Mechanisms of Duration Change*

K. S. Harris†
Haskins Laboratories, New Haven, Conn.

ABSTRACT

Changes in stress, speaking rate, and terminal consonant are known to modify the duration of vowels in spoken utterances (Lehiste, 1970). Electromyographic investigation permits a detailed examination of the mechanisms underlying these observed effects; in particular, it is possible to test Lindblom's (1963) hypothesis that context-dependent vowel color alterations result from a change in the timing of the signals to the articulators, rather than from a reorganization of the articulatory process. Results suggest that the articulatory process itself is reorganized, and that reorganization is different for the three types of change.

INTRODUCTION

In 1963, Lindblom wrote a classic paper on the effects of variations in word stress on the target formant position of vowels. The object of his experiments was to show that the so-called "vowel neutralization phenomenon" could be derived from a very simple model of upper vocal tract control. The phenomenon itself is well-known; briefly, as a syllable is destressed, its vowels will tend to be more neutralized, as well as shorter in duration. Lindblom suggested that the neutralization is a consequence of the shortening. He made a number of spectrographic measurements, showing a regular relationship between duration of vowels and their target formant positions. The relationship is consonant with a model in which the signals sent to the articulators are determined by a stored template for each vowel, independent of its stress position; if signals are sent to the articulators at rates greater than some critical value, target position is not attained before new signals arrive; thus, the shorter the vowel, the greater the target undershoot. In his original paper, Lindblom (1963) suggests that the same model can be applied to the effects of changes in speaking rate.


†Also the Graduate School of the City University of New York.

Acknowledgment: The series of experiments reported in part here were performed in close association with L. Raphael, T. Gay, and T. Ushijima.

and changes in stress on vowel color. The model presumably could be extended to apply to any context effect that might be expected to cause changes in vowel duration, such as the well-known effect of the voicing status of the final consonant on the vowel. Lindblom did not suggest the level at which constant signals are presumed to be sent to the articulators. The simplest suggestion is that the signals to the muscles might be constant. If this were true, we might expect electromyographic (EMG) signals to the muscles to be of equal size, under conditions of varying stress, speaking rate, and voicing status of the final consonant. A secondary result of Lindblom's model is that the timing relationship between consonant and vowel signals may be expected to change, as the duration of the vowel changes.

The experiment described here was the latest in a series of tests of Lindblom's hypothesis; the results will be compared with a number of related experiments.

**EXPERIMENT**

**Procedure**

A single speaker recorded the four-syllable nonsense utterances /apiapia/ and /apiaiba/ under several conditions: stress was placed on either the first or the second /i/; there were two speaking rates; and thus, there was a total of eight utterance types. Utterances were arranged in random lists of 28, with slow and fast lists alternated. After the removal of faulty utterances, there were between 24 and 33 utterances of each type for averaging.

Hooked-wire electrodes were inserted bilaterally into the anterior portion of the genioglossus muscle (GGA). A single electrode was inserted into the posterior portion of the genioglossus muscle (GCP). Results from other placements for this experiment will not be discussed here. Electrode construction and placement are discussed in Hirose (1971).

EMG data were amplified and recorded on 16-channel instrumentation tape, together with the acoustic signal and code pulses for later computer analysis. After inspection for artifacts, EMG signals were processed and averaged by techniques previously described by Port (1971) and Kewley-Port (1973).

Wide-band spectrograms were made of 56 utterances, half from near the beginning and half from near the end of the recording session. Thus, acoustic records were available of seven utterances of each type. On the spectrograms, measurements were made of the duration of the two syllables containing /i/, from the release of closure to closure or the end of voicing. The second and third formants were measured in each syllable at their peak frequency.

**Results**

Averaged EMG curves for one of the GGA leads are shown in Figure 1. The point on the time line marked "zero" is the averaging lineup point for each utterance, and is at the /p/ closure of the second syllable. The duration of the acoustic events is indicated above each figure. Clearly the EMG signal associated with each stressed syllable has a somewhat larger peak height, and a somewhat longer duration, than its unstressed counterpart. This is in accord with earlier results (Harris, 1973). Further, there is a systematic tendency
for the fast speaking condition to show smaller peak heights than the slow condition (Gay, Ushijima, Hirose, and Cooper, 1974). Although it is less obvious, there is no overall systematic trend for vowel peaks associated with terminal /p/ and /b/.

Correlation coefficients were calculated between various measures. Results are shown in Table 1. Correlations are quite high, and are uniformly significant

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( F_2 )</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>duration</td>
</tr>
<tr>
<td>( F_2 )</td>
</tr>
<tr>
<td>( F_3 )</td>
</tr>
<tr>
<td>( GGAR )</td>
</tr>
<tr>
<td>( GGA_L )</td>
</tr>
</tbody>
</table>

at the .05 level or greater. The correlations between the formant levels and duration are essentially a reiteration of Lindblom's results, in a somewhat different form—that is, formants reach a more extreme value as duration lengthens. The correlations between peak height and duration, and peak height and formant value, however, are a contradiction of Lindblom's hypothesis.

While the overall correlations are rather high, they are somewhat misleading, since the effects of the terminal /b/ on half the utterances are masked. More detailed results, for the second syllable only, are presented in Table 2.

<table>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{S}_2 )</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>duration in msec</td>
</tr>
<tr>
<td>( F_2 ) in Hz</td>
</tr>
<tr>
<td>( F_3 ) in Hz</td>
</tr>
<tr>
<td>( GGAR ) in ( \mu V )</td>
</tr>
<tr>
<td>( GGA_L ) in ( \mu V )</td>
</tr>
<tr>
<td>( GGP ) in ( \mu V )</td>
</tr>
</tbody>
</table>

196
An inspection of the table shows the following results:

1) The expected effects of stress, speaking rate, and voicing on the duration of the second syllable are obtained.

2) Values of F2 and F3 are more extreme for slow speech and for stressed production. However, there is no systematic tendency for the values to differ for terminal /p/ and /b/. The result is not surprising in view of the classic literature. So far as I know, it has never been suggested that vowels are more neutral before voiceless consonants.

3) Stress and speaking rate affect the peak values of muscular activity. In ten of twelve possible comparisons, peaks are higher for terminal /p/ than for terminal /b/. The result suggests a systematic trend. However, Raphael (1974) has examined the peak heights associated with vowel production in a large number of utterances in which the terminal consonant is a voiced or voiceless stop or fricative. His results show a prolongation of the vowel signal before voiced consonant, but no systematic differences in amplitude. Obviously, this result requires more systematic examination.

DISCUSSION

It is interesting to consider Lindblom's explanation of the stress effect in light of the picture it gives of the organization of running speech. In his model, pushed to an extreme, a series of signals are sent to the articulators, which depend for their identity on the phonetic specification of the segments. Changes of stress or speaking rate will affect the timing of the arrival of these signals (as will certain segmental characteristics of the sequence, by extension) but not their relative size. The resulting acoustic output will vary, not because of variation in the signal size, but because of changes in the relative timing of, for example, successive vowel and consonant signals. Furthermore, differences between contexts vary along the single dimension of time.

The results described above suggest that the real picture is substantially more complex. Signal size for vowels varies systematically with duration for changes in stress and speaking rate, and does not (apparently) vary with the duration changes conditioned by the voicing status of the terminal consonant. Let us examine the stress and speaking rate variations first, since Lindblom's model is intended to apply only to them. Is the target position observed due entirely to the size difference observed, or may the result be due in part to Lindblom's proposed mechanism? In Lindblom's model, any duration change automatically generates a change in target position, unless it is counteracted by some other adjustment. Therefore, if vowels are longer preceding /b/, then signals should be smaller, if the acoustic target for the formants is to be the same. As noted above, in the present experiment, there is a trend in this direction, although the same trend is not seen in other experiments. This point must be examined further.

In the present experiment, the effects of stress and speaking rate are at least qualitatively homogeneous. However, there is some evidence that consonant and vowel signals do not behave in parallel ways under the two manipulations.
Some years ago, we found that consonants associated with heavily stressed syllables will be produced with stronger associated articulation (Harris, Gay, Sholes, and Lieberman, 1968). Gay et al. (1974) have found that faster speaking rates are associated with higher consonant peak heights. These effects are in opposite directions to the associated vowel articulations. Our information is, however, concerned with the somewhat special circumstance of the labial consonant surrounding the vowel, and should be examined in more varied environments.

Lindblom's model is in one respect similar to a much earlier formulation proposed by Cooper, Liberman, Harris, and Grubb (1958). In both models, constant signals yield a variable output, due to variability in timing. In all the experiments we have performed in manipulating stress, speaking rate, and context, we seem to get the opposite result. The relative timing of consonant and vowel gestures to different articulators seems to be very closely time locked, while the amplitude and duration of gestures for particular segments vary substantially. We will be interested to see what happens in those conditions where the same articulator is involved in both consonant and vowel gestures.

REFERENCES


