Effect of speaking rate on vowel formant movements

Thomas Gay

University of Connecticut Health Center, Farmington, Connecticut 06032
and Haskins Laboratories, New Haven, Connecticut 06511
(Received 13 June 1977; revised 13 September 1977)

The purpose of this experiment was to study the effects of changes in speaking rate on both the attainment of acoustic vowel targets and the relative time and speed of movements toward these presumed targets. Four speakers produced a number of different CVC and CVVCVC utterances at slow and fast speaking rates. Spectrographic measurements showed that the midpoint formant frequencies of the different vowels did not vary as a function of rate. However, for fast speech the onset frequencies of second formant transitions were closer to their target frequencies while CV transition rates remained essentially unchanged, indicating that movement toward the vowel simply began earlier for fast speech. Changes in both speaking rate and lexical stress had different effects. For stressed vowels, an increase in speaking rate was accompanied primarily by a decrease in duration. However, stressed vowels, even if they were of the same duration as quickly produced stressed vowels, were reduced in overall amplitude, fundamental frequency, and to some extent, vowel color. These results suggest that speaking rate and lexical stress are controlled by two different mechanisms.

PACS numbers: 43.70.Gv, 43.70.Bk

INTRODUCTION

Within certain limits, speech perception does not appear to be constrained by rate of speech production; the information-bearing elements of segmental units are preserved across a wide range of speaking rates. Does the perceptual mechanism adapt to a different acoustic representation of these elements during fast speech, or are the articulatory gestures reorganized to produce a constant acoustic output? Some physiological evidence exists to support the latter. In a series of experiments (Gay and Birose, 1973; Gay et al., 1974; Gay and Ushijima, 1975) it was shown that the motor patterns underlying articulatory movements for fast speech were not only different than those during slow speech, but were reorganized in complex ways. In general, electromyographic activity associated with tongue body movements during vowel production decreased during fast speech, while activity associated with both labial and alveolar stop consonant production increased with an increase in speaking rate. At the movement level, while it would appear that vowel targets are not always reached during fast speech (Gay et al., 1974), the tradeoffs between articulatory displacement and velocity seem to vary for individual speakers (Kuehn and Moll, 1976).

While it is apparent that changes in both motor programming and articulatory movements occur for changes in speaking rate, it is not known how these changes are reflected in the acoustic signal. Are acoustic targets the same for speech produced at slow and fast rates, or are these presumed targets systematically centralized, or otherwise shifted in frequency, as a function of rate? What are the temporal properties of CV transition movements for different speaking rates; do onset frequencies and rates of transition movements change for fast speech? The experiment reported in this paper was designed to study these questions by mapping the acoustic vowel space of several speakers across changes in speaking rate. A second purpose of the experiment was to study the acoustic effects of changes in speaking rate in relation to those for lexical stress to determine whether the two features can be accounted for by the same duration control mechanism.

I. METHOD

A. Subjects and speech material

Speakers were four adults, three males (WE, TG, LR) and one female (KB), all native speakers of American English. Two (TG, LR) spoke a New York dialect, one (KB) a New England dialect, and one (WE) a General American (West Coast) dialect. While all four speakers were phonetically trained and experimentally sophisticated, none, except the author, knew the specific research goals.

Three different types of speech samples were constructed. The main set consisted of CVC syllables containing the nine vowels, /i, a, a, u, /, in a /b p/ environment. This frame was used for two reasons: One was that it paralleled that of an earlier physiological experiment (Gay et al., 1974), and the other is that it would probably produce minimal contextual effects. A second set consisted of a corresponding CVC subset with the point vowels, / a, u, in a /b p/ environment, and a third consisted of CVVCVC sequences of the type, /kitap/, /'kitap/, /kapip/, /ka'pip/. The second set was used to provide a voicing contrast to parallel points of the main set, while the third was used to study the effects of changes in both speaking rate and lexical stress on the same syllable types. The 18 speech samples were arranged, randomly within each set, into a list. Each utterance was embedded in the carrier phrase, "It's a ____ again." Five such lists were constructed, one for each of five repetitions by each speaker.
measurements were made from the spectrograms at points indicated in Fig. 1. Duration measurements were made for the stop gap closure of the initial consonant, CV transition, a combined measurement of the vowel nucleus and, if present, the final VC transition, and closure for the final consonant. The VC transition component was included in the vowel nucleus measurement because it was difficult to segment out. Formant frequency measurements \(F_1, F_2, F_3\) were made at the time of release of the initial consonant, the vowel midpoint, and at the time of closure for the final consonant. If the CV transition was not visible at the time of consonant release, its position was straight-line extrapolated. The vowel midpoint was defined in one of three ways: (1) the point where the \(F_2\) transition reached a steady state, (2) if a steady state was not reached, the point where the \(F_2\) transition reached maximum displacement before changing direction, or (3) if the transition was unidirectional, at a point midway between the onset of voicing and closure for the final consonant. Most of the vowels met criteria 1 or 2; however, for two speakers, both \(/u/\) and \(/o/\), consistently, and \(/i/\) and \(/e/\), occasionally, were characterized by unidirectional glides from initial consonant to final consonant. The method of tracking and locating the measuring points was the usual one: A pencil line was drawn through the center of both the transition and steady-state (where present) portions of the formant; measurements were made where the line intersected the point of consonant release (for the transition onset) and at either the center of the steady-state portion or where the line intersected that of the CV transition.

First and second formants were visible for all subjects, but \(F_3\) did not always appear clearly for speakers TG and LR, and for most samples of the stressed syllables. As would be expected, CV movements for \(F_1\) were small, while those for \(F_2\) provided the most reliable and useful transition movement information. Repeated measurements of selected samples revealed fairly consistent error ranges. Duration measurements were accurate to \(\pm 10\) ms, while formant frequency measurements were accurate to \(\pm 25\) Hz. The latter range is consistent with that of other reports (Lindblom, 1961; Öhman, 1965). For the most part, the token-to-token variability for each subject was small. Durations (as measured from closure of the initial consonant to release of the final consonant) usually varied by no more than 25 ms within each rate compared to an average of 45 ms for across-rate differences. Similarly, token-to-token formant frequency variations usually fell within the range of error measurement (\(\pm 25\) Hz). This stability might be due to the highly structured nature of the utterances and the stress pattern of the sentence. There were, of course, some exceptions to this general stability; these exceptions will be discussed in the following section.

## II. RESULTS

The effects of differences in speaking rate on the durations of each of the segments of the main set of CVC syllables are summarized in Table I. The purpose of this table is to show how the overall decrease in syllable
TABLE I. Mean durations of consonant closure (1), vowel, and consonant closure (2), pooled over all repetitions and speakers. For each vowel, values for the slow rate appear on the first row, those for the fast rate on the second row.

<table>
<thead>
<tr>
<th>Consonant closure (1)</th>
<th>Vowel</th>
<th>Consonant closure (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>Ratio</td>
<td>Duration</td>
</tr>
<tr>
<td>i</td>
<td>100</td>
<td>0.95</td>
</tr>
<tr>
<td>95</td>
<td>90</td>
<td>80</td>
</tr>
<tr>
<td>l</td>
<td>105</td>
<td>0.90</td>
</tr>
<tr>
<td>95</td>
<td>95</td>
<td>85</td>
</tr>
<tr>
<td>e</td>
<td>105</td>
<td>0.90</td>
</tr>
<tr>
<td>95</td>
<td>105</td>
<td>90</td>
</tr>
<tr>
<td>æ</td>
<td>105</td>
<td>0.90</td>
</tr>
<tr>
<td>95</td>
<td>125</td>
<td>70</td>
</tr>
<tr>
<td>a</td>
<td>100</td>
<td>0.90</td>
</tr>
<tr>
<td>90</td>
<td>115</td>
<td>75</td>
</tr>
<tr>
<td>ɔ</td>
<td>105</td>
<td>0.95</td>
</tr>
<tr>
<td>100</td>
<td>130</td>
<td>70</td>
</tr>
<tr>
<td>ʊ</td>
<td>105</td>
<td>0.95</td>
</tr>
<tr>
<td>95</td>
<td>90</td>
<td>75</td>
</tr>
<tr>
<td>ɔ</td>
<td>100</td>
<td>0.95</td>
</tr>
<tr>
<td>95</td>
<td>90</td>
<td>75</td>
</tr>
<tr>
<td>ʌ</td>
<td>100</td>
<td>0.90</td>
</tr>
<tr>
<td>90</td>
<td>85</td>
<td>75</td>
</tr>
<tr>
<td>Mean</td>
<td>105</td>
<td>130</td>
</tr>
<tr>
<td>95</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

duration during faster speech is absorbed by each of the constituent segments. The table shows the durations of initial /p/ closure, CV transition, vowel nucleus, and final /p/ closure, pooled over both utterance repetitions and speakers. Each value represents the mean of 20 (5 repetitions×4 speakers) tokens. While the actual durations of the different syllables varied for the four speakers, relative differences, both among vowels and between rates, are represented in the means.

The table shows, not unexpectedly, that the reduction in duration during fast speech is reflected primarily in the duration of the vowel, although perhaps not to as great a degree as might be intuitively expected. Differences in the durations of the vowel nuclei for slow and fast speech ranged from 20 to 35 ms, depending on the particular vowel. While the overall duration of the vowel varied from 105 ms for /i/ to 165 ms for /æ/ at the slow rate, the percentage change did not vary as a function of duration. For example, the phonetically long vowels /ø/ and /ø/ were not reduced to any greater or lesser percent during fast speech than the phonetically short vowels, /i/ or /a/. The slow rate vowel durations obtained in this experiment are considerably shorter than those reported by Peterson and Lehiste (1960) (245 ms) for CVC words in a similar carrier, but slightly longer than those reported by Klatt (1975) (110 ms) for vowels spoken in connected discourse. These differences are probably related to differences in phonetic context as well as utterance position in the sentence carrier.

Although the vowel nucleus absorbed most of the decrease in duration during fast speech, the consonant segments were also consistently shorter as well. However, differences in duration for initial and final consonant closure were considerably less between the two rates; also, greater variability and even some overlap occurred in certain instances at the individual token level. Pre-stressed initial /p/ is consistently, and usually substantially, longer than post-stressed final /p/, for both rates, as expected. Interestingly, although the vowel portion is most affected during fast speech, the contributions of the initial and final consonants account for at least one third of the total reduction in syllable duration. It was also found that the transition durations within each rate were relatively stable across the different vowels. However, transition time was reduced somewhat during fast speech, to about the same degree as that for consonant closure, some 5–10 ms. Transition times ranged from 40 to 50 ms for the slow rate and 35 to 45 ms for the fast rate. The stable transition times across vowels are consistent with the articulatory data of both Kent and Moll (1969) and Kuehn and Moll (1976), but shorter and less variable than those reported by Lehiste and Peterson (1961) and Öhman (1965). These differences might be due to differences in overall duration, phonetic context, and carrier structure.

The major question of interest in this paper is whether the acoustic targets of vowels (as measured at the midpoint) vary, either systematically or unsystematically, as a function of speaking rate. Spectrographic measurements of first, second, and third formant frequencies show that they do not. Figure 2 shows the F1–F2 vowel space for all nine vowels produced by each speaker.

![Figure 2. F1–F2 vowel space for midpoint measurements for both speaking rates.](image-url)
Each data point represents the mean of the five repetitions by that speaker at each rate. The overall picture is one of little variability. For the most part, the means for each rate (both F1 and F2) are quite close, usually falling well within the range of error measurement. However, some instances of variability between the slow and fast rates do occur. For speaker WE, F1 for /e/ is approximately 50 Hz lower for the fast rate. This difference, statistically significant at the 0.02 level of confidence (t test for independent means), might be explained by jaw undershoot during fast speech; however, if this is true, it is curious why the other open vowels are not similarly affected. Speaker LR also showed some front vowel variability, but in this case, only F2 for /e/ is significantly different (0.01 level of confidence). The greater variability between rates appears in the data for the single female speaker (KH). While the range of variability is still small, it is nonetheless greater than that for the other speakers. In addition, variability within each rate is greater for this speaker than for any of the others. The wider range of variability for this speaker might be a simple consequence of the increased probability of error encountered when measuring female formant frequencies, a speculation that is supported by the fact that only the means of F1 for /e/ were significantly different (0.05) as a function of rate.

While vowel durations for three of the four speakers were distributed essentially bimodally between the two speaking rate conditions, one speaker (TG) often showed considerable temporal variability, with vowel durations distributed along a continuum. One such instance was for the vowel /a/. Figure 3 shows the first, second, and third formant frequencies (measured at the midpoint) for /a/ plotted as a function of duration (transition + nucleus). The 20 tokens represent all of those produced at both rates for both the /p/ and /b/ syllables for this speaker. It is apparent from this figure, that the effect of speaking rate on the attainment of acoustic vowel targets is negligible, even over a wide (55 ms) range of durations. The ranges of formant frequency variations are 50 Hz for F1, 75 Hz for F2, and 75 Hz for F3. It should be noted that these midpoint frequencies are attained through CV transitions that originate, at consonant release, some 300–400 Hz lower in frequency. It is also apparent that the observed variability is not correlated with duration along any part of the continuum.

Assuming that the same acoustic target is reached for a vowel spoken at both slow and fast rates, the question becomes how the underlying articulatory gesture is modified to achieve that constant end. Either of two types of adjustments seem likely: First, articulator movement toward the vowel target can begin earlier in time, that is, closer to the time of initial consonant closure, or second, the movement can be produced at a faster rate of speed. Evidence for the first can be seen spectrographically by a change in the position of the onset frequency of the CV transition as a function of rate; specifically, the onset frequency of the formant transition would be closer to that of the target frequency for faster speech. An estimation of the second could be made by simply calculating the overall velocity of the formant transition itself. These measurements appear for the vowel /a/ for all speakers in Table II. This table shows the mean durations for /p/ closure, the CV transition and the vowel nucleus, and the F2 onset and midpoint frequencies and F2 rate of change, for both speaking rates. The means are pooled over utterance repetitions. The F2 rate was calculated by dividing the difference between the F2 midpoint and F2 onset frequencies by the transition duration.

For all subjects, the onset frequency of the second formant transition is higher for the fast rate condition, while the F2 midpoint frequencies and F2 rates of change remain essentially unaffected across the two rates. The differences in mean onset frequencies are statistically significant (at various levels of confidence) for all but the female speaker. The mean differences range from 40 Hz for subject WE to 80 Hz for subject TG. The corresponding differences in midpoint frequencies, however, are only of the order of 15–25 Hz for all speakers. It should also be noted that these shifts occur in syllables whose closure durations are shorter, a condition which, theoretically at least, would tend to minimize the observed effects.

### Table II. Duration and formant frequency measurements for the vowel /a/, pooled over utterance repetitions.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Duration</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Closure</td>
<td>Transition</td>
</tr>
<tr>
<td>WE</td>
<td>95</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>40</td>
</tr>
<tr>
<td>TG</td>
<td>95</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>50</td>
</tr>
<tr>
<td>KH</td>
<td>105</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>110</td>
<td>45</td>
</tr>
<tr>
<td>LR</td>
<td>110</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>105</td>
<td>40</td>
</tr>
</tbody>
</table>

*aSignificant at 0.10.  *Significant at 0.05.
TABLE III. Duration and formant frequency measurements for
voiced–voiceless contrasts. Values are pooled over all repeti-
tions for the three male speakers.

<table>
<thead>
<tr>
<th>Utterance (b/p)</th>
<th>Closure duration</th>
<th>Transition duration</th>
<th>F2 onset</th>
<th>F2 midpoint</th>
<th>F2 rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>pip</td>
<td>100</td>
<td>50</td>
<td>1810</td>
<td>2120</td>
<td>6.6</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>45</td>
<td>1885</td>
<td>2120</td>
<td>6.5</td>
</tr>
<tr>
<td>blp</td>
<td>105</td>
<td>45</td>
<td>1960</td>
<td>2120</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>40</td>
<td>1985</td>
<td>2105</td>
<td>3.2</td>
</tr>
<tr>
<td>pap</td>
<td>95</td>
<td>50</td>
<td>1340</td>
<td>1170</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>45</td>
<td>1325</td>
<td>1180</td>
<td>2.0</td>
</tr>
<tr>
<td>bap</td>
<td>105</td>
<td>...</td>
<td>1130</td>
<td>1150</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>...</td>
<td>1160</td>
<td>1170</td>
<td>...</td>
</tr>
<tr>
<td>pap</td>
<td>95</td>
<td>40</td>
<td>1250</td>
<td>1000</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>40</td>
<td>1200</td>
<td>990</td>
<td>3.9</td>
</tr>
<tr>
<td>bup</td>
<td>100</td>
<td>45</td>
<td>1065</td>
<td>990</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>35</td>
<td>1030</td>
<td>990</td>
<td>1.2</td>
</tr>
</tbody>
</table>

A difference in second formant transition onset frequen-
cies between slow and fast speech is also evident
where /b/ is the initial consonant. Measurements for
/b/ appear in Table III, which contains essentially the
same data as in Table II but for the /pip–bip/, /pap–
bap/, and /pip–bip/ contrasts, pooled over the three
male speakers. For these three vowels, at least, the
rate effects that exist for /p/, exist for /b/ as well, ex-
cept perhaps to a slightly lesser extent. Again, for all
three vowels preceding /b/, the onset frequency of the
second formant transition is closer to the midpoint fre-
quency, while the midpoint frequencies, themselves, and
the F2 rates remain largely unaffected. The differences
in onset frequencies is slightly less for /b/ than /p/.
For /bip/, the shift across rates is 25 Hz, while for
/pip/, the shift averages 60 Hz. For /a/ and /u/, how-
ever, the shifts are more comparable.

More obvious than the rate differences are the differ-
ences in second formant transition onset frequencies and
rates of change between /p/ and /b/, within each speak-
ing rate condition. For all three vowels, it is apparent
that the onset frequencies are considerably closer to the
midpoint frequencies, and the F2 rate is correspondingly
slower, for /b/ as opposed to /p/. All frequency differ-
ences between /p/ and /b/ were statistically significant
for all four speakers at either the 0.05 and 0.01 level of
confidence. The absence of a CV transition for /bap/
indicates that the movement towards /a/ from /b/ was
probably completed before release of the consonant oc-
curred, much earlier than the corresponding transition
from /p/. The greater range of F2 onsets for /b/ as op-
posed to /p/ is consistent with Fant’s (1969) calcula-
tions for similar CV syllables; even the extent of the
differences is similar. One explanation for this frequency
shift is that because of the greater tenseness associated
with /p/, the tongue is not as free to move toward the
vowel as if it is for the more lax /b/; in other words, the
tongue is more free to articulate with the following
vowel during /b/ than it is during /p/. A second possi-
bility is that the voiceless consonant /p/ occurs earlier
in time during the vowel-to-vowel movement than /b/,
producing transitions that are temporally offset to the
left. It is also conceivable that part of the frequency dif-
ference between /pap/ and /bap/ is due to the presence
of an additional subglottal formant associated with /p/
release (Fant, 1972). In any event, while changes in
phonetic context affect the pattern of movements to-
ward the vowel target, the frequencies of the targets,
themseves, do not appear to be affected, nor is the ba-
sic rate effect on the movements different. The rate
effects seem to be superimposed on the context de-
pendent articulatory movements, and are not affected by,
or assimilated into, these movements.

The final set of measurements is related to the ques-
tion of how changes in both speaking rate and lexical
stress affect the acoustic properties of vowels. Are the
effects of these two features, both of which affect vowel
duration, additive or independent? The duration and
frequency measurements for both the /a/ and /a/ stress
contrasts appear in Tables IV and V. These tables show
the measurements of vowel duration, relative overall
amplitude, fundamental frequency, and first and second
formant frequencies for the second syllable of the utter-
ance, for each speaker. All frequency and amplitude
measurements were made at the vowel midpoint. The
amplitude measurements are in dB relative to the least
intense utterance (=0) for each speaker. All values are
pooled over the five utterance repetitions.

For both /a/ and /u/, a number of differences appear

TABLE IV. Duration, relative amplitude, and frequency measurements for the vowel /I/ as a function of
both stress and speaking rate.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Duration</th>
<th>Rel. Amp.</th>
<th>F0</th>
<th>F1</th>
<th>F2</th>
<th>Duration</th>
<th>Rel. Amp.</th>
<th>F0</th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>WE</td>
<td>140</td>
<td>6</td>
<td>140</td>
<td>300</td>
<td>2150</td>
<td>115</td>
<td>0</td>
<td>110</td>
<td>300</td>
<td>2125</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>6</td>
<td>150</td>
<td>300</td>
<td>2120</td>
<td>95</td>
<td>0</td>
<td>110</td>
<td>325</td>
<td>2100</td>
</tr>
<tr>
<td>TG</td>
<td>90</td>
<td>6</td>
<td>125</td>
<td>315</td>
<td>2155</td>
<td>80</td>
<td>5</td>
<td>95</td>
<td>335</td>
<td>2085</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>8</td>
<td>135</td>
<td>325</td>
<td>2125</td>
<td>70</td>
<td>0</td>
<td>95</td>
<td>330</td>
<td>2050</td>
</tr>
<tr>
<td>KH</td>
<td>125</td>
<td>7</td>
<td>225</td>
<td>340</td>
<td>2710</td>
<td>100</td>
<td>3</td>
<td>150</td>
<td>425</td>
<td>2530</td>
</tr>
<tr>
<td></td>
<td>105</td>
<td>4</td>
<td>220</td>
<td>320</td>
<td>2670</td>
<td>75</td>
<td>0</td>
<td>160</td>
<td>430</td>
<td>2505</td>
</tr>
<tr>
<td>LR</td>
<td>115</td>
<td>10</td>
<td>140</td>
<td>315</td>
<td>2085</td>
<td>95</td>
<td>0</td>
<td>110</td>
<td>320</td>
<td>2090</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>7</td>
<td>135</td>
<td>335</td>
<td>2120</td>
<td>80</td>
<td>0</td>
<td>110</td>
<td>330</td>
<td>2010</td>
</tr>
</tbody>
</table>

between the stressed and unstressed pairs for each speaking rate, but virtually no differences emerge between rates for the same stress condition. The slow and fast pairs within each stress condition are characterized by essentially the same overall amplitude, fundamental frequency, and first and second formant frequencies. However, the corresponding unstressed syllables at each rate are consistently lower in overall amplitude and fundamental frequency, and somewhat reduced in vowel color. Fundamental frequency differences were statistically significant for all speakers (0.01) for the stressed–unstressed pairs while reduction of F1 and F2 was significant (0.05) only for speaker KH.

The overall stress findings are, of course, consistent with those of a number of earlier studies (Fry, 1955; Lieberman, 1960; Lindblom, 1963; Brown and McGlone, 1974). Of particular interest in the present data, however, is that these differences are apparently not related primarily to differences in duration. For example, while the “unstressed-slow” syllables are, in at least half the cases, roughly comparable in duration to the “fast-stressed” syllables, they are nonetheless considerably reduced in fundamental frequency, and somewhat reduced in overall amplitude and vowel color with respect to their “fast-stressed” counterparts. These data seem to indicate that while speaking rate and lexical stress both affect the duration of vowel segments, they have different effects on several acoustic parameters, and are probably independently controlled by different physiological mechanisms.

III. DISCUSSION

The results of this experiment show that differences in vowel duration due to changes in speaking rate do not seem to have a substantial effect on the attainment of acoustic vowel targets. The formant frequencies of these presumed targets remained essentially unchanged across changes in speaking rate. It was also shown that the probable mechanism by which these targets were achieved was an earlier onset of the transition movement from consonant to vowel. The speed of movement from the consonant to the vowel, however, did not seem to change. While these patterns were consistent across all five speakers, they were observed for only a small number of phonetic samples produced by phonetically trained speakers in a precise manner. Further, the present results might also be affected by several additional factors that were not studied in the present experiment. One such complicating factor might be differences in phonetic context. For example, the effect of an increase in speaking rate on transition movements might be different depending on whether the movement is from an alveolar or labial consonant. Likewise, because vowel duration is conditioned by factors other than speaking rate, differences in speech material, phonetic context, and word position in a sentence (for example, Klatt, 1976), shorter segment durations associated with one of these factors might produce a different pattern of CV transition movement.

Differences in overall duration might account for the differences between the present results and the articulatory data of Kuehn and Moll (1978). Kuehn and Moll showed that different speakers can use different strategies to control speaking rate, with one such observed strategy being an adjustment of articulatory velocity. This obvious inconsistency between the two sets of data might be related to differences in corresponding across-rate durations. In the present experiment, transition durations for fast speech were approximately 90% of those for slow speech, while the corresponding fast speech durations measured by Kuehn and Moll were on the order of 50% of those for slow speech. Thus, it might be suggested that if changes in articulatory velocity (and corresponding transition rate of change) appear, they might do so primarily at very fast rates of speech.

The present acoustic data are also inconsistent with earlier EMG data (Gay et al., 1974; Gay and Ushijima, 1975) that showed a change in the level of muscle activity for vowels in response to a change in speaking rate. The EMG data showed that the activity levels of the genioglossus muscle for the vowel /a/ decreased with an increase in speaking rate. The genioglossus is a prime mover of the tongue and is active during, and probably responsible for, the bunched and protruding movement of the tongue for /a/. The decrease in activity implies either, or a combination of both, a decrease in articulatory displacement or a decrease in the speed of articulatory movement. However, the present acoustic data
show that neither of these parameters (as reflected in the acoustic signal) are substantially affected. The different interpretations that arise from the physiological and acoustic data might be explained in a number of ways, none of which seem entirely satisfactory. First, the reduction in EMG activity might not reflect a corresponding difference in articulatory displacement, that is, undershoot at the muscle contraction level might not produce undershoot at the articulator movement or acoustic level. Second, a totally different motor strategy using different muscles in different ways might come into play during fast speech. Third, the peak of the integrated EMG envelope might not provide an accurate indication of the maximum strength contraction of an active muscle when the duration of the muscle contraction is charged. Changes in the peak of the EMG envelope are usually interpreted as reflecting (without being able to separate) changes in either the displacement of the articulator that the muscle is acting directly upon, or changes in the speed of movement of that articulator. However, it is also possible that the summed potentials of the integrated signal might peak differently if the duration of the contraction changes. Thus, with all other parameters held constant, a reduction in the peak of the integrated EMG envelope might also reflect simply a reduction in the contraction time of that muscle.

The inconsistencies between the physiological and acoustic data aside, it would appear from the data of this experiment that the coordination of articulatory movements is adjusted in some way in order to preserve the information bearing elements of segmental units across changes in speaking rate. The reduction in duration of all segments (Ref. Table I) coupled with the relative constancy of acoustic (vowel) targets, suggest that this adjustment involves primarily a horizontal compression along the time dimension. This type of compression, the existence of which was suggested some 30 years ago (Jongs, 1948), is a nonlinear one, and one that causes both a decrease of duration within segments and an increase in coarticulation between segments. It also appears that temporal restructuring for changes in rate is superimposed on the basic serial ordering process.

The control of, and effects of changes in, speaking rate and lexical stress seem to be different in a number of ways. The data of this experiment show that for stressed vowels, only duration is reduced to any substantial degree. However, de-stressed vowels, even if they are of the same duration as quickly produced stressed vowels, are reduced in overall amplitude, fundamental frequency, and to some extent, vowel color.

The finding that a de-stressed vowel was not substantially reduced in color toward the neutral schwa does not completely coincide with either Lindblom's (1963) acoustic data or Harris' (1975) EMG data. These differences are probably due to the fact that in the present experiment, speakers were explicitly instructed to maintain the phonetic identity of the vowel during de-stressing. It might be suggested that if extended stress contrasts were studied in the present experiment, greater degree of vowel reduction might have been observed. While differences in the degree of vowel reduction between Lindblom's (1963) findings and the present data can be explained, the question of the relationship between reduction and duration is more difficult to resolve. Because vowel reduction appeared for changes in both stress and speaking rate in Lindblom's data, he concluded that reduction (and undershoot) was caused solely by changes in duration, and not the suprasegmental features of stress and rate, for example. However, the present data lead to the opposite conclusion. Because the tendency for formant frequencies to be reduced toward the neutral schwa occurs only for an un-stressed vowel, even if it is of the same duration as its stressed counterpart, the present data suggest that the degree of reduction is linked to stress, regardless of the relative or absolute duration of the segment. The suggestion that stress, and not duration, determines target attainment has also been put forth by both Harris (1975) and Nord (1975).

In this experiment, it was also shown that de-stressing affected the fundamental frequency and overall amplitude of the vowel, indeed, even more so than the formant structure. These effects, which are consistent with those described by Fry (1955) and Lieberman (1960) among others, are compatible with an "extra effort" model of stress, such as the one proposed by Ohman (1967). A reduction of overall articulatory effort can result in corresponding reductions in the four parameters measured in this experiment: fundamental frequency, overall amplitude, duration, and vowel color. Thus, the findings of this experiment suggest that two separate and independent physiological mechanisms control changes in speaking rate and lexical stress, one that horizontally compresses the string and the other that modulates overall articulatory effort. A change in duration is a deliberate strategy of the first, while only a consequence of the second.

ACKNOWLEDGMENTS

This research was carried out while the author was on leave at the Department of Speech Communication, Royal Institute of Technology, and Department of Linguistics, Stockholm University, Stockholm, Sweden. The comments and suggestions of Gunnar Fant and Bjorn Lindblom of these institutions are gratefully acknowledged. This research was supported, in part, by grants from the National Science Foundation (BN-7516894) and the National Institute of Neurological and Communicative Disorders and Stroke (NS-10424).


