4

THE INFLUENCE OF SPEAKING RATE ON VOWEL FORMANT TRACK SHAPE AS MODELED BY LEGENDRE POLYNOMIALS*

Abstract

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. Legendre polynomial functions were used to model and quantify the shape of normalized formant tracks. No differences in normalized formant track shapes were found that could be attributed to differences in speaking rate. But a higher F_1 frequency in fast-rate speech relative to normal-rate speech was found that can be explained as the result of a uniform change in frequency. These results indicate a much more active adaptation to speaking rate than implied by the target-undershoot model. Within each speaking rate, there was only evidence of a weak leveling off of the F_1 tracks of the open vowels /E A a/ *with shorter durations. These same conclusions were reached when sentence-stress was taken into consideration and when vowel realizations from a more uniform, alveolar-vowel-alveolar, context were examined separately. In the alveolar context, a small rise in F_2 of the vowel* /o/ *might indicate more coarticulation in fast-rate speech.*

*Adapted from: Van Son, R.J.J.H. & Pols, L.C.W. (1991). "The influence of speaking rate on vowel formant track shape as modeled by Legendre polynomials", *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 15, 43-59.

Introduction

Vowel duration is generally considered an important parameter in determining the pronunciation of vowels and therefore of vowel formant tracks (e.g., Lindblom, 1963; Broad and Fertig, 1970; Gay, 1978, 1981; Lindblom, 1983; Broad and Clermont, 1987; Di Benedetto, 1989a; Lindblom and Moon, 1988; Moon, 1990). Vowel duration is important for the shape of the overall formant tracks. The target-undershoot model (Lindblom, 1963, 1983) is often cited to explain vowel formant behaviour under different speaking conditions. It predicts more coarticulation when vowels become shorter. In a large sample of normal speech, with typical utterances, this averages out to more spectral reduction, i.e. more schwa-like formant values in the vowel nucleus and more level (less curved) formant tracks (cf. Koopmans-van Beinum, 1980; Van Bergem, 1993). In a previous study we found that there was no evidence for an increased reduction or more coarticulation in fast-rate speech of a highly experienced speaker (chapter 1; Van Son and Pols, 1990), at least not in the vowel nucleus.

Relatively few studies have considered the relation between vowel formant dynamics and duration (exceptions are Broad and Fertig, 1970; Broad and Clermont, 1987; Di Benedetto, 1989a) 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 (but see chapter 3; Van Son and Pols, 1989, 1992). 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 or, conversely, an amplification 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 conventionally measured using two to four points from each formant track (Di Benedetto, 1989a; Strange, 1989 a, 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 was developed to characterize formant track shapes. First vowel formant tracks were sampled (16 points, adapted from Broad and Fertig, 1970). Second, the global "shape" of the sampled formant tracks was modeled with Legendre polynomials of order 0-4 (see section 4.1.1). This modeling approach was used to investigate the effects of speaking rate on vowel formant track shape. In chapter 3, this problem was studied using the 16 equidistant points directly (cf. Van Son and Pols, 1992).

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, 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 modeled. The effects of stress and vowel context were also taken into account.

4.1 Methods

The work presented in this chapter used the sampled formant track values obtained in chapter 3. We refer to that chapter for a description of the vowel segments and the methods used to obtain the sampled formant tracks. For convenience we reproduce the table with the number of vowel realizations used (table 4.1, which is identical to table 3.1).

4.1.1 Measuring differences between formant tracks

Legendre polynomial coefficients of order 0-4 were used as measures of formant track shape, see table 4.2 and figure 4.1 (appendix B; Churchhouse, 1981; Abramowitz and Stegun, 1965, pp.773-802). The Legendre polynomials are the simplest set of orthogonal polynomials and are generally easier to use than other sets. For practical reasons, we used the shifted Legendre polynomials which are defined on the base [0,1] instead of [-1,1].

An analysis using Legendre polynomials is a kind of regression analysis. The Legendre polynomial coefficients are calculated as a linear combination of the formant track sample points (see appendix B). Therefore, when the data points have a Gaussian distribution, all the coefficients also have a Gaussian distribution and the corresponding statistics can be used to test for differences between Legendre coefficients. The coefficients include the mean value (order 0) and linear regression slope (order 1). The second-order coefficient measures the parabolic excursion within a vowel realization, independent of the overall slope of the formant track. The third- and fourth-order coefficients measure, among other things, the amount of "stability" in the central part of the vowel (c.f. figure 4.1). The Legendre polynomials are orthogonal, meaning that the Legendre polynomial coefficients that describe track shape are mathematically independent. Because the zeroth-

Table 4.1: 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 parenthesis for some vowels (Dutch alveolar consonants are /n t d s z l r/, *see text).*

vowel	stressed	unstressed	unequal stress	total	
E	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)
´	0	21	0	21	
u	4	7	5	16	
y	5	6	1	12	
total	116	381	52	549	(135)

Figure 4.1.a: The first five Legendre polynomials, L_0-L_4. The polynomials are drawn with different Legendre coefficients P_i (actually the function $P_i \cdot L_i$ is drawn): P_0=1, P_1=P_2= -0.5, P_3=P_4= -0.25.

order measures the mean formant frequency, the results for this order should be identical to those found with the Averaging method in Van Son and Pols (1990) which uses the same speech data (see chapter 2).

Calculation of the Legendre polynomial coefficients was done by integration of the product of the sampled formant track and the appropriate Legendre polynomial function. We used the closed-type Newton-Cotes formulas to perform the numerical integration (Abramowitz and Stegun, 1965 p.886; appendix B). Because no 15th-order version of the Newton-Cotes formulas was available, we integrated the 15 intervals between the 16 track samples in two parts with the Legendre functions. The first part with the leading eight intervals (eighth-order Newton-Cotes formula) and the second part with the trailing seven intervals (seventh-order Newton-Cotes formula).

Legendre polynomials are used to model data points. The remaining variance after the fit is calculated by subtracting the variances of the various order polynomials, defined as $P_i \cdot P_i/\{1+2 \cdot i\}$ (P_i is the Legendre polynomial coefficient and i the order, Abramowitz and Stegun, 1965 pp.773-802; Churchhouse, 1981), from the original variance of the function. The remaining error (i.e., the RMS error) is the square-root of the remaining variance. The precision of the coefficients, especially the higher order ones, is limited by the precision of the calculations and the incomplete equivalence between the integration of continuous functions and the numerically integration of sampled data. However, this proved to be no problem.

Table 4.2: First five shifted Legendre polynomials and their slope at three points.
The polynomials, $L(\tau)$, are defined between 0 and 1 (inclusive). Next to the expressions the slope values of the polynomials are given for three points in the first half of the interval. The relative time τ is defined as time/duration ($0 \le \tau \le 1$). $L_i(0) = 1$ for even-order polynomials and $L_i(0) = -1$ for odd-order polynomials, $L_i(1) = 1$ for all polynomials. Even-order polynomials are symmetrical and odd-order polynomials are anti-symmetrical, i.e. if $-0.5 \le \varepsilon \le 0.5$ and $L_i' = dL_i/d\tau$ then $L_i(0.5+\varepsilon) = L_i(0.5-\varepsilon)$ and $L_i'(0.5+\varepsilon) = -L_i'(0.5-\varepsilon)$ if i is even and $L_i(0.5+\varepsilon) = -L_i(0.5-\varepsilon)$ and $L_i'(0.5+\varepsilon) = L_i'(0.5-\varepsilon)$ if i is odd (Adapted from Abramowitz and Stegun, 1965).

order	L_i ($0 \le t \le 1$)	$L_i'(0)$	$L_i'(0.25)$	$L_i'(0.5)$
0	1	0	0	0
1	$2 \cdot \tau - 1$	2	2	2
2	$6 \cdot \tau^2 - 6 \cdot \tau + 1$	-6	-3	0
3	$20 \cdot \tau^3 - 30 \cdot \tau^2 + 12 \cdot \tau - 1$	12	0.75	-3
4	$70 \cdot \tau^4 - 140 \cdot \tau^3 + 90 \cdot \tau^2 - 20 \cdot \tau + 1$	-20	3.125	0

Figure 4.1.b: Example of Legendre polynomials and their use in modeling functions. Tracks composed of different Legendre polynomials, using the same coefficient as in 4.1.a. Top: $1L_0 - 0.5L_1 - 0.25L_3$, bottom: $1L_0 - 0.5L_2 - 0.25L_4$. When formant frequency tracks are modeled, the horizontal axis represents the normalized time and the vertical axis the formant frequency in Hz. Note that tracks are shaped like formant tracks.

4.2 Results

The formant tracks 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), all statistical tables from Abramowitz and Stegun (1965 pp.966-990). Correlation coefficients were recalculated to a Student's *t* (Ferguson, 1981) to determine significance. To prevent repeated-test results from containing spurious errors, a two tailed threshold level for statistical significance of p•0.1% was chosen for testing Legendre polynomial coefficients (five values per formant 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 low numbers of realizations prevented us from distinguishing between speaking rates.

Vowel tokens spoken at a fast rate were 15% shorter (on average) than tokens spoken at a normal rate. The difference was consistent for all vowels except /´/ and statistically significant for /E A a i o/ (p•0.1%). The correlation between vowel durations at different speaking rates was high and statistically significant (p•0.1%, 0.64•r•0.89 except for /´/).

4.2.1 Goodness of fit

The Legendre polynomials were meant to model formant track shape. It was therefore important to know how well they fit the formant tracks and how much each order contributes to the overall fit (see section 4.1.1). In table 4.3, the proportion of variance (in percent), explained by each component was calculated for individual tokens and then averaged over all tokens. The contribution of the zeroth-order component (the mean formant frequency) represents the variance around zero frequency, which is not instructive for models of formant track *shape*. Therefore, the zeroth order component was left out: the variance was calculated around the mean frequency. Also, the remaining part of the variance left after the fit (the RMS error) was calculated.

In table 4.3 it can be seen that the bulk of the variance in the individual formant tracks could be explained by the first- and the second-order polynomials (65% - 93%). The remaining variance, left after fitting all Legendre polynomials up to order 4, was between 1% and 12%. The proportion of the variance that remained after the fit, tended to be higher when there was less movement in the formant tracks, i.e. when there was only a small variance to explain (e.g., F_1 of /u o y i/). For most vowel formant tracks, the amount of variance explained decreases with the order of the Legendre coefficient. Exceptions are the F_1 tracks of the vowels /E A a/, and the F_2 track of the vowel /i/. For these formant tracks the second-order coefficient explains most of the variance (up to 66%, table 4.3), making it the determining factor of track shape.

Table 4.3: Mean percentage of formant track variance around the mean formant frequency (i.e., excluding the zeroth-order Legendre coefficient) explained by the higher order Legendre polynomials (order 1-4) for each vowel. In the last column (rest), the mean percentage of the remaining (i.e., not explained) variance is given. Tokens from both speaking rates are pooled.

vowel		1	2	3	4	rest
E	F_1	39	54	3	2	2
	F_2	51	32	9	4	4
A	F_1	31	61	5	2	2
	F_2	67	17	8	3	5
a	F_1	25	66	4	2	3
	F_2	62	23	7	4	5
i	F_1	51	21	15	6	7
	F_2	38	42	7	5	7
o	F_1	40	29	17	5	9
	F_2	47	32	10	7	5
´	F_1	58	32	6	3	1
	F_2	56	26	9	5	4
u	F_1	47	18	14	9	12
	F_2	60	31	4	3	2
y	F_1	37	37	14	6	6
	F_2	82	10	3	2	3

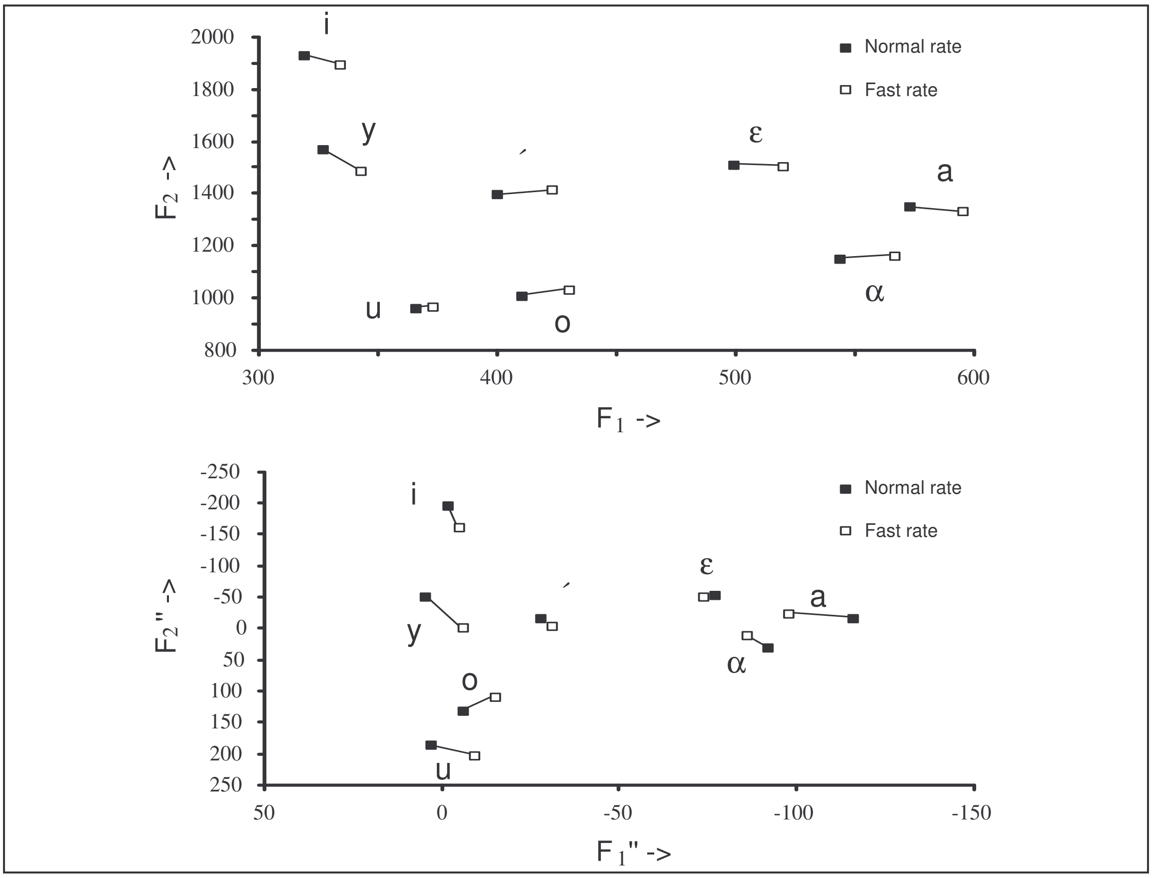

Figure 4.2: Vowel space (F_1/F_2 space) constructed by plotting mean Legendre polynomial coefficient values for the second formant frequency against the mean coefficient values for the first formant frequency for all vowels used. Filled squares: normal-rate tokens, open squares: fast-rate tokens. Upper panel: Zeroth-order Legendre polynomial coefficients P_0 (i.e., mean formant frequency within the realization). This plot results in the normal vowel triangle. Lower panel: Second-order Legendre polynomial coefficients P_2 (F_n", note reverse axes).

4.2.2 Legendre polynomial coefficients and their interpretation

In table 4.4 the mean values of the Legendre coefficients are presented for the orders 0-2. Of all polynomial coefficients, only the zeroth- and second-order coefficient values differed systematically (i.e., statistically significant for both speaking rates) from zero. Almost all mean first-order coefficient values were negative but only a few values were statistically significantly different from zero for both speaking rates (F_2 of /A/). Therefore, the first order polynomial coefficient, which corresponds to the linear regression slope, was important for describing the shape of each individual formant track (see previous section), but the sign of the coefficient (i.e., the slope) was not determined for any vowel.

The zeroth-order coefficient corresponds to the mean formant frequency. It is known that the value of the mean formant frequency is a strong cue to vowel identity (e.g., see chapter 1; Van Son and Pols, 1990). The value of the second-order coefficient can be interpreted as an excursion size relative to a straight line, i.e. the difference between maximum and minimum value of the second-order polynomial.

From the formulae in table 4.2 it follows that this excursion size is 1.5 times the value of the second-order coefficient (in Hz). For F_1, the values of the mean second-order coefficient were between 5 and -116 (table 4.4.a), which amounts to excursion sizes of between 0 and about 180 Hz. For F_2, the mean second-order coefficient values were between -196 and +203 (table 4.4.b), which corresponded to excursion sizes (absolute values) between 0 and approximately 300 Hz. These values are in line with the differences between formant values of vowel onset and nucleus found by Di Benedetto (1989a) for F_1, and Krull (1989) and Weismer et al. (1988) for F_2. These studies also show that much larger excursion sizes are found when speaking styles other than reading a text are involved (reference speech in Krull, 1989), or with certain consonant-vowel combinations that were hardly or not at all present in the speech material used here (e.g., /w/ context in Weismer et al., 1988; /u/ in Krull, 1989). The fact that in a variable context the mean excursion size of some vowels was systematically, and substantially, different from zero indicates that formant excursion size could be used to determine vowel identity (see below).

The mean third- and fourth-order coefficient values were not statistically significantly different from zero, except the fourth-order coefficient values of F_1: $\underline{9}$, $\underline{16}$ for /a/ and F_2: $\underline{32}$, $\underline{37}$ for /o/, normal and fast respectively (data not shown). Also, the contribution of the third- and fourth-order polynomials to the total fit were small and often negligible (table 4.3). Therefore, we will not discuss them in the remaining part of this paper. We did use them to estimate the slope values (see below).

From the polynomial coefficients, the normalized slope at each point in

Table 4.4.a: Mean values of first formant (F_1) in Hz.Legendre polynomial coefficients (order 0-2) and calculated mean value of normalized slope at τ = 1/4 and τ = 3/4 (SL 1/4 and SL 3/4 in Hz/segment, see table 4.2) Mean values that are statistically different from zero are underlined (Student's t-test, p≤0.1%). Whenever the fast-rate value differs significantly from the normal-rate value, this is indicated with a "" (Student's t-test on difference, p≤0.1%). Normal-rate: top row (N), fast-rate: bottom row (F).*

vowel		0	1	2	SL 1/4	SL 3/4
E	N	$\underline{499}$	$\underline{-33}$	$\underline{-77}$	$\underline{-161}$	$\underline{-297}$
	F	* $\underline{520}$	* -9	$\underline{-74}$	$\underline{199}$	$\underline{-241}$
A	N	$\underline{544}$	-21	$\underline{-92}$	$\underline{236}$	$\underline{-324}$
	F	* $\underline{567}$	-15	$\underline{-86}$	$\underline{213}$	$\underline{-280}$
a	N	$\underline{573}$	$\underline{-24}$	$\underline{-116}$	$\underline{252}$	$\underline{-338}$
	F	* $\underline{595}$	-10	* $\underline{-98}$	$\underline{249}$	$\underline{-287}$
i	N	$\underline{319}$	$\underline{-12}$	-2	-21	-21
	F	* $\underline{334}$	-11	-5	-6	-24
o	N	$\underline{410}$	$\underline{-14}$	-6	18	$\underline{-62}$
	F	* $\underline{430}$	-10	$\underline{-15}$	41	$\underline{-65}$
´	N	$\underline{400}$	-32	-28	16	-139
	F	$\underline{423}$	-33	$\underline{-31}$	18	$\underline{-144}$
u	N	$\underline{366}$	-11	-3	14	-57
	F	$\underline{373}$	-26	-9	-15	-82
y	N	$\underline{327}$	13	5	5	54
	F	$\underline{343}$	5	-6	12	23

the original formant tracks was approximated by summing the values of the slopes of the individual Legendre polynomials at these points (table 4.2), multiplied by the corresponding Legendre coefficient. We calculated the normalized slopes at points at one-fourth (SL1/4) and three-fourths (SL3/4) of the normalized duration of each vowel and averaged them just like the Legendre coefficients (table 4.4, last two columns). These two points are positioned to lie in the on-and off-glide of the vowels, except for the long vowels, /a o/, where they may occasionally lie in the vowel nucleus.

The slopes in the on- and off-glide parts of the vowels, as estimated from all five Legendre polynomials, differed in a systematic way from zero for many vowels but were nevertheless difficult to interpret. Often the absolute values of the slopes on the onglide of the tokens were very different from those on the offglide (table 4.4). This difference showed that vowel formant track shapes were generally asymmetric.

The differences in slope of the formant tracks between fast- and normal-rate tokens (after time-normalization) were never statistically significant and thus did not help us to determine the effects of speaking rate on formant track dynamics.

4.2.3 Relations between polynomial components

The mean values of the zeroth- and second-order coefficients were linked together: higher zeroth-order coefficient values were accompanied by lower (more negative) second-order coefficients. Negative second-order coefficients imply a maximum in the formant track, positive coefficients imply a minimum. This correlation was statistically significant for all vowels pooled ($|r| = 0.6$, p•0.1%). In the upper panel of figure 4.2, the mean zeroth-order coefficient values are plotted, F_2 against F_1, for both speaking rates (compare figure 2.1; Van Son and Pols, 1990). In the lower panel, the second-order coefficients are presented. For both orders, the mean coefficient values of the individual vowels form the familiar vowel

Table 4.4.b: As table 4.4.a. Second formant (F_2)

vowel		0	1	2	SL 1/4	SL 3/4
E	N	1507	-55	-53	23	-249
	F	1500	-35	-49	41	-192
A	N	1146	-51	31	-160	-31
	F	1159	-40	* 11	* -89	-69
a	N	1349	-38	-16	-65	-117
	F	1329	-26	-23	2	-121
i	N	1929	-67	-196	447	-724
	F	1892	-40	-162	358	-528
o	N	1009	-30	132	-339	221
	F	1031	-35	111	-305	156
´	N	1396	-7	-15	55	-85
	F	1414	1	-4	88	-60
u	N	960	-35	187	-605	432
	F	962	2	203	-603	597
y	N	1568	-157	-49	-145	-471
	F	1487	-157	-1	-388	-219

triangle. For the zeroth-order coefficient values this was expected, for the second-order coefficient values this was new. Presupposing random ordering, the probability of just this constellation for the mean second-order coefficients is less than 0.1% (in the upper panel /i y u o A a E ´/ are ordered in a spiral, the probability of just such a spiral in the lower panel is 4·8/8!•0.0008, allowing for the freedom to choose the signs of the axes (2·2=4) and the ambiguity of the order of a single pair (/u o/: 8)).

Figure 4.2 suggests that in the F_1 direction the second-order coefficient values could be interpreted as a measure of openness: closed has value zero, e.g. the vowels /u y i/. In the F_2 direction it could be interpreted as a measure of front- versus back-articulation: schwa has value zero (i.e., flat), /u/ is positive (i.e., a minimum) and /i/ is negative (i.e., a maximum). Based on the second-order polynomial coefficient and the vowels used here, the vowels could be grouped in distinguishable sets. This meant that the vowel-sets /u o/, /y/, /i/, /E A a/ and /´/ could be distinguished from each other with statistical significance (p•0.1%, Students-*t* test on means of F_1 or F_2), by only using the value of the second-order coefficient of individual vowel realizations. This fact and the large contribution to the overall shape of the formant tracks (especially F_1, see section 4.2.3) suggested that the second-order coefficient could be an important cue of the relation between vowel identity and vowel formant track shape.

The correlation between zeroth- and second-order Legendre coefficients was not statistically significant for the tokens of any *single* vowel (|r| • 0.15 none significant, not shown), contrary to what was found when all vowel realizations were pooled. Therefore, zeroth- and second-order Legendre coefficient values can be considered to be independent apart from being both related to the vowel identity.

Correlations between different orders of Legendre polynomial coefficients were not always small. Of all correlations between all different order coefficient values from tokens of the same vowel, approximately 7% was statistically significant (p•0.01% each). However, we could not find any pattern in these correlations (data not shown). From this we inferred that the contributions of polynomials of different orders were indeed independent from each other, but that extraneous (e.g., textual) factors could have caused correlations between polynomial coefficients of different orders that depended on the distribution of these factors in the text.

4.2.4 Effects of speaking rate

The zeroth-order component (i.e., mean formant value) of F_1 from the vowels /E A a o/ (table 4.4.a) showed a higher fast-rate value compared to the normal-rate value. The other, higher order, components rarely showed statistically significant differences between speaking rates, only first-order F_1 of the vowel /E/, and second-order F_1 of the vowel /a/ and F_2 of the vowel /A/ (table 4.4.a, b). From this we can conclude that the F_1 frequency of fast spoken vowels is higher than the F_1 frequency of tokens spoken at a normal rate. The difference is uniform and irrespective of vowel identity.

Correlations between speaking rates of the zeroth-order (mean value) component were high and statistically significant (p•0.1%, table 4.5). First-order coefficient values showed significant correlations between speaking rates, but generally with lower correlation coefficients than those of the zeroth-order components. Second-, third- and fourth-order components often showed statistically significant correlations between speaking rates, especially for F_2 (table 4.5, only second-order is shown). The correlation coefficients of F_2 were higher than those of F_1 in most vowels. The correlation coefficients decreased with increasing order but still remained quite high (up to r=0.74 for /o/, third-order F_2, not shown). These results led to the conclusion that higher order components of formant tracks contained information that was preserved between speaking rates. All different order components could be used to investigate the effects of duration on vowel formant shape.

Generally, there was no extra information to extract from the on- and off-glide slopes. Between-speaking-rate correlation coefficients of the slope values were almost always lower than those of the first-order component.

4.2.5 Relation between polynomial coefficients and vowel duration

The polynomial coefficient values found for the formant tracks were correlated with vowel duration. This correlation was performed for both speaking rates independently (not shown). Generally, the correlation coefficients between Legendre coefficient values and vowel duration were small and statistically not significant for both speaking rates. An exception were the second-order Legendre coefficients of the F_1 of the vowels /E A a/ (r•0.33-0.52, p•0.1%). These coefficient values were almost as high as the between-speaking-rate correlation coefficients (cf. table 4.5). The correlations between duration and second-order components of F_1 implied a decrease in

Table 4.5: Correlation coefficients between speaking rates of Legendre polynomial coefficients (order 0-2) and of calculated mean values of normalized slope at $\tau = 1/4$ and $\tau = 3/4$ (SL 1/4 and SL 3/4, see Table 4.2). Correlation coefficients that are statistically different from zero are underlined (coefficients recalculated for Student's t-test, p≤0.1%).

vowel		0	1	2	SL 1/4	SL 3/4
E	F_1	0.62	0.47	0.47	0.46	0.41
	F_2	0.87	0.76	0.54	0.69	0.44
A	F_1	0.86	0.67	0.46	0.64	0.49
	F_2	0.91	0.86	0.68	0.81	0.61
a	F_1	0.71	0.59	0.55	0.47	0.52
	F_2	0.85	0.85	0.67	0.85	0.56
i	F_1	0.57	0.69	0.46	0.42	0.51
	F_2	0.32	0.50	0.25	0.29	0.04
o	F_1	0.85	0.69	0.70	0.66	0.60
	F_2	0.87	0.78	0.76	0.68	0.75
´	F_1	0.55	0.36	0.40	0.74	0.28
	F_2	0.95	0.83	0.19	0.66	0.55
u	F_1	0.04	0.75	0.26	0.06	0.58
	F_2	0.73	0.86	0.73	0.83	0.75
y	F_1	0.73	0.62	0.54	0.39	0.19
	F_2	0.84	0.88	0.72	0.32	0.81

curvature (or excursion size) for shorter durations, i.e. shorter vowels had more level formant tracks.

The correlation coefficients between on- and offglide slopes and vowel duration that were statistically significant were all comparable in size to those between the second-order coefficients and vowel duration. The former relation can most likely be explained from the latter. All other correlation coefficients were small and not statistically significant for either speaking rate.

4.2.6 Effects of context

A subset of the tokens of the most numerous vowels /E A a i o/ in an all alveolar CVC context was analysed separately (i.e., C is one of /n t d s z r l/). For each vowel, the number of tokens available in an alveolar context was quite small (between 16 and 38, see table 4.1). 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.

The mean values of the Legendre polynomial coefficients (order 0-2) and the estimated slope at 1/4 and 3/4 of the vowel did not differ much from those found for the tokens of the parent set (table 4.4). The second-order Legendre coefficients of the F_1 tracks of the vowels /E A a/ might be an exception. The tokens of these three high F_1 target vowels had a somewhat higher (up to 20%) mean second-order coefficient value for both speaking rates and the slopes at both points inside the tokens were somewhat steeper.

The fast-rate tokens of this subset had a uniformly higher F_1 than the normal-rate tokens (p•0.1% for /A o/, zeroth-order). The vowel /o/ also showed a slightly higher F_2 in the fast-rate tokens (42 Hz p•0.1%, zeroth-order). The between-speaking-rate correlation coefficients of the Legendre coefficients were high for both F_1 and F_2, often higher than those for the parent set. The trends were the same as in the parent set of tokens (table 4.5).

The correlation coefficients between Legendre polynomial coefficients or slope and vowel duration were generally higher in the subset of tokens in alveolar context than in the parent set (section 4.2.2). Still, only few correlation coefficients were statistically significant (p•0.1%, fast-rate F_1: second-order coefficient of /E A a/ and slope at 1/4 of /E/) or larger than the corresponding correlation between speaking rates (c.f. table 4.5). An exception was the second-order Legendre coefficients of the F_1 tracks of the fast-rate tokens of the vowels /E A a/. Here the correlation coefficients were higher (|r|•0.60-0.75, p•0.1%) than the coefficients obtained from the corresponding correlation between the two speaking rates.

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.

4.2.7 Effects of stress

The previous analyses were repeated on token-pairs of the vowels /E A 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•0.1%). The differences in vowel duration between speaking rates were comparable for stressed and unstressed tokens (i.e., 15%).

For the F_1, zeroth- and (negative) second-order Legendre coefficient values of the stressed tokens of the high F_1-target vowels /E A a/ were higher than those of the unstressed tokens at both rates (p•1% for vowels pooled). The vowel space of the stressed tokens was larger, i.e. less reduced, in the F_1 direction (/i/ to /a/) than that of the unstressed tokens, both for zeroth-order (5%) and second-order coefficients (25%). The slopes of the F_1 tracks of stressed tokens were generally steeper than those of unstressed tokens. Both the fast-rate stressed and unstressed tokens had a uniformly higher F_1 than the normal-rate tokens (zeroth-order, p•0.1%, stressed /E a/, unstressed all individual vowels).

Due to the lower number of realizations, the second-order coefficient values and track slopes of the F_2, were often not statistically significantly different from zero for the stressed tokens of vowels that did show significant values for the unstressed tokens. There was no indication that, compared to stressed tokens, unstressed tokens are spectrally reduced with respect to the F_2.

Generally, correlation coefficients, both for vowel duration and formants between speaking rates and between formants and vowel duration, were higher in stressed tokens than in unstressed tokens. The comparison was difficult because results for the stressed tokens were often statistically not significant due to the small number of stressed tokens. No other difference between stressed and unstressed tokens was found. As far as could be checked, the results obtained from all tokens pooled were equally valid for both of these subsets of tokens.

4.3 Discussion

The results found here generally are in agreement with those found using a more conventional type of analysis based on a direct comparison of the 16 equidistant points per vowel segment. These latter results are discussed in chapter 3 (see also Van Son and Pols, 1989, 1992). In this chapter we discuss specifically coordinated, whole track differences between speaking rates, instead of "local" point-by-point differences.

4.3.1 Effects of speaking rate

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 track shape parameters measured at different speaking rates. This means that, after normalization for duration, a difference in speaking rate did not result in systematic differences in formant track shape. Only the F_1 frequency is higher in vowels spoken at a fast rate than in vowels spoken at a normal rate, see figure 4.2. This rate-dependent rise in F_1 frequency was found irrespective of vowel identity. It was also limited to the zeroth-order Legendre polynomial (i.e., mean formant value).

This means that the F_1 frequencies in all parts of the fast-rate tracks were raised by roughly the same amount. This means that the equivalent results found by Van Son and Pols (1990; see chapter 2) for "static" measurements, in which method Average is identical to using the zeroth-order coefficient, must be attributed to an uniform increase in formant frequency over the whole F_1 track in fast-rate speech. It cannot be attributed to an increase in only the vowel nucleus or only the transition parts, which would also have changed the *shape* of the formant tracks (i.e., higher order Legendre coefficient values).

4.3.2 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. This means that duration should explain a significant part of the variance in formant track parameters (i.e., the variance in track parameters would be systematic and linked to the variance in duration). However, correlation coefficients between formant frequencies and vowel duration were only significant for the F_1 tracks of the high F_1 target vowels (/E A a/), see section 4.2.5. The correlations implied a leveling off of the F_1 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 second-order Legendre coefficients (0.33•|r|•0.52). Between-speaking-rate correlations for these three vowels sometimes explained up to 70% of the variance in F_1 formant track parameters (zeroth order,|r|•0.86, table 4.5). This indicates that a very large part of the variance in formant track parameters is indeed systematic and reproduced for each "reading" of the text, independent of speaking-rate. The fact that the correlation between formant track parameters and vowel duration is much weaker than the between-speaking-rate correlation indicated that duration is not a major determinant of overall vowel formant track shape in read speech.

There is one area where the correlation between formant track parameters and vowel duration is as strong as the between-speaking-rate correlation and where duration might indeed explain much of the systematic variance. For the second-order Legendre polynomial coefficients of the F_1, the between-speaking-rate correlation coefficients were not larger (i.e., 0.46•|r|•0.55, table 4.5) than those between Legendre coefficients and duration. This indicates that much of the *systematic* variance of the second-order Legendre coefficients of the F_1, as measured by the between-speaking-rate correlation, might indeed have been determined by vowel duration. The correlation between second-order Legendre coefficients and vowel duration was as predicted by the target-undershoot model, i.e. shorter duration were combined with more level formant tracks. But again, the absolute size of the effect of duration on track shape is minimal, generally explaining less than a quarter of the *total* variance observed.

F_2 formant tracks do not show any sizeable correlation between track parameters and vowel duration.

4.3.3 Effects of context and stress

The context in which a vowel is spoken might be important for the effects produced by changes in speaking rate (or changes in duration). We compared the differences in duration and in formant track shape between speaking rates for stressed with the differences for unstressed token-pairs and also the differences between speaking rates for tokens from an alveolar context with those from all tokens pooled.

Stressed vowel tokens were generally longer than the unstressed tokens and less reduced spectrally (at least for F_1). No differences between stressed and unstressed tokens were found when 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 F_1 formant frequency between stressed and unstressed tokens but no difference between speaking rates. This indicates that the vowel duration alone is not enough to explain the differences between stressed and unstressed vowel realizations, confirming the results of Nord (1987).

For tokens from an alveolar CVC context, we would expect the largest effects on the open vowels /E A a/ for the F_1 tracks and on the back vowel /o/ for the F_2 tracks (see section 3.1.2 of chapter 3). For fast-rate tokens we found an increase in the correlation between the second-order Legendre coefficient of the F_1 tracks of the vowels /E A a/ and vowel duration. This suggests that the constraints on F_1 formant movements might have been tighter for vowel realizations spoken at fast rate than for realizations spoken at normal rate in this extreme consonant context, i.e. closed-open-closed. The same uniformly higher F_1 frequency in the fast-rate tokens was found as in the parent set. For vowels in an alveolar context we found the same lack of effect of either speaking rate or duration on the F_2, except that in this context the F_2 of the vowel /o/ showed a small, uniform, increase in fast-rate speech. Therefore, there might have been more coarticulation or "target-undershoot" in the F_2 in this extreme context (alveolar-/o/-alveolar). But because only one vowel was affected it is difficult to interpret the change.

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. This shows that the effects of speaking rate on vowel realizations is to a large extent independent of sentence-stress and (alveolar) context.

4.4 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, which predicts more reduction (both static and dynamic) in vowel articulation with a faster speaking rate, especially when vowel durations are quite short to begin with. Even the refined versions of the target-undershoot model that incorporate alternative articulation strategies (Gay, 1981) and increased effort (Lindblom, 1983) on a global level, would predict some measurable differences in formant track shape or frequency values between speaking rates. That neither was found indicates that these theories are not universally valid for all speakers using continuous read speech. We cannot rule out the possibility that these theories might explain some aspects of the relation between vowel duration and formants within a single speaking style or when strong coarticulation is predicted. However, our study indicates that their explanatory power is limited and probably speaker specific. Based on these results, articulation models are needed that acknowledge a much more active behaviour of the speaker in adapting to a high speaking rate.