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 CVCVC utterances at slow and fast speaking rates. Spectrographic measurements showed that the midpoint format 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, destressed 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.Gr, 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 suggest the latter. In a series of experiments (Gay and Hirose, 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 (KH), all native speakers of American English. Two (TG, LR) spoke a New York dialect, one (KH) 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, /iɪɛæɑɔʊuʌ/, in a /p_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, /i a u/ in a /b_p/ environment, and a third consisted of CVCVC sequencies of the type, /kipap´/, /ki´pap/, /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 16 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.

 0001-4966/78/6301-0223$00.80 ©1978 Acoustical Society of America

224

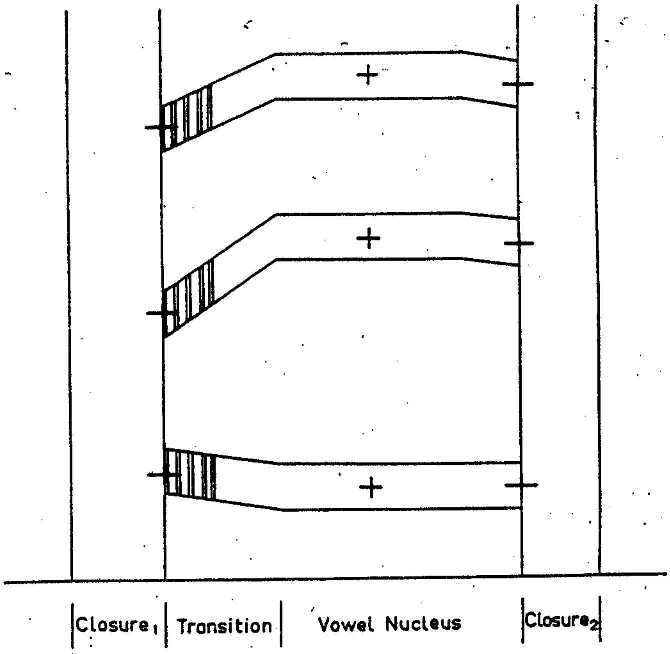

FIG. 1. Schematic illustration indicating points where duration and formant frequency measurements were made.

B. Data recording

The original protocol called for three different speaking rates to be used: one slow, or normal, and two fast. However, none of the speakers could comfortably or reliably produce speech at two different "fast" rates. An additional rate, slower than that of the normal one, was also considered, but it, too, was unnatural and resulted in obviously contrived renditions. Apparently, these four speakers, at least, have two natural rate modes, one normal (or slow) and one fast, with rates outside those two being difficult to control or maintain. Thus, only two rates were studied, both determined by each speaker's own judgment of natural slow and fast rates. Each speaker read each of the five utterance lists through, first at the slow rate and then at the fast rate. For the CVC utterances, subjects were instructed to place sentence stress on the test word. For the stress contrasts, the speakers were instructed to destress the appropriate syllable while still maintaining its phonetic identity, in other words, not to the point where the vowel would be reduced in color to a schwa. All subjects received detailed instructions about the tasks before the recording session began, and had ample practice time with the utterances. All recordings were made in a sound-treated room.

C. Data analysis

A total of 640 utterances were analyzed (16 samples ×5 repetitions×2 rates×4 speakers). Spectrograms were made for each utterance on a Voice-print Laboratories Sound Spectrograph, using the extended frequency scale, wide-band filter (300 Hz), and highshaping setting. In addition, fundamental frequency and overall amplitude measurements were made for the stress constrasts using a specially modified pen writing oscillograph (Mingograf, 34T). Duration and formant frequency 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 ($F1$, $F2$, $F3$) 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 $F2$ transition reached a steady state, (2) if a steady state was not reached, the point where the $F2$ 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 /ɔ/, consistently, and /ɪ/ and /ɛ/, 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 $F3$ did not always appear clearly for speakers TG and LR, and for most samples of the destressed syllables. As would be expected, CV movements for $F1$ were small, while those for $F2$ provided the most reliable and useful transition movement information. Repeated measurements of selected samples revealed fairly consistent error ranges. Duration measurements were accurate to ±10 ms, while formant frequency measurements were accurate to ±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 (±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.

	Consonant closure (1)		Vowel		Consonant closure (2)	
	Duration	Ratio	Duration	Ratio	Duration	Ratio
i	100 95	0.95	120 90	0.75	80 70	0.88
I	105 95	0.90	105 85	0.81	80 75	0.94
ɛ	105 95	0.90	130 105	0.81	90 80	0.89
æ	105 95	0.90	155 125	0.81	80 70	0.88
a	100 90	0.90	145 115	0.79	80 75	0.94
ɔ	105 100	0.95	165 130	0.79	80 70	0.88
ʊ	105 100	0.95	110 90	0.82	85 75	0.88
u	100 95	0.95	120 90	0.75	80 65	0.81
ʌ	100 90	0.90	115 85	0.74	80 75	0.94
Mean	105 95		130 100		80 75	

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 /ɪ/ to 165 ms for /ɔ/ at the slow rate, the percentage change did not vary as a function of duration. For example, the phonetically long vowels /æ/ and /ɔ/ were not reduced to any greater or lesser percent during fast speech than the phonetically short vowels, /ɪ/ or /ʊ/. 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. Prestressed initial /p/ is consistently, and usually substantially, longer than poststressed 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.

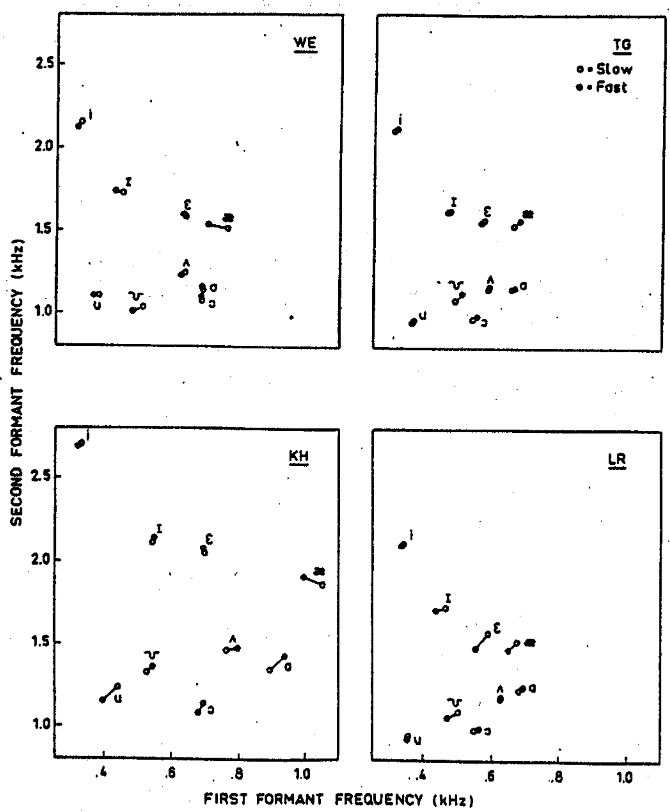

FIG. 2. $F1$–$F2$ vowel space for midpoint measurements for both speaking rates.

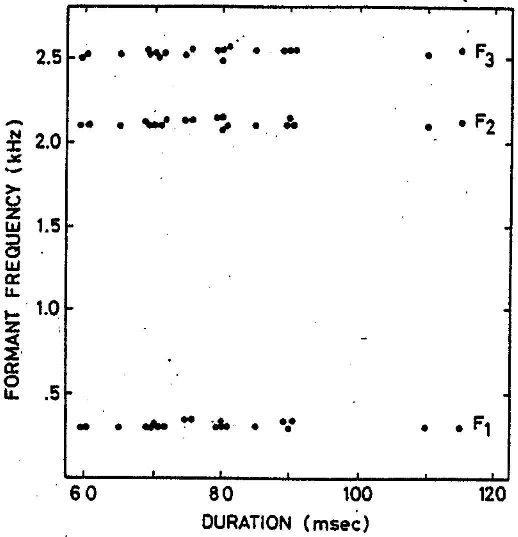

FIG. 3. Midpoint formant frequencies for $F1$, $F2$, and $F3$, as a function of duration, for the vowel /i/ (speaker TG).

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 /æ/ 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 /ɛ/ 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 /æ/ 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 /i/. Figure 3 shows the first, second, and third formant frequencies (measured at the midpoint) for /i/ 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 /i/ 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 measures 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 /i/, pooled over utterance repetitions.

Speaker (S/F)	Duration			Frequency		
	Closure	Transition	Vowel	$F2$ onset	$F2$ midpoint	$F2$ rate
WE	95	45	115	1925	2150	4.5
	80	40	95	1965[a]	2125	4.3
TG	95	50	95	1765	2105	7.7
	90	50	75	1845[c]	2115	6.9
KH	105	55	140	2230	2700	8.8
	110	45	105	2300	2685	8.9
LR	110	50	130	1735	2100	7.7
	105	40	95	1780[b]	2115	8.4

[a]Significant at 0.10.
[b]Significant at 0.05.
[c]Significant at 0.01.

TABLE III. Duration and formant frequency measurements for voiced–voiceless contrasts. Values are pooled over all repetitions for the three male speakers.

Utterance (S/F)	Closure duration	Transition duration	F2 onset	F2 midpoint	F2 rate
pip	100	50	1810	2120	6.6
	90	45	1885	2120	6.5
bip	105	45	1960	2120	3.6
	95	40	1985	2105	3.2
pap	95	50	1340	1170	2.1
	85	45	1325	1180	2.0
bap	105	...	1130	1150	...
	95	...	1160	1170	...
pup	95	40	1250	1000	3.6
	85	40	1200	990	3.9
bup	100	45	1065	990	0.8
	90	35	1030	990	1.2

A difference in second formant transition onset frequencies 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 /pup–bup/ 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, except perhaps to a slightly lesser extent. Again, for all three vowels preceded by /b/, the onset frequency of the second formant transition is closer to the midpoint frequency, while the midpoint frequencies, themselves, and the *F*2 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/, however, the shifts are more comparable.

More obvious than the rate differences are the differences in second formant transition onset frequencies and rates of change between /p/ and /b/, within each speaking rate condition. For all three vowels, it is apparent that the onset frequencies are considerably closer to the midpoint frequencies, and the *F*2 rate is correspondingly slower, for /b/ as opposed to /p/. All frequency differences 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 occurred, much earlier than the corresponding transition from /p/. The greater range of *F*2 onsets for /b/ as opposed to /p/ is consistent with Fant's (1969) calculations 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 it is for the more lax /b/; in other words, the tongue is more free to coarticulate with the following vowel during /b/ than it is during /p/. A second possibility 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 difference 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 toward the vowel target, the frequencies of the targets, themselves, do not appear to be affected, nor is the basic rate effect on the movements different. The rate effects seem to be superimposed on the context dependent articulatory movements, and are not affected by, or assimilated into, these movements.

The final set of measurements is related to the question 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 /i/ 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 utterance, 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 /i/ and /a/, 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.

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

228

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

Speaker (S/F)	Stressed					Unstressed				
	Duration	Rel. Amp.	F0	F1	F2	Duration	Rel. Amp.	F0	F1	F2
WE	150	4	140	665	1125	130	2	110	650	1150
	105	4	145	675	1150	85	0	115	640	1125
TG	120	3	110	675	1155	115	1	90	660	1190
	100	6	125	675	1165	100	0	95	625	1175
KH	145	5	230	910	1400	140	0	155	850	1450
	115	5	220	880	1380	95	1	165	800	1500
LR	155	4	140	665	1230	120	1	110	650	1240
	125	6	130	660	1250	100	0	110	600	1250

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 (1976). 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 /i/ 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 bunching and protruding movement of the tongue for /i/. 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 summated 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 (Joos, 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, destressed 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 destressed 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 destressing. It might be suggested that if extended stress contrasts were studied in the present experiment, a 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, *per se*. 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 unstressed 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 destressing 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 Öhman (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 (BNS-7616954) and the National Institute of Neurological and Communicative Disorders and Stroke (NS-10424).

Brown, W. S., and McGlone, R. (1974). "Aerodynamic and acoustic study of stress in sentence productions," J. Acoust. Soc. Am. 56, 971–974.

Fant, G. (1969). "Stops in CV syllables," Speech Transmission Laboratory, QPSR 4/1969, R.I.T., 1–25.

Fant, G., Ishizaka, K., Lindqvist, J., and Sundberg, J. (1972). "Subglottal formants," Speech Transmission Laboratory, QPSR 1/1972, R.I.T., 1–13.

Fry, D. (1955). "Duration and intensity as physical correlates of linguistic stress," J. Acoust. Soc. Am. 27, 765–768.

Gay, T., and Hirose, H. (1973). "Effect of speaking rate on labial consonant production: A combined electromyographic high-speed motion picture study," Phonetica 27, 203–213.

Gay, T., and Ushijima, T. (1975). "Effect of speaking rate on stop consonant-vowel articulation," Proc. Speech Communication Seminar, 1974, edited by G. Fant (Almqvist and Wiksell, Stockholm), pp. 205–209.

Gay, T., Ushijima, T., Hirose, H., and Cooper, F. S. (1974). "Effect of speaking rate on labial consonant-vowel articulation," J. Phonetics 2, 47–63.

Harris, K. S. (1975). "Mechanisms of duration change," Proc. Speech Communication Seminar, 1974, edited by G. Fant (Almqvist and Wiksell, Stockholm), pp. 299–305.

Joos, M. (1948). "Acoustic Phonetics," Language, Suppl. 136 pp.

Kent, R., and Moll, K. (1969). "Vocal-tract characteristics of the stop cognates," J. Acoust. Soc. Am. 46, 1549–1555.

Klatt, D. H. (1975). "Vowel lengthening is syntactically determined in a connected discourse," J. Phonetics 3, 129–140.

Klatt, D. H. (1976). "Linguistic uses of segmental duration in English," J. Acoust. Soc. Am. 59, 1208–1221.

Kuehn, D., and Moll, K. (1976). "A cineradiographic study of VC and CV articulatory velocities," J. Phonetics 4, 303–320.

Lehiste, I., and Peterson, G. E. (1969). "Transitions, glides, and diphthongs," J. Acoust. Soc. Am. 33, 268–277.

Lieberman, P. (1960). "Some acoustic correlates of word stress in American English," J. Acoust. Soc. Am. 32, 451–454.

Lindblom, B. (1961). "Accuracy and limitations of Sonagraph measurements," Proc. Fourth Int. Congr. Phonetic Sci. (Mouton, s'Gravenhage), pp. 208–213.

Lindblom, B. (1963). "Spectrographic study of vowel reduction," J. Acoust. Soc. Am. 35, 1773–1781.

Nord, L. (1975). "Vowel reduction: Centralization or contextual assimilation?" Proc. Speech Communication Seminar, 1974, edited by G. Fant (Almqvist and Wiksell, Stockholm), pp. 149–154.

Öhman, S. (1966). "Coarticulation in VCV utterances: Spectrographic measurements," J. Acoust. Soc. Am. 39, 151–168.

Öhman, S. (1965). "Durations of formant transitions," Speech Transmission Laboratory, QPSR 1/1965, R.I.T., 10–14.

Öhman, S. (1967). "Word and sentence intonation: A quantitative model," Speech Transmission Laboratory QPSR 2-3/1967, R.I.T., 20–54.

Peterson, G. E., and Lehiste, I. (1960). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. 32, 693–703.