

Formant movements of Dutch vowels in a text, read at normal and fast rate

R. J. J. H. Van Son and Louis C. W. Pols

Institute of Phonetic Sciences, University of Amsterdam, Herengracht 338, 1016 CG Amsterdam, The Netherlands

(Received 13 December 1990; accepted for publication 24 February 1992)

Speaking rate in general, and vowel duration more specifically, is thought to affect the dynamic structure of vowel formant tracks. To test this, a single, professional speaker read a long text at two different speaking rates, fast and normal. The present project investigated the extent to which the first and second formant tracks of eight Dutch vowels varied under the two different speaking rate conditions. A total of 549 pairs of vowel realizations from various contexts were selected for analysis. The formant track shape was assessed on a point-by-point basis, using 16 samples at the same relative positions in the vowels. Differences in speech rate only resulted in a uniform change in $F1$ frequency. Within each speaking rate, there was only evidence of a weak leveling off of the $F1$ tracks of the open vowels /a a/ with shorter durations. When considering sentence stress or vowel realizations from a more uniform, alveolar-vowel-alveolar context, these same conclusions were reached. These results indicate a much more active adaptation to speaking rate than implied by the target undershoot model.

PACS numbers: 43.70.Bk, 43.70.Fq, 43.72.Ar

INTRODUCTION

In the target undershoot model of vowel articulation, vowel duration is considered an important parameter in determining the actual realization of the vowel formants (e.g., Lindblom, 1963, 1983; Broad and Fertig, 1970; Gay, 1978; Gay, 1981; Broad and Clermont, 1987; Di Benedetto, 1989; Lindblom and Moon, 1988; Moon, 1990). Vowel duration is important both for the formant frequency inside the vowel nucleus (for use of "vowel nucleus," see Krull, 1989) as well as for the shape of the complete formant tracks. The target undershoot model predicts more spectral reduction when vowels become shorter, i.e., more schwa-like formant values in the vowel nucleus and more level, i.e., less curved, formant tracks.

Inside the vowel nucleus, the formant frequencies appeared to be correlated to vowel duration in the way predicted by the target undershoot model, at least when speaking style was held constant (Broad and Fertig, 1970; Broad and Clermont, 1987; Lindblom and Moon, 1988; Moon, 1990). In contrast, formant frequencies were only weakly correlated to vowel duration, or not at all, when the speaking style differed (e.g., clear speech versus citation form speech, Lindblom and Moon, 1988; Moon, 1990; fast rate speech versus normal rate speech, Van Son and Pols, 1990). Several studies did not find speaking-rate-dependent differences between formant frequencies that were in any way connected to vowel identity (e.g., Gay, 1978; Gopal and Syrdal, 1988; Den Os, 1988; Engstrand, 1988; Van Son and Pols, 1990; Fourakis, 1991). Van Son and Pols (1990) did find a systematic higher $F1$ in fast rate speech, but this difference occurred in all vowels (even /a/). This rise in $F1$ cannot be interpreted as vowel reduction in the sense of the target un-

dershoot model. These studies suggest that there are two kinds of durational differences between vowel realizations. The first type of durational differences are those found between vowels spoken in the same speaking style and at the same rate. These differences in vowel duration are (cor-) related to spectral differences as predicted by the target undershoot model. The other type of durational differences are the differences between vowels spoken in different speaking styles or at different rates. These latter differences in duration are not related to spectral differences between vowels.

Relatively few studies have considered the relation between vowel formant dynamics and duration (e.g., Broad and Fertig, 1970; Broad and Clermont, 1987; Di Benedetto, 1989; Van Son and Pols, 1989) and these were limited to only one speaking style. Studies that did use different speaking styles or different speaking rates generally only measured formant frequencies within the vowel nucleus. Therefore, it is not clear whether fast-rate speech is just "speeded-up" normal-rate speech, or whether different articulation strategies (as proposed by Gay, 1981) or a higher speaking effort (Lindblom, 1983) are used. Differences in articulation or speaking effort should result in different shapes of the formant tracks, e.g., a levelling-off of the formant movements in fast rate speech.

Formant track shape is generally characterized by the lengths and slopes of vowel on- and off-glide which are measured using two to four points from each formant track (Di Benedetto, 1989; Strange, 1989a,b; Duez, 1989; Krull, 1989). However, it is very difficult to determine the boundaries of the stationary part (Benguerel and McFadden, 1989) and to measure formant track slopes accurately. Therefore, another method to characterize formant track shapes was chosen. We performed a point-by-point analysis

on sampled vowel formant tracks (16 points, adapted from Broad and Fertig, 1970) and compared the formant frequencies on comparable, relative, positions in the vowel realizations.

Differences between speaking rates are best studied by using vowel realizations that differ *only* in speaking rate. In order to obtain a large and varied inventory of such vowel pairs, a long text was read twice by a single professional speaker (a well-known newscaster), once at a normal rate and once at a fast rate (Van Son and Pols, 1990). With these vowels, we have tested whether vowel formant track shape depends on vowel duration and speaking rate and how this relation can be modelled. Also the effects of stress and vowel context were taken into account.

Using a single, professional speaker will make it difficult to generalize the results of this study to other, more "naive," speakers. However, the way an experienced newscaster, who speaks standard Dutch and whose pronunciation is perceived as "correct," reacts to speaking rate differences will be very likely an "accepted" way of doing so. General theories of articulation do not consider personal skill or experience as a factor of importance. Therefore, when our speaker does not utter vowels in the way predicted then we have, for nonaberrant speech, a counter example to the general theories of articulation. We do acknowledge that large sections of the population might react in a different way to speaking rate changes. Our experiment should be viewed only as a test on the predictive power of articulation theories on the effects of speaking rate.

I. METHODS

The present project investigated a subset of the material used in our previous study (Van Son and Pols, 1990). Here, we will only summarize the procedures used.

A. Speech material and segmentation

A meaningful text of 844 words (1440 syllables) was read twice by an experienced speaker, once as fast as possible, once at a normal rate (i.e., as for an audience). The speech was recorded on a commercial Sony PCM-recorder, low-pass filtered at 4.5 kHz and digitized at 10 kHz, with 12-bit resolution. Subsequent storage, handling, and editing were done in digital form only. Reading the text took 330 s for the normal speaking rate and 220 s for the fast speaking rate (4.4 and 6.6 syll./s including pauses). The overall reduction in duration of the fast rate as compared to the normal-rate realization was one-third when pauses longer than 200 ms were included, and one-fourth when these longer pauses were excluded. A subjective evaluation did not reveal differences in reading style between speaking rates.

Based on the orthographic form of the original text, we selected putative realizations of the vowels we wanted to study. These vowel realizations were localized in the speech recordings and the segment boundaries were placed with the help of a visual display of the waveform and auditory feedback. The vowel boundaries were chosen at a zero crossing in the speech waveform. A whole number of pitch periods was used. Any pitch period that could be attributed to the target

vowel, and not to the neighboring phonemes, was considered to be part of that vowel realization. The segments were copied with a leading and trailing edge of 50 ms of speech. Vowel realizations that could not be separated from their context with confidence were not used, contrary to what was done in Van Son and Pols (1990). The tokens were labeled for sentence accent and actual phoneme realization. Stress and phoneme labels at the two rates were not always identical but the differences between the speaking rates were not systematic.

B. Vowels used

Seven of the 12 Dutch monophthongs were used: /i y u o a ϵ /. These vowels were selected because of their rather high frequency of use in Dutch and their representativeness in the vowel space. Five of the vowels used are short or half-long vowels (/i y u o ϵ /) and two are long vowels (/o a/).

As a neutral "anchor" in the vowel space, a small number of realizations of the schwa was selected as well. These schwa realizations came from the words "HET" = / ət / (English: "THE") and "ER" = / ər / or / dər / (English: "THERE"). Some other vowels which were reduced to schwa, were included in this group of schwa vowels as well.

The various numbers of vowels thus obtained are listed in Table I. Out of 1178 isolated tokens, only equally paired tokens that could be segmented with confidence were used in this study, leaving 549 pairs of tokens.

To assess the importance of stress and vowel context, more homogeneous subsets of realizations of the vowels / ϵ a i o/ were selected from the total set of tokens and analyzed separately: We used tokens with and without sentence stress and those tokens that occurred in a CVC context in which both C's were alveolar consonants (i.e., one of /n t d s z l r/, Table I). Alveolar consonants can be considered as closed and fronted phonemes, from an articulatory viewpoint close to the vowel /i/. The target-undershoot model predicts the largest influence of duration when the articulatory distance between consonant and vowel is largest. Therefore, we

TABLE I. Number of vowel pairs matched on normal versus fast rate. Both tokens in a pair are from the same text item. Only pairs with comparable vowel realizations that could be reliably segmented are presented, 38 pairs from the original material were not used and are not included in this Table (see text). The schwa is never stressed. In the last column the number of tokens in an alveolar-vowel-alveolar context is added between parentheses for some vowels (Dutch alveolar consonants are /n t d s z l/, see text).

Vowel	Stressed	Unstressed	Unequal stress	Total
ϵ	23	85	12	120 (21)
a	23	79	8	110 (33)
a	21	70	11	102 (27)
i	23	57	4	84 (38)
o	17	56	11	84 (16)
ə	0	21	0	21
u	4	7	5	16
y	5	6	1	12
total	116	381	52	549 (135)

would expect the largest coarticulatory effects on the $F1$ tracks of the open vowels / ϵ α a / and the $F2$ tracks of the back vowel / o /. There were not enough tokens in another (nonalveolar) homogeneous context to merit analysis.

Of the three other vowels, there were too few stressed tokens or realizations in an alveolar context to enable analysis.

C. Spectral analysis and formant track sampling method

The vowel segments were analyzed with a 10-pole LPC analysis, using a 25.0-ms Hamming window, which shifted in 1-ms steps (Vogten, 1986). The formant analysis was based on the Split-Levinson algorithm, which gives continuous formant tracks (Willems, 1986).

The formant tracks obtained from the different vowels were sampled at 16 equidistant points, including both boundaries. The linear formant frequency, in Hz, was used. Two tokens (both / i /) were shorter than 16 ms and thus gave less than 16 different frames in a track. From these we doubled some frames to obtain the 16 desired values. Symmetry was preserved by the doubling.

II. RESULTS

The formant values and vowel durations were compared for the two speaking rates. Comparisons were done between pairs of tokens taken from readings of the same text items at different speaking rates.

All statistical tests are from Ferguson (1981), and all statistical tables from Abramowitz and Stegun (1965, p. 966–990). Correlation coefficients were recalculated to a Student t test to determine significance. To prevent repeated test results from containing spurious errors, a two-tailed threshold level for statistical significance of $p \leq 0.01\%$ was chosen for testing the point-by-point formant data (16 points per formant per vowel) and a threshold level of $p \leq 0.1\%$ was chosen for testing differences in duration (1 value per vowel). When the two speaking rates were tested in parallel, i.e., not pooled, only results that were statistically significant at both speaking rates were considered, because the methods used were not well qualified to distinguish between speaking rates.

A. Duration

Mean differences of duration between speaking rates were tested (Table II). As was to be expected, the fast rate tokens were shorter than the normal rate tokens. The difference was around 15% for all vowels combined, intrinsic long vowels (/a, o/) showed a shortening of around 20% at a higher speaking rate.

Mean duration was statistically significantly ($p \leq 0.1\%$) shorter for fast rate tokens than for normal rate tokens for the vowels / ϵ α a i o / (Table II). Realizations of the schwa did not differ in length between speaking rates. This could be explained by the fact that they were already extremely short. The vowels / u y / showed no significant differences, probably because of their small numbers (see Table I). From the results presented in Table II it was found that the mean dura-

TABLE II. Mean duration (in ms) of tokens for both speaking rates, and mean difference in duration between speaking rates. The mean duration of short vowels (/ ϵ α i u y /, all tokens pooled) was 86 ms (normal rate) and 76 ms (fast rate). The mean duration of long vowels (/a o/, all tokens pooled) was 128 ms (normal rate) and 104 ms (fast rate). Last column: Correlation coefficient of vowel duration between tokens of the same text item at both speaking rates. Statistical significance is tested with a Student's t test on difference. Correlation coefficients were recalculated to a Student's t test variable before testing. Statistical significant differences and correlation coefficients are underlined (level $p \leq 0.1\%$, last two columns), the others are not significant.

Vowel	Normal	Fast	Normal-fast	Corr. coeff.
ϵ	85	74	<u>11</u>	<u>0.78</u>
α	87	77	<u>10</u>	<u>0.74</u>
a	127	102	<u>26</u>	<u>0.79</u>
i	86	74	<u>13</u>	<u>0.64</u>
o	129	107	<u>23</u>	<u>0.80</u>
$\text{\textcircled{a}}$	56	54	2	-0.02
u	89	82	8	<u>0.89</u>
y	92	81	11	<u>0.86</u>
Total	99	84	<u>15</u>	<u>0.82</u>

tion of the long vowels ($V:$) was related to the mean duration of the short and half-long vowels (V , excluding / $\text{\textcircled{a}}$ /) as: $V: = a V - d$, in which "a" and "d" are speaking-style independent constants (Fant and Kruckenberg, 1989; Koopmans-van Beinum, 1990; they found $V: = 1.9 V - 45$ ms and $V: = 2.05 V - 38$ ms, respectively). As only two speaking conditions were available, the coefficient "a" could not be determined reliably from our data and was chosen to lie between the two published values, i.e., "a" = 2. The constant "d" was found to be 45 ms in normal rate speech and 47 ms in fast rate speech.

The correlation between vowel duration values of tokens spoken at normal and fast rate was significant for all vowels tested, except for the vowel / $\text{\textcircled{a}}$ /, and correlation coefficients were larger than 0.71 for all vowels except for the vowels / i $\text{\textcircled{a}}$ / (Table II). This meant that the within-speaking-rate variation in duration is preserved between different speaking rates. The lack of correlation between durations of the schwa at fast and normal rate, could possibly be attributed to the restricted contexts from which these tokens were extracted and the lack of differences between realizations at the two speaking rates.

B. Effects of speaking rate on formant frequencies

Speaking rate differences resulted in differences in vowel durations and probably also in formant values. Mean formant frequency differences between speaking rates proved to be rather small. In Fig. 1, the differences in formant values between speaking rates are displayed as the normal rate formant frequency subtracted from the corresponding fast rate formant value, so any deviation from a straight line at 0 value might be interesting. For each vowel, the differences between tokens spoken at different rates, corresponding to a

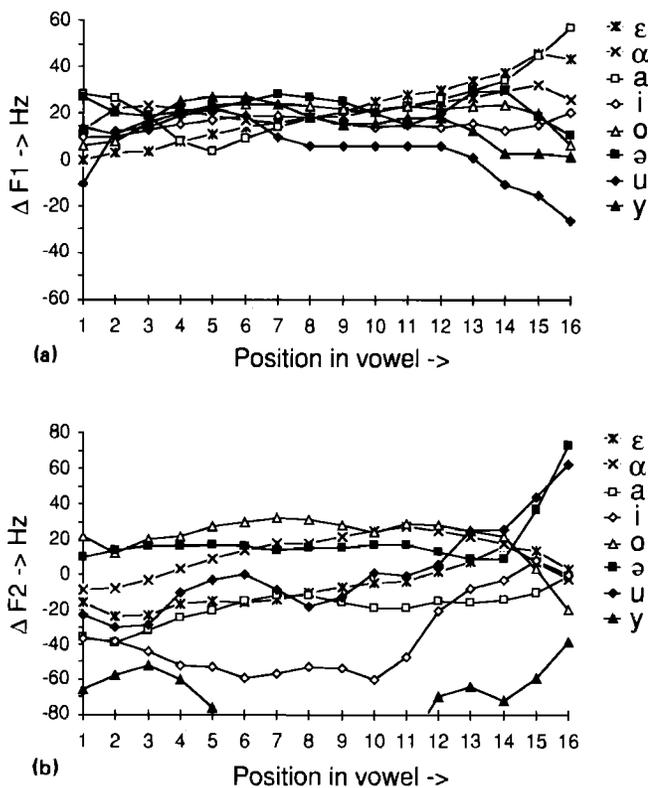


FIG. 1. Mean differences in formant frequency values in Hz (fast-rate value minus normal-rate value) for all 16 points within the vowels. Statistical significance is determined by a Student *t* test on difference ($p < 0.01\%$). (a) First formant (*F1*). The differences are significant at the points / ϵ /: 7-16; / α /: 2-15; / a /: 2, 8-16; / i /: 3-8; / o /: 3-14. (b) Second formant (*F2*). The differences are significant at the points / a /: 10-12.

certain point in the token (points 1 through 16), were averaged and the statistical significance was determined by a Student's *t* test on the difference. Statistical significance for individual points was indicated in the legend of Fig. 1.

For *F1*, the differences were statistically significant in more than half of the vowel segment (more than 8 points) for the vowels / ϵ α a o / and in less than half of the vowel segment in / i / [see Fig. 1(a)]. The differences in *F1* were small, on the average 20 Hz. The parts showing significant differences did not correspond to a certain position within the vowel. Thus fast-rate tokens showed a slightly higher *F1* value than normal-rate tokens in all parts of the vowel, irrespective of vowel identity.

Despite quite large differences between mean *F2* values [Fig. 1(b)], statistically significant differences were only found in a small part in the second half of / a /. Thus no consistent differences in frequency were found between *F2* values from vowels spoken at a fast rate as compared with those spoken at a normal rate. This result suggests that there were no large, systematic effects of speaking rate on the shape of the second formant track.

C. Correlation between speaking rates

The two readings resulted in two correlated sets of formant measurements. The context of each text item was identical in both readings so the formant frequency values measured in tokens of the same text item at different speaking

rates might very well be correlated. The correlation coefficient over pairs of tokens of the same vowel is then a measure of the amount of context dependent variance captured with the measurements (see also Van Son and Pols, 1990). These correlations were calculated for each point in the vowels and the resulting correlation coefficients were plotted in Fig. 2.

The values measured at both speaking rates from the same text item, indeed showed high correlation coefficients. The correlations were statistically significant for *F1* in all parts of the vowels / ϵ α a i o / [Fig. 2(a)]. For *F1*, the correlation coefficients surpassed 0.71 (more than 50% of variance explained) in most parts of the vowels / α o / and were larger than 0.5 (more than 25% of variance explained) in the vowels / ϵ a i /, i.e., in those vowels that showed significant correlations ($p < 0.01\%$) between *F1* values. The vowels / ϵ u y / did not show significant correlations between speaking rates, despite some fairly high correlation coefficients (e.g., for / y / tokens).

For the second formant [*F2*, see Fig. 2(b)], the tokens of / ϵ α a o ϵ / showed significant correlations between speaking rates ($p < 0.01\%$) in all or most parts of the vowels, the vowels / i u y / only in small parts. Except for the vowel / i /, the values of the statistically significant correlation coefficients were almost all above 0.71 and thus explained more than half of the variance in most parts of the vowels. Note

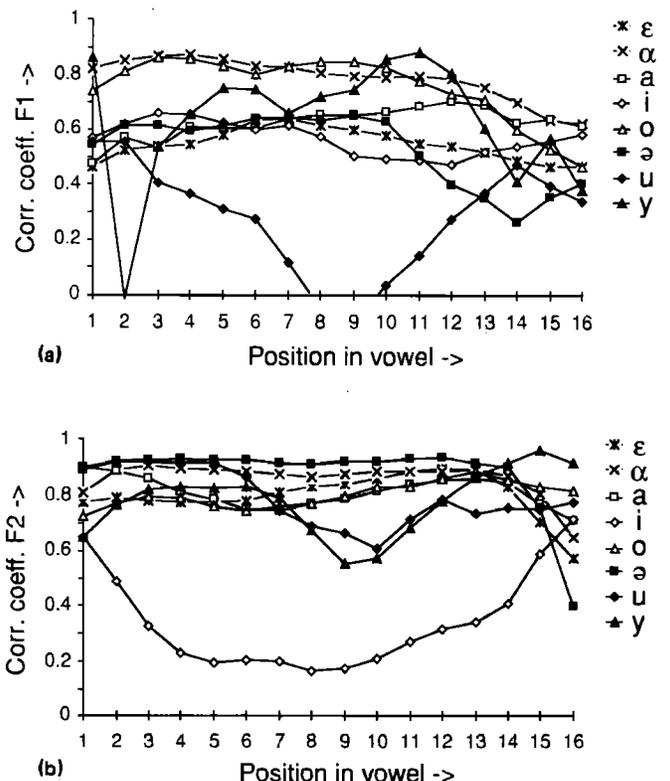


FIG. 2. Correlation coefficients between formant frequency values measured in fast-rate tokens and the corresponding values measured in normal-rate tokens for all 16 points within the vowels. Statistical significance is determined by recalculating the correlation coefficients to a Student *t* test ($p < 0.01\%$). (a) First formant (*F1*). The correlations are significant at all 16 points within the vowels / ϵ α a i o /. (b) Second formant (*F2*). The correlations are significant at all 16 points within the vowels / ϵ α a o /, and at the points / i /: 1, 2, 15, 16; / ϵ /: 1-15; / u /: 1-6; / y /: 14-16.

that the correlation coefficients between the formant values of vowels spoken at normal and fast rate (Fig. 2) were often larger than the corresponding correlation coefficients between vowel durations (Table II).

These results indicate that a large fraction of the variation in vowel formant values within each speaking rate was indeed systematic and reproduced when the text was reread.

D. Effects of duration on formant frequencies

Because durations differed between speaking rates (cf. Sec. II A) and $F2$ values did not seem to (cf. Sec. II B), it would not have been prudent to pool tokens from both speaking rates to calculate correlation coefficients between vowel duration and vowel formant frequency. Therefore, correlation coefficients between formant values and vowel durations were calculated for each speaking rate independently (not shown). The strength of the correlation between formant frequency values and vowel duration denotes the importance of the duration in determining vowel formant frequency (and vice versa). The stronger the relation between formant frequencies and vowel duration, the higher the absolute value of the correlation coefficient between both values. It must be remembered that a lot of variance could be explained due to the strong correlation between speaking rates, both for duration (Sec. II A) and formant values (Sec. II B).

The correlation coefficient values between $F1$ frequency and vowel duration generally were positive in the center and smaller or negative in the on- and offglide for the open vowels / ϵ α a / (not shown). This means that realizations of these high $F1$ vowels that have a longer duration also have higher $F1$ frequencies in the center and equal or lower $F1$ frequencies in the on- and offglide part. This can also be described as a decrease in the difference between the center and the on- and offset frequencies of the $F1$ track with a decrease of duration. This indicates a leveling of the formant track with a shorter duration. However, significant correlation strengths between $F1$ values and duration were reached for both normal rate and fast rate for the vowels / α a / only (not shown), and there only in a small part (2–8 points) in the center of the vowels. Only in fast-rate tokens of the vowel / a / did the correlation coefficient surpass 0.5, but then for three sample points only ($|r| \leq 0.55$). This indicated that the amount of variance explained this way (i.e., less than 25%) was small but could still be of importance.

For $F2$, none of the vowels showed a statistically significant correlation between formant values and vowel duration for both speaking rates (not shown). There was no measurable relation between vowel duration and $F2$ frequency values.

E. Effects of context

The tokens of the vowels / ϵ α a i o / in an all alveolar CVC context (C is one of / n t d s r l /) were also analyzed. The number of tokens per vowel available in an alveolar context was quite small ($n = 16$ – 38 , Table I). For small numbers, the estimated parameter values will have a large

error. Therefore, we concentrated on the relation between the tokens in the subset and those of the parent set and not on the actual sizes of the differences between the two sets. For this analysis, a threshold level of significance of $p \leq 0.1\%$, reached at two or more points within a vowel, was sufficient.

The fast-rate tokens of this subset had a uniform higher mean $F1$ frequency than the normal-rate tokens but the difference was not statistically significant ($p > 0.1\%$ at all points). The between-speaking-rate correlation coefficients of the formant frequencies were high for both $F1$ and $F2$, often higher than those for the parent set. The trends were the same as in the parent set of tokens.

The correlation coefficients between formant frequencies and vowel duration were generally higher in the subset of tokens in alveolar context than in the parent set, especially for $F1$ of / α a /. Still, only a few correlation coefficients were statistically significant ($F1$ in the center of / α /, $p \leq 0.1\%$ for more than 2 points).

These results show that the tokens from the subset of vowels in alveolar context were not different from the complete parent set of vowel tokens.

F. Effects of stress

The previous analyses were repeated on token pairs of the vowels / ϵ α a i o / for which both tokens were stressed or unstressed (data not shown). This was done to check whether sentence stress might be significant with respect to the effects of differences in speaking rate or duration.

Stressed tokens were 30% longer than the unstressed ones for both speaking rates ($p \leq 0.1\%$). The differences in vowel duration between speaking rates were comparable for stressed and unstressed tokens (i.e., 15%). The mean duration of the long vowels (\bar{V}) was related to that of the short vowels (\bar{v}) as $\bar{V} \approx 2\bar{v} - 54$ ms in stressed tokens and $\bar{V} \approx 2\bar{v} - 43$ ms in unstressed tokens (cf. Sec. II A).

For the $F1$, formant frequencies of the stressed tokens were generally higher than those of the unstressed tokens at both rates. This difference was largest for the high $F1$ -target vowels ($p \leq 0.01\%$ in the center of / α / for both speaking rates). The vowel space of the stressed tokens was larger, i.e., less reduced, in the $F1$ direction (/i/ to /a/) than that of the unstressed tokens. There was no indication that, compared to stressed tokens, unstressed tokens are spectrally reduced with respect to the $F2$. The fast rate stressed and unstressed tokens had a uniform higher $F1$ than the normal rate tokens. For unstressed tokens the difference was statistically significant ($p \leq 0.01\%$). For stressed tokens the difference was smaller than for unstressed tokens and not statistically significant ($p > 0.1\%$).

Correlation coefficients between speaking rates were higher in stressed tokens than in unstressed tokens and statistically significant for both ($p \leq 0.01\%$). The reverse was found for the correlation between formant values and vowel duration. For both stressed and unstressed tokens the correlation between formant values and vowel duration was never statistically significant ($p > 0.1\%$) for both speaking rates. As far as could be checked, the results obtained from all tokens pooled were equally valid for both subsets of tokens individually.

III. DISCUSSION

The results presented in this paper were obtained from sampled formant tracks, analyzing the formant frequency samples in a point-by-point way. Especially when correlated changes in formant frequency occur between different parts of the formant tracks, this method might be not sensitive enough. Therefore, we also modeled the formant tracks of all vowel realizations with polynomials (using Legendre polynomials, see Nossair and Zahorian, 1991 for an example of how these can be used). These new, polynomial, formant track parameters were analyzed in the same way as the point-by-point data. On all accounts, the results of the analysis using polynomial parameters appeared to be a duplication of those obtained from the point-by-point data. Therefore, we did not include them in this paper (results are available upon request).

A. Effects of speaking rate

The difference in vowel duration between tokens spoken at normal and fast rate was small but consistent. In fact, the difference was only half of what would have been expected from the overall difference in duration of both readings, which was 25% (see Sec. I A). For both readings the mean duration of long vowels (\bar{V}) was twice the mean duration of short vowels (\bar{V}) minus a constant, i.e., $\bar{V}: \approx 2\bar{V} - 46$ ms. From this relation it follows that the absolute difference in vowel duration between speaking rates should have been approximately twice as large for long vowels than for short vowels. But this relation does not explain why the overall differences were so small. A possible explanation could be that vowels are more resistant to durational compression than other phonemes. Indeed, this was found by Eefting (1991) using the same speaker.

In other studies, larger differences in vowel duration were found between speaking styles and rates (e.g., Lindblom and Moon, 1988) than in the present study. These studies used speech which contained longer vowel realizations than did our speech material. Starting with (much) shorter vowel realizations from a long read text, the small reductions in vowel duration found in this study were likely to strain the articulatory capabilities of our speaker more than did the much larger reductions of vowel duration in studies which used isolated words or sentences. As the articulatory models discussed before emphasize articulatory effort as an important factor influencing vowel formant tracks, even this relatively small reduction should have had a measurable effect on vowel formant tracks.

Despite the fact that the fast-rate vowel realizations are generally (and consistently) shorter than the normal-rate realizations, there is hardly a difference between the formant frequency values measured at different speaking rates. This means that a difference in speaking rate did not result in systematic differences in formant values. Only the $F1$ frequency is higher in vowels spoken at a fast rate compared to vowels spoken at a normal rate. This rate-dependent rise in $F1$ frequency was present irrespective of vowel identity and it was uniform (independent of the position inside the vowel). This means that the equivalent results found by Van Son

and Pols (1990) for vowel nucleus measurements cannot be attributed to a change in formant track shape due to speaking rate. It also indicates that our speaker increased articulation speed when he spoke faster. This increase in articulation speed matched the decrease in vowel duration.

B. Effects of duration on formant tracks

A simple, one-way, relation between vowel formant tracks and vowel duration would result in a clear-cut, and strong, correlation between these two. However, correlation coefficients between formant frequencies and vowel duration were only significant for the $F1$ tracks of the high $F1$ target vowels ($/a a/$). The correlations implied a leveling off of the $F1$ tracks with shorter durations of the tokens. This is predicted by the target-undershoot model. However, the correlation coefficients were rather small in all cases. The correlation between formant frequency and vowel duration hardly explains more than 30% of the variance in formant frequencies ($|r| \leq 0.55$, Sec. II C). Between-speaking-rate correlations for these three vowels, which measure the context dependent variation captured by the measurements, sometimes explained up to 70% of the variance in $F1$ formant frequencies [$|r| \leq 0.85$, Fig. 2(a)]. This difference in correlation indicated that duration is not a major determinant of overall vowel formant track shape in read speech.

$F2$ formant tracks do not show any sizeable correlation between formant track frequency and vowel duration.

C. Effects of context and stress

The context in which a vowel is spoken might be of importance for changes in speaking rate (or changes in duration). We compared the results for stressed with those for unstressed token pairs and also the results for tokens from an alveolar context with those from all tokens pooled.

Stressed vowel tokens were generally longer than the unstressed tokens and spectrally less reduced (at least for $F1$). No differences between stressed and unstressed tokens were found when the effects of changes in speaking rate or duration were considered. The difference in duration between stressed and unstressed tokens was twice the difference between speaking rates. There was a difference in $F1$ formant frequency between stressed and unstressed tokens but stressed and unstressed tokens did not differ in the way speaking rate affected their formant frequencies, i.e., $F1$ was higher in fast rate speech, although the size of the effect of speaking rate might have been smaller in stressed tokens than in unstressed tokens. All this indicates that vowel duration alone is not enough to explain the differences between stressed and unstressed vowel realizations. This confirms the results of Nord (1987).

For tokens from an alveolar CVC context, the same uniform higher $F1$ frequency in the fast rate tokens was found as in the parent set. There was the same lack of effect of either speaking rate or duration on the $F2$. These results indicate that if coarticulation from an all-alveolar context was stronger in fast-rate speech than in normal-rate speech, the difference was too small to be measured by the methods used in this paper. We were only able to test a subset of Dutch

vowels and consonants. It is still possible that other CVC combinations are more strongly affected by speaking rate changes.

To summarize, the trends observed in vowel realizations in our parent set were also present in the stressed and unstressed realizations and in the realizations from an alveolar-vowel-alveolar context. Therefore, we conclude that the variation of these textual factors in our data did not influence the results we obtained.

D. Conclusions

This study was limited in that only one speaker was used who read aloud a single text. From the results we conclude that this speaker did not behave as predicted by the target-undershoot model. Even the refined versions of the target-undershoot model that incorporate alternative articulation strategies (Gay, 1981) and increased effort (Lindblom, 1983) would predict some measurable differences in formant frequency values between speaking rates. That these differences were not found indicates that these theories are not universally valid for all speakers using continuous read speech. We found evidence that they might explain some aspects of the relation between vowel duration and formants within a single speaking style. However, our study indicates that their explanatory powers are limited and probably speaker specific.

The results presented here indicate that the articulatory effects of differences in vowel duration *between* speaking rates (and probably speaking styles) are not the same as the effects of differences in vowel duration *within* a single speaking rate (or style). This difference should be addressed by articulation theories based on the target undershoot model. It is also clear that our speaker was readily able to actively adapt his articulation to a fast speaking rate. It is therefore unlikely that articulation speed is a limiting factor in his vowel pronunciation as is implied by the target undershoot model.

ACKNOWLEDGMENTS

The authors wish to thank Dr. A. C. M. Rietveld of the Catholic University of Nijmegen, The Netherlands, for performing the sentence accent labeling of the speech material used in this study. The text used was selected by Dr. W. Eefting of the State University of Utrecht, The Netherlands, and the speech was recorded by her and Dr. J. Terken of the Institute of Perception Research, Eindhoven, The Netherlands. We also want to thank Dr. B. Lindblom, Dr. G. Bloothoof, and an anonymous reviewer for their helpful comments. This research project was part of the Dutch national program "Analysis and synthesis of speech" funded by the Dutch program for the Advancement of Information Technology (SPIN).

Abramowitz, M., and Stegun, I. A. (1965). *Handbook of mathematical functions* (Dover, New York, NY).

Benguerel, A.-P., and McFadden, T. U. (1989). "The effect of coarticulation on the role of transitions in vowel perception," *Phonetica* 46, 880-896.

- Broad, D. J., and Clermont, F. (1987). "A methodology for modelling vowel formant contours in CVC context," *J. Acoust. Soc. Am.* 81, 155-165.
- Broad, D. J., and Fertig, R. H. (1970). "Formant-frequency trajectories in selected CVC-syllable nuclei," *J. Acoust. Soc. Am.* 47, 1572-1582.
- Den Os, E. A. (1988). *Rhythm and tempo of Dutch and Italian; a contrastive study* (Ph.D. thesis, University of Utrecht, The Netherlands).
- Di Benedetto, M. G. (1989). "Vowel representation: Some observations on temporal and spectral properties of the first formant frequency," *J. Acoust. Soc. Am.* 86, 55-66.
- Duez, D. (1989). "Second formant locus-nucleus patterns in spontaneous speech: some preliminary results on French," *Phonetic Experimental Research Institute of Linguistics, University of Stockholm (PERILUS) X*, 109-114.
- Eefting, W. (1991). "The effect of information value and accentuation on the duration of Dutch words, syllables and segments," *J. Acoust. Soc. Am.* 89, 412-424.
- Engstrand, O. (1988). "Articulatory correlates of stress and speaking rate in Swedish VCV utterances," *J. Acoust. Soc. Am.* 85, 1863-1875.
- Fant, G., and Kruckenberg, A. (1989). "Preliminaries to the study of Swedish prose reading and reading style," *STL-QPSR* 2/1989, 1-83.
- Ferguson, G. A. (1981). *Statistical Analysis in Psychology and Education, International Student Edition* (McGraw-Hill, New York), pp. 381-406.
- Fourakis, M. (1991). "Tempo, stress, and vowel reduction in American English," *J. Acoust. Soc. Am.* 90, 1816-1827.
- Gay, T. (1978). "Effect of speaking rate on vowel formant movements," *J. Acoust. Soc. Am.* 63, 223-230.
- Gay, T. (1981). "Mechanisms in the control of speech rate," *Phonetica* 38, 148-158.
- Gopal, H. S., and Syrdal, A. K. (1988). "Effects of speaking rate on temporal and spectral characteristics of American English vowels," *Speech Communications Group Working Papers VI, Research Laboratory of Electronics MIT*, 162-180.
- Koopmans-van Beinum, F. J. (1990). "Spectro-temporal reduction and expansion in spontaneous speech and read text: the role of focus words," *Proceedings ICSLP 90 Vol. 1*, 21-24.
- Krull, D. (1989). "Second formant locus patterns and consonant-vowel coarticulation in spontaneous speech," *Phonetic Experimental Research Institute of Linguistics, University of Stockholm (PERILUS) X*, 87-108.
- Lindblom, B. (1963). "Spectrographic study of vowel reduction," *J. Acoust. Soc. Am.* 35, 1773-1781.
- Lindblom, B. (1983). "Economy of speech gestures" in *The Production of Speech*, edited by P. F. MacNeilage (Springer-Verlag, New York, NY), pp. 217-246.
- Lindblom, B., and Moon, S.-J. (1988). "Formant undershoot in clear and citation-form speech," *Phonetic Experimental Research Institute of Linguistics, University of Stockholm (PERILUS) VIII*, 21-33.
- Moon, S.-J. (1990). "An acoustic and perceptual study of formant undershoot in clear- and citation-form speech," *J. Acoust. Soc. Am. Suppl. 1* 88, S129.
- Nord, L. (1987). "Vowel reduction in Swedish" in *Papers from the Swedish Phonetics Conference*, edited by O. Engstrand (Uppsala), pp. 16-21.
- Nossair, Z. B., and Zahorian, S. A. (1991). "Dynamic spectral shape features as acoustic correlates for initial stop consonants," *J. Acoust. Soc. Am.* 89, 2978-2991.
- Strange, W. (1989a). "Evolving theories of vowel perception," *J. Acoust. Soc. Am.* 85, 2081-2087.
- Strange, W. (1989b). "Dynamic specification of coarticulated vowels spoken in sentence context," *J. Acoust. Soc. Am.* 85, 2135-2153.
- Van Son, R. J. J. H., and Pols, L. C. W. (1989). "Comparing formant movements in fast and normal rate speech," in *Eurospeech 89, the European Conference on Speech Communication and Technology*, edited by J. P. Tubach and J. J. Mariani (CEP Consultants, Edinburg, UK), Vol. 2, pp. 665-668.
- Van Son, R. J. J. H., and Pols, L. C. W. (1990). "Formant frequencies of Dutch vowels in a text, read at normal and fast rate," *J. Acoust. Soc. Am.* 88, 1683-1693.
- Vogten, L. L. M. (1986). "LVS speech processing programs on IPO-VAX11/780," *Manual 67, Institute for Perception Research, Eindhoven, The Netherlands*.
- Willems, L. F. (1986b). "Robust formant analysis," *Annual progress report 21, Institute for Perception Research, Eindhoven, The Netherlands*, pp. 34-40.