would seem to depend upon a complete description provided only by simultaneous measures.

Other instances where simultaneous analysis of the three levels is more useful than any combination of the two are related to dynamic measurements of vowel production. That is, simultaneous measures not only allow for inter-articulator timing measurements, such as tongue and jaw relationships, but

* A version of this paper was presented at the annual convention of the American Speech and Hearing Association, November 21-24, 1980, Detroit, Michigan.
+ Also University of Connecticut, Storrs, Connecticut.
Acknowledgment. This research was supported by NIH Grant NS13617 to Haskins Laboratories.
also allow for intra-articulator timing measurements, for example, genio- 
glossus muscle activity and tongue fronting. Furthermore, high correlations 
between patterns of EMG activity and movement lend support to the notion that 
the relationship between EMG activity and movement of the muscle-articulator 
system under study is causal.

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. A single speaker of American English produced isolated 
syllables of the form /apVp/, using ten different vowels. We will consider 
here only the dynamics associated with tongue movements for these syllables. 
More specifically, we will show that the timing of vertical tongue movements 
for both front and back vowels was time-locked to some component of the 
inital consonant, while the timing of horizontal movements began much earlier 
for back vowels than for front vowels. For back vowels, horizontal tongue 
movement began before voice onset for the schwa, whereas for front vowels 
horizontal tongue movement began at about the same time as their vertical 
movements. In addition, we will show that the differentiation in horizontal 
tongue movements during schwa production was perceptually significant.

PROCEDURE

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. In addition, a gold 
chain was laid on the floor of the nasal tract for monitoring velar movements. 
However, we will consider here only movements of the tongue dorsum.

EMG signals were recorded from the orbicularis oris muscle and from two 
muscles of the tongue, the genioglossus and superior longitudinal. The paths 
of insertion of the hooked wire electrodes for these muscles are also 
indicated in Figure 1. Good quality acoustic recordings were made by using 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. 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. The subject's utterances from the 
experiment 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.
Figure 1. Schematic representation of lead pellets attached to the tongue tip, blade, and dorsum, and to the upper and lower incisors. Also shown is a gold chain laid on the floor of the nasal tract for monitoring velar movements. The arrows indicate the paths of insertion of the hooked wire electrodes for the genioglossus and superior longitudinal muscles.
RESULTS

The next three figures demonstrate the good agreement among the three types of measures made in this study.

Figure 2 shows results of acoustic measurements on vowels produced during the X-ray run. 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 vowels /i/ and /e/ higher and more forward than /I/ and /ɛ/.

Figure 3 shows the 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). Movements along all of these trajectories, except the one for /I/, are in an ascending direction and away from the center. The pattern of locations of these trajectories grossly resembles the vowel pattern in the acoustic domain just shown.

Figure 4 shows the pattern of peak EMG activity for the genioglossus muscle for each of the ten vowels. Greatest activity is noted for /i/ and /e/ and somewhat less for /u/ and /o/. These vowels, traditionally termed tense, are also observed to be highest in the acoustic and articulatory domains. Among the remaining vowels, there is somewhat more activity for the front than for the back.

Next we turn our attention to articulator timing measurements. Simultaneous monitoring of different levels of speech production, namely muscle activity, articulator movement, and acoustic, allow for both intra- and inter-articulator timing measurements. As an example of intra-articulator measures, Figure 5 demonstrates the relationship between genioglossus EMG activity and tongue movements. This figure 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. High correlations of this type, genioglossus EMG with tongue fronting and bunching movements in this example, lend support to the notion that the relationship between EMG activity and movement of the muscle-articulator system under study was causal.

Similar patterns of genioglossus activity were reported by Raphael and Bell-Berti (1974) for the same talker producing six of these vowels in a similar frame. The Raphael and Bell-Berti study, in addition, reports data from additional lingual muscles. Their data, as well as our own, demonstrate that the onset of genioglossus activity never preceded the onset of voicing for the 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, the EMG data suggest a timing difference for backing and fronting maneuvers.

With these comments in mind, we turn our attention to interarticulator timing measurements. Figure 6 shows sagittal plane trajectories for the tongue dorsum pellet for four vowels. The time interval for these plots begins at the voice onset of the schwa and ends at lip contact for the final consonant. The number of vowels has been limited here to simplify the figure.
Figure 2. Peak center frequency values in Hz for the ten vowels used in this study. Each data point represents the average of the two tokens produced during the X-ray run.
Figure 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 /o/, movements along the trajectories are in an ascending direction and away from the center. Each trajectory represents the average movement of two tokens.
Figure 4. Peak genioglossus EMG activity for each of the ten vowels. Each data point represents the average of two tokens produced during the X-ray run.
Figure 5. Genioglossus EMG activity with tongue dorsum horizontal movement (top left) and with tongue dorsum vertical movement (bottom left) during /i/. Correlation functions between the EMG curve and the respective movement curves are shown on the right.
Figure 6. Movement trajectories of the tongue dorsum pellet during the interval beginning with the 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.
Lines have been superimposed on the trajectories in Figure 6 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 can see that the region is long and flat; that is, anticipatory movements for the vowel occur primarily in the horizontal direction but very little in the vertical direction. Moving into the /p/ closure region, the trajectories continue to spread horizontally and also lower. Lowering movements during bilabial stops have been noted previously (Houde, 1967). It is unclear whether this movement is active or passive. In either case, there is a movement apparently related to the consonant that makes it difficult to determine the onset of vowel-related movements. Finally, the trajectories, moving upward and out toward the extremes of the space, demonstrate vowel-related movements.

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 7. In this plot, the lineup 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 is the same for all vowels during the interval preceding implosion. The curves begin to diverge from each other at this point. Therefore, the onset of vertical vowel-related movements appears to be time-locked to some component of the consonant, so that they appear in these utterances at about the time of implosion.

Horizontal movements shown in Figure 8 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 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 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 from vertical movements in high back vowels.

Why they should be controlled independently, however, cannot be determined from the above data alone. Several explanations are possible. It may be that backing movements are intrinsically slower than raising and fronting movements and therefore must begin earlier. Other explanations might rest on acoustic or aerodynamic grounds. However, the results show, for this speaker, that front-back information about the vowel is available before high-low information, and that the information is available at the beginning of the syllable.
Figure 7. 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.
TONGUE DORSUM HORIZONTAL POSITION

Figure 8. 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.
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. Schwa segments from three productions of /æp/p/, and from a single production of /æp/p/, /æp/p/, and /æp/p/ were excised by computer. Each of the six stimuli was about 25 msec in duration and consisted of about three pitch periods. Using the Haskins Pulse Code Modulation system, the six stimuli were digitized, and AX discrimination and labeling tests were prepared and presented to 12 subjects. The results of the discrimination test are shown in Figure 9. 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 consisted of the three schwas taken from three different productions of /æp/p/ (hereafter referred to as the /i/ schwas) and one schwa taken from /æp/p/ (hereafter the /I/ schwa), and a back group that consisted of one schwa each taken from a single production of /æp/p/ and /æp/p/ (the /a/ and /u/ schwas, respectively). For instance, the first row shows that when the first token of one of the three /i/ schwas, il, was paired with front group schwas, i2, i3, and I schwas, discrimination performance was at chance level, 46 percent correct. However, when the il schwa was paired with back group schwas (the /a/ and /u/ schwas), discrimination performance improved to 82 percent correct. The summary data shown at the bottom of the figure demonstrate that discrimination performance across all front-back AX pairs was well above chance at 85 percent correct, whereas discrimination performance across front-front pairs was at a chance level of 46 percent correct. However, also note that discrimination performance across back-back pairs was also well above chance at 86 percent correct. Finally, note that overall discrimination performance, which included same as well as different AX pairs, was at 79 percent correct. 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 based on the acoustic consequences of articulatory parameters other than fronting and backing alone, since discrimination performance between the back group schwas, as well as overall discrimination performance, was very high.

Based on 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 either /i/, /I/, /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 results are shown in Figure 10. 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 percent occurrence. Overall correct performance, represented by scores of the main diagonal, is 42 percent correct, which is well above chance. 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.
"X" STIMULUS

<table>
<thead>
<tr>
<th></th>
<th>E₁</th>
<th>E₂</th>
<th>E₃</th>
<th>I</th>
<th>A</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>E₁</td>
<td></td>
<td>15/33</td>
<td></td>
<td>46</td>
<td>18/22</td>
<td></td>
</tr>
<tr>
<td>F R O N T</td>
<td>E₂</td>
<td></td>
<td>11/33</td>
<td></td>
<td>22/22</td>
<td></td>
</tr>
<tr>
<td>E₃</td>
<td></td>
<td>33</td>
<td></td>
<td>39</td>
<td>21/22</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
<td>13/33</td>
<td></td>
<td>64</td>
<td>17/22</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td>33/44</td>
<td></td>
<td>64</td>
<td>11/11</td>
<td></td>
</tr>
<tr>
<td>B A C K</td>
<td>A</td>
<td></td>
<td>75</td>
<td></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>U</td>
<td></td>
<td>38/44</td>
<td></td>
<td>8</td>
<td>11/11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>86</td>
<td></td>
<td>73</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ACROSS GROUPS = 149/176 = 85%
FRONT VS FRONT = 60/132 = 46%
BACK VS BACK = 19/22 = 86%
TOTAL CORRECT = 520/660 = 79%

Figure 9. 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 "/i/ schwa," and across a back group consisting of a single "/a/ and /u/ schwa." The symbol "E" represents the vowel /i/.
Figure 10. 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. The symbol "E" represents the vowel /i/.

<table>
<thead>
<tr>
<th>STIMULUS</th>
<th>E 240</th>
<th>I 240</th>
<th>U 240</th>
<th>A 240</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>72</td>
<td>71</td>
<td>57</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>30</td>
<td>24</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>54</td>
<td>93</td>
<td>33</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>39</td>
<td>14</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>27</td>
<td>126</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>11</td>
<td>53</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>63</td>
<td>35</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>26</td>
<td>15</td>
<td>47</td>
</tr>
</tbody>
</table>

404/960
42
Finally, notice that the subjects appeared to have more difficulty labeling the front schwas than the back. The /i/ stimulus, for example, was labeled as /i/ 72 times and as /I/ 71 times, whereas the /u/ and /a/ stimuli were labeled correctly 126 and 113 times, respectively. Although it is quite probable that other vocal tract parameters contributed to the increased accuracy in which the back schwas are labeled, we submit that the anticipatory backing gesture observed in the movement data during schwa production is at least one of the articulatory parameters contributing to this effect. That is, the anticipatory tongue backing during schwa production appears to be perceptually significant.

In conclusion, the major findings of this experiment indicate that studies of coarticulation must consider the different components of tongue movement since they appear to have different constraints, and that the consequences of the anticipatory tongue movements appear to be perceptually significant.

REFERENCES