Vowel Stress and Articulatory Reorganization*

Katherine S. Harris+
Haskins Laboratories, New Haven

If a speaker is asked to produce a word which contains a particular speech sound, it can be shown that there will be a great deal of variability in what is produced. Some of this variability depends on the immediately neighboring speech sounds; some depends on the stress and intonation pattern in which the word is imbedded. A principal thrust of recent physiological investigation has been towards showing that at least part of this variability can be accounted for by relatively low-level rules. One formulation of this sort is the suggestion that a shape template, or target, for a speech sound is stored in the nervous system, and that the effects of coarticulation can be described as due to the overlapping effects of several targets at any moment in time (MacNeilage, 1970; Ohman, 1967). The most careful working out of this sort of formulation is probably Lindblom's (1963) ingenious theory of vowel neutralization.

This theory was developed to account for the changes in vowel color which accompany changes in stress. If a vowel is destressed, it will tend to be of shorter duration and to move in vowel color towards the neutral schwa; the latter phenomenon is called vowel neutralization. Lindblom's proposal is that the neutralization is a consequence of the accompanying shortening. Briefly, in a CVC sequence, although the signals sent to the articulators are constant, the response of the articulators is sluggish. If signals arrive at the muscles too fast, the articulators will start towards the vowel target but will be deflected towards the subsequent consonant target— that is, there will be undershoot. Lindblom tested his theory by having subjects produce sentences containing CVC monosyllables. The effect of rearranging the sentences was to change the stress on one "word" and consequently to change the vowel duration. He made careful measurements of the most extreme positions of the first and second formants, as a function of the vowel length. He found that as vowels lengthened, the formants tended towards a target frequency which could be described as a target articulation.

Lindblom's theory seemed to us to be elegant and testable, if one substitutes for "signals" the more specific "muscle contractions." A reformulation in electromyographic terms would then perhaps be: "Under conditions of changing stress (or rate of articulation) the electromyographic signals associated with any vowel will remain constant. Only the spacing between them will change."

+
Also The Graduate Center of the City University of New York.
Some time ago, we performed an experiment on consonants which is relevant here (Harris, Gay, Sholes, and Lieberman, 1968). Subjects produced sentences with one word containing /p/; this word was either heavily contrastively stressed or not. Thus, we would compare "It's the keeper" with "It's the keeper." An electrode in the orbicularis oris muscle measured the strength of the closure contraction, or "signal" in Lindblom's terms.

The results showed a contrast in the amplitude of the EMG signal for the two conditions. However, the effect was quite small—about 20 percent difference between conditions. Furthermore, even this difference was obtained only under conditions of very strong contrastive stress, perhaps stronger than we would observe in ordinary running speech.

We wanted to repeat the experiment with vowels, using stress contrasts more like those in ordinary running speech. The genioglossus muscle, which is active for high vowels, seemed suitable for examination (Harris, 1971; Raphael, 1971; Smith, 1971).

Figure 1 shows the genioglossus muscle. It is a large, fanshaped muscle, which is generally described as bunching and fronting the tongue. The arrow shows the general direction of electrode insertion into the muscle body. Electrode preparation and insertion procedures are described in detail elsewhere (Hirose, 1971).

We constructed a set of nonsense trisyllables, with stress on either the first or the second syllable. The vowel in one syllable was always /i/, while the vowel in the other syllable was /ɔ/ or /u/; /i/ appeared equally often in the first or second syllable and was equally often stressed and unstressed. All conditions were repeated with /p/ as an intervocalic consonant and with /k/. Typical trisyllables, then, would be /pikupa/ and /pupipə/. The subject read sixteen lists in which these nonsense words appeared in random order. The resulting electromyographic signals were recorded and averaged by the usual techniques (Port, 1971).

To return to Lindblom's model, it would lead us to expect a constant muscle signal for the vowel, /i/, with changes in timing of adjacent signals, depending on stress context.

Figure 2 shows the utterances /pikupa/ and /pikupa/. As usual, time runs along the abscissa and the ordinate indicates amplitude of muscle signal. Zero is the point corresponding to the end of voicing in the first syllable. The pair of utterances contrast in whether the first or the second syllable is stressed. If /i/ in the first syllable is stressed, the amplitude for /i/ increases. If /u/ in the second syllable is stressed, /i/ amplitude will decrease. (The vowel /u/ also shows some genioglossus activity, since it is a high vowel.) The amplitude of the stressed syllable is greater than the amplitude of a corresponding unstressed syllable. Of course, we see changes in timing, as well.

Peak heights of the genioglossus activity, averaged over various conditions, are shown in tabular form in Figure 3. This slide shows mean peak height values for four conditions—when /i/ is stressed and unstressed, in the first syllable and in the second. Overall, stress produces greater activity.
Figure 1: Electrode insertion into the genioglossus muscle.
Figure 2: Averaged EMG signals from the genioglossus muscle for tri-syllables which are stressed on the first and second syllable.
<table>
<thead>
<tr>
<th>Syllable Position</th>
<th>Stress Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>S₁</td>
<td>137</td>
</tr>
<tr>
<td>S₂</td>
<td>169</td>
</tr>
</tbody>
</table>

Figure 3: Peak heights of genioglossus activity under various stress conditions.
One further question may be asked. If differences in signal size contribute to stressing differences, is there any evidence that the duration mechanism works as well? Lindblom's model says, in essence, that the longer the vowel, the less neutral. To consider this question, we must extend the model to yet another situation, the vowel duration differences which accompany the shift from voiceless to voiced terminal consonants.

This phenomenon is extremely well known. Briefly, the vowel before a voiceless stop or fricative is shorter than before the corresponding voiced consonant. Now, let us assume that Lindblom's mechanism is at work in running speech. If the time distance between the vowel signal and the consonant signal is shorter for voiceless than for voiced consonants, then one of two things must happen: either the vowel must be more neutralized before voiceless stops, or alternatively, there should be an adjustment of peak activity to compensate for the duration difference. There is no evidence, either in our own work or, so far as I know, in the extensive literature on the voicing effect, that the vowel before a voiceless consonant is more neutralized than before a voiced stop, although we should, of course, check spectrograms, which has not yet been done. Some data collected by Raphael (1971) allow us to examine the second possibility.

Figure 4 shows genioglossus activity for four high front vowels in the frame /pVp/. There is substantial genioglossus activity for /i/ and /e/ before /p/ for this subject but relatively little activity for their so-called "lax" counterparts. Since the genioglossus is apparently a chief determiner of vowel color for /i/ and /e/, we would expect an adjustment in peak height to compensate for the difference in vowel length before voiced and voiceless consonants. On the other hand, we have no such anticipation with respect to /ɛ/ and /ɪ/, since they show very little activity.

Figure 5 shows peak heights for the four vowels before a series of voiced and voiceless consonant pairs. Overall, peak activity is lower for the voiced member of the pair, although there is one case of approximate equality. The situation is reversed before the lax vowels--I have no idea why. For long vowels this result can be interpreted as a tendency to compensate for duration differences, with peak size changes, for "essential" muscles. This compensation anticipates the duration difference, that is, the speaker seems to make some sort of anticipatory calculation.

Figure 6 shows peak values for a second subject, who used relatively high values of genioglossus activity for all four vowels (though notice that /e/ is strongly diphthongized for this speaker, so that only the second peak corresponding to /i/ or /ɪ/, is high). We would, therefore, expect compensation for voicing distinctions in all four vowels.

Figure 7 shows peak heights for the four vowels. We looked only at two sets of voiced/voiceless pairs for this subject. There are two entries for /e/, the diphthongized vowel, one for each peak. We would expect greater activity for the voiceless member of the pair for all four vowels, and indeed, this is about what we get, though there is one case of approximate equality. For /e/, only the second peak shows voicing compensation.

Let me summarize at this point. We have produced some rather preliminary evidence that stressing may affect the size of the contraction signals to
Figure 4: Averaged EMG signals from the genioglossus muscle for four high front vowels; speaker LJR.
<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>b</th>
<th>t</th>
<th>d</th>
<th>k</th>
<th>g</th>
<th>s</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>653</td>
<td>606</td>
<td>702</td>
<td>644</td>
<td>694</td>
<td>625</td>
<td>532</td>
<td>537</td>
</tr>
<tr>
<td>I</td>
<td>236</td>
<td>319</td>
<td>316</td>
<td>360</td>
<td>268</td>
<td>494</td>
<td>223</td>
<td>294</td>
</tr>
<tr>
<td>e</td>
<td>643</td>
<td>533</td>
<td>710</td>
<td>538</td>
<td>643</td>
<td>571</td>
<td>530</td>
<td>500</td>
</tr>
<tr>
<td>E</td>
<td>140</td>
<td>154</td>
<td>179</td>
<td>331</td>
<td>206</td>
<td>481</td>
<td>172</td>
<td>195</td>
</tr>
</tbody>
</table>

Figure 5: Peak heights of genioglossus activity for four high front vowels before voiced/voiceless pairs; speaker LJR.
Figure 6: Averaged EMG signals from the genioglossus muscle for four high front vowels; speaker KSH.
<table>
<thead>
<tr>
<th>Vowel</th>
<th>p</th>
<th>b</th>
<th>k</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>177</td>
<td>158</td>
<td>199</td>
<td>111</td>
</tr>
<tr>
<td>I</td>
<td>190</td>
<td>148</td>
<td>110</td>
<td>111</td>
</tr>
<tr>
<td>e</td>
<td>95/179</td>
<td>135/136</td>
<td>103/116</td>
<td>95/93</td>
</tr>
<tr>
<td>e</td>
<td>163</td>
<td>127</td>
<td>102</td>
<td>95</td>
</tr>
</tbody>
</table>

Figure 7: Peak heights of genioglossus activity for four high front vowels before voiced/voiceless pairs; speaker KSH.
muscles, as well as their timing, although, by generous overinterpretation of the data, we can find some evidence for the effectiveness of a timing change mechanism, as well. However, if we presume that the "extra energy" mechanism works at all, it really originates more problems than it solves, since it leaves the question of what is invariant about a vowel under two stress conditions. Presumably, each vowel would be characterized by a pattern of contractions; however, if the size of one member of the pattern changes, what happens to the others?

Vowel height can be shown to be a joint product of tongue height and jaw opening. If genioglossus activity changes under stress, does the activity of the anterior belly, and the other muscles which open the jaw, increase proportionately? It seems far more likely that, for any vowel, only a selected group of muscles increase activity under stress. If this is indeed so, then the pattern of activity for any vowel becomes different, not only in "size" but in configuration, for changes in stress.

How does all this affect our views of speech mechanisms? The most common model for afferent feedback is that there is, for any phone, a "target" articulation, which is represented either as a position in the mouth or, more specifically, as a set of muscle lengths of each phone. These two hypotheses differ in their specificity. In the second case, not only is a target required, but the target must be reached by the same set of muscle adjustments each time. A recent observation (Lindblom and Sundberg, 1971) shows that if a speaker must attain a given tongue height with a jaw opening that is constrained by a block holding the jaws open at a fixed distance, he will use a compensating adjustment of the muscles to raise the tongue. Some data of Borden's (1972) can be interpreted to mean that if one of a set of muscles is partially paralyzed, other muscles will attempt to compensate by more than normal activity. These observations seem to me to indicate that a target representation in muscle length terms is probably not a sensible one. The simple continuous gamma loop correction models, depending on attainment of a set of lengths, would seem to fall with this evidence. "Targets" must somehow be specified in position coordinates which allow for configuration flexibility. The study we reported here seems to indicate that a given vowel must be represented as a series of targets which differ from some neutral point, and which are arrived at by different muscle action patterns. Single loop correction does not seem capable of operating successfully on targets which change in this fashion.

REFERENCES


