Institute of Phonetic Sciences,
University of Amsterdam,
Proceedings 15 (1991), 43-59.

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

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

Abstract

Speaking rate and vowel duration are generally thought to affect the dynamic structure of vowel formant tracks. This idea was tested by letting a single professional speaker read a long text at two different speaking rates, fast and normal. The extent to which the shape of the first and second formant tracks of 8 Dutch vowels varied under the two different speaking rate conditions was investigated. 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 F1 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 F1 tracks of the open vowels /ɛ, ɑ, 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 F2 of the vowel /o/ might indicate more coarticulation in fast rate speech.

Introduction

The pronunciation of vowels, and therefore the shape of their formant tracks, is generally considered to be determined to an important extend by vowel duration (e.g. Lindblom, 1963; Broad and Fertig, 1970; Gay, 1978; Gay, 1981; Lindblom, 1983; Broad and Clermont, 1987; Di Benedetto, 1989; Lindblom and Moon, 1988; Moon, 1990). The target undershoot model (Lindblom, 1963; Lindblom, 1983) is often cited to explain vowel formant behaviour under different speaking conditions. It 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. In a previous study we found that there was no evidence for an increased reduction or more coarticulation in fast rate speech (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 (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, 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 modelled with Legendre polynomials of order 0-4 (see section I.D). This modelling approach was used to investigate the effects of speaking rate on vowel formant track shape. A study of this problem using the 16 equidistant points directly will be published elsewhere, apart from this point the structure of both papers is very similar (Van Son and Pols, submitted).

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 modelled. Also the effects of stress and vowel context were taken into account.

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.

I.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 seconds for the normal speaking rate and 220 seconds for the fast speaking rate (4.4 and 6.6 syll./sec. including pauses, cf. Koopmans-van Beinum 1990). 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 neighbouring phonemes, was considered to be part of that vowel realization. The segments were

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 brackets for some vowels (Dutch alveolar consonants are /n, t, d, s, z, l, r/, see text).

vowel	stressed	unstressed	unequal stress	total	
ɛ	23	85	12	120	(21)
ɑ	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)

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.

I.B Vowels used

Seven of the twelve Dutch monophthongs were used: /i, y, u, o, a, ɑ, ɛ/. These vowels were selected according to their 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, ɑ, ɛ/) 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. 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 /ɛ, ɑ, 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 would expect the largest coarticulatory effects on the F1 tracks of the open vowels /ɛ, ɑ, a/ and the F2 tracks of the back vowel /o/. There were not enough tokens in another (non-alveolar) homogeneous context to merit analysis.

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

TABLE II. First five shifted Legendre polynomials and their slope at three points.
The polynomials 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 ($0 \le \tau \le 1$). $P_i(0) = 1$ for even order polynomials and $P_i(0) = -1$ for odd order polynomials, $P_i(1) = 1$ for all polynomials. Even order polynomials are symmetric and odd order polynomials are anti-symmetric, i.e. if $-0.5 \le \varepsilon \le 0.5$ and $P_i' = dP_i/d\tau$ then $P_i(0.5+\varepsilon) = P_i(0.5-\varepsilon)$ and $P_i'(0.5+\varepsilon) = -P_i'(0.5-\varepsilon)$ if i is even and $P_i(0.5+\varepsilon) = -P_i(0.5-\varepsilon)$ and $P_i'(0.5+\varepsilon) = P_i'(0.5-\varepsilon)$ if i is odd.
Adapted from Abramowitz and Stegun (1965).

order	P_i ($0 \le \tau \le 1$)	$P_i'(0)$	$P_i'(0.25)$	$P_i'(0.5)$
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

I.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).

A linear frequency scale was used for formant frequencies and for calculating Legendre polynomial coefficients. We also tested Bark and logarithmic scales, but their performance was not different from linear scales.

The formant tracks obtained from the different vowels were sampled at 16 equidistant points, including both boundaries. 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.

I.D Measuring differences between formant tracks

Legendre polynomial coefficients of order 0-4 were used as measures of formant track shape, see table II and figure 1 (Churchhouse, 1981; Abramowitz and Stegun, 1965, p773-802). 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. Therefore, when the data points have a Gaussian distribution, all the coefficients also have a Gaussian distribution and the corresponding statistics can be used. 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 1). The Legendre polynomials are orthogonal, meaning that the Legendre polynomial coefficients that describe track shape are mathematically independent. Because the zeroth order measures the mean formant

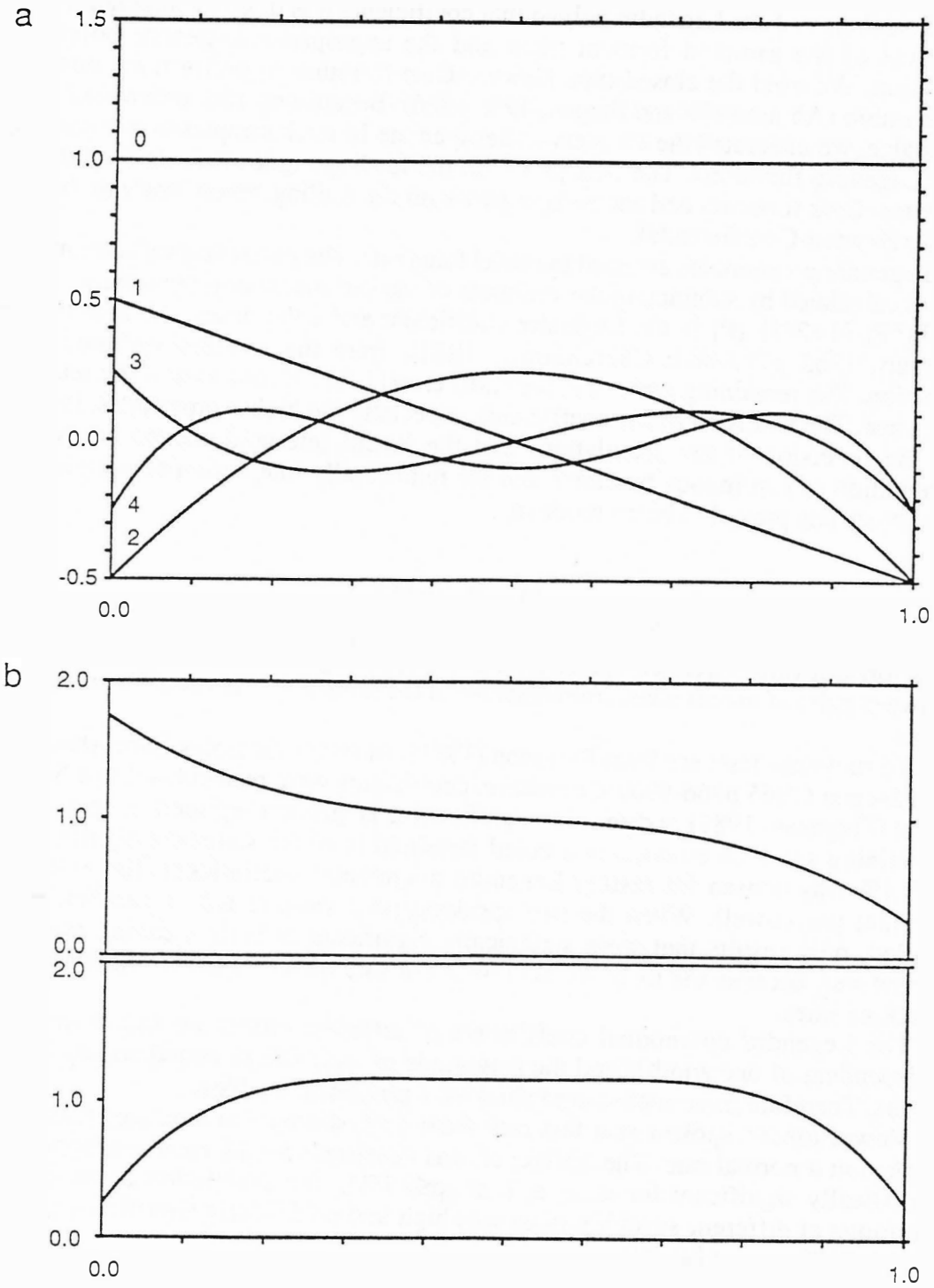


Figure 1: Example of Legendre polynomials and their use in modelling functions. When formant frequency tracks are modelled, the horizontal axis represents the normalized time and the vertical axis the formant frequency in Hz.
a. The first five Legendre polynomials, L_0-L_4. The polynomials are drawn with different Legendre coefficients P_i (actually the function P_i*L_i is drawn): $P_0=1$, $P_1=P_2=-0.5$, $P_3=P_4=-0.25$.
b. Tracks composed of different Legendre polynomials, using the same coefficient as in 1.a. Top: $1L_0 - 0.5L_1 - 0.25L_3$, bottom: $1L_0 - 0.5L_2 - 0.25L_4$. Note that tracks are shaped like formant tracks.

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.

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-Cote formulas to perform the numerical integration (Abramowitz and Stegun, 1965 p886). Because no 15th order version 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-Cote formula) and the second part with the trailing seven intervals (seventh order Newton-Cote formula).

Legendre polynomials are used to model functions. The remaining variance after the fit, is calculated by subtracting the variances of the various order polynomials, defined as $P_i*P_i/\{1+2*i\}$ (P_i is the Legendre coefficient and i the order, Abramowitz and Stegun, 1965 p773-802; 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.

II Results

The formant values 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 p966-990). Correlation coefficients were recalculated to a Student's t-test (Ferguson, 1981) to determine significance. To prevent repeated test results from containing spurious errors, a two tailed threshold level for statistical significance of $p \leq 0.1\%$ 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 methods used were not well qualified to distinguish between speaking rates.

The Legendre polynomial coefficients of different orders are (mathematically) independent of one another and the magnitude of their values varied widely between orders. Therefore these coefficient values were presented in tables.

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 /ə/ and statistically significant for /ɛ, ɑ, a, i, o/ ($p \leq 0.1\%$). The correlation between vowel durations at different speaking rates was high and statistically significant ($p \leq 0.1\%$, $0.64 \leq r \leq 0.89$ except for /ə/).

II.A 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 contributed to the overall fit (see section I.D). In table III, the fraction of variance (in percent), explained by each component was calculated for individual tokens and then averaged over all tokens. For clarity, the contribution of the zeroth order component

TABLE III. 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
ɛ	F1	39	54	3	2	2
	F2	51	32	9	4	4
ɑ	F1	31	61	5	2	2
	F2	67	17	8	3	5
a	F1	25	66	4	2	3
	F2	62	23	7	4	5
i	F1	51	21	15	6	7
	F2	38	42	7	5	7
o	F1	40	29	17	5	9
	F2	47	32	10	7	5
ə	F1	58	32	6	3	1
	F2	56	26	9	5	4
u	F1	47	18	14	9	12
	F2	60	31	4	3	2
y	F1	37	37	14	6	6
	F2	82	10	3	2	3

(the mean formant frequency) was left out: the variance was calculated around the mean frequency. Also, the remaining fraction of variance left after the fit (the RMS error) was calculated. In table III it can be seen that the bulk of the variance within 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, was between 1% and 12%. The fraction 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. F1 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 F1 tracks of the vowels /ɛ, ɑ, a/, and the F2 track of the vowel /i/. For these formant tracks the second order coefficient explains most of the variance (up to 66%, table III), making it the determining factor of track shape.

II.B Legendre polynomial coefficients and their interpretation

In table IV the mean values of the Legendre coefficients were 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 (F2 of /ɑ/).

For zeroth order (i.e. mean formant frequency) the value of the mean formant frequency is a strong cue to vowel identity (e.g. see: 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 formula of table II it follows that this excursion size is 1.5 times the value of the second order coefficient (in Hz). For F1, the values of the mean second order coefficient were between 5 and -116 (table IV, upper part), which conforms to excursion sizes of between 0 and about 180 Hz. For F2, the mean

TABLE IV. Mean values of 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 II).
Mean values that are statistically different from zero are underlined (Student's t-test, $p \leq 0.1\%$). Whenever the fast rate value differs significantly from the normal rate value, this is indicated with a "+" (Student's t-test on difference, $p \leq 0.1\%$).
Normal-rate: top row (N), fast-rate: bottom row (F).

First formant (F1)

vowel		0	1	2	SL 1/4	SL 3/4
ɛ	N	499	-33	-77	-161	-297
	F	+ 520	+ -9.3	-74	199	-241
ɑ	N	544	-21	-92	236	-324
	F	+ 567	-15	-86	213	-280
a	N	573	-24	-116	252	-338
	F	+ 595	-10	+ -98	249	-287
i	N	319	-12.2	-1.9	-21	-21
	F	+ 334	-10.5	-4.9	-5.8	-24
o	N	410	-14	-5.8	18	-62
	F	+ 430	-10.4	-15	41	-65
ə	N	400	-32	-28	16	-139
	F	423	-33	-31	18	-144
u	N	366	-11	-3.1	14	-57
	F	373	-26	-9	-15	-82
y	N	327	13	4.6	5	54
	F	343	5.3	-5.9	12	23

Second formant (F2)

vowel		0	1	2	SL 1/4	SL 3/4
ɛ	N	1507	-55	-53	23	-249
	F	1500	-35	-49	41	-192
ɑ	N	1146	-51	31	-160	-31
	F	1159	-40	+ 11.2	+ -89	-69
a	N	1349	-38	-16	-65	-117
	F	1329	-26	-23	2.3	-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
ə	N	1396	-7.4	-15	55	-85
	F	1414	1	-4	88	-60
u	N	960	-35	187	-605	432
	F	962	1.8	203	-603	597
y	N	1568	-157	-49	-145	-471
	F	1487	-157	-0.9	-388	-219

second order coefficient values were between -196 and +203 (table IV, lower part), 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 (1989, F1), Krull (1989, F2), and Weismer et al. (1988, F2). These studies also show that much larger excursion sizes are found when other speaking styles 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 vowel identity could be important in determining formant track shape (see below).

The mean third and fourth order coefficient values were not statistically significantly different from zero, except the fourth order coefficient values of F1: 9, 16 for /a/ and F2: 32, 37 for /o/, normal and fast respectively (others not shown). The contribution of the third and fourth order polynomials to the total fit were small and often negligible (table III) and the mean coefficient values are not significantly different from zero. 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 the original formant tracks was approximated by summing the values of the slopes of the individual Legendre polynomials at these points (table II), 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 IV, 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 one side of the tokens were very different from those on the other side (table IV). This difference indicated that vowel formant track shapes were often asymmetric, and probably curved.

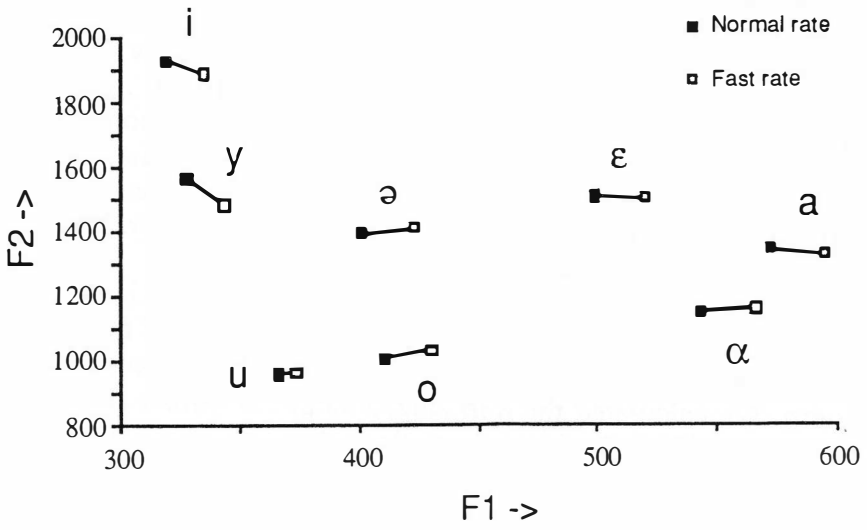
The differences in slope of the formant tracks between fast and normal rate tokens were never statistically significant and thus did not help us to determine the effects of speaking rate on formant track dynamics.

II.C 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 ($|r| = 0.6$, $p \leq 0.1\%$). In figure 2.a, the mean zeroth order coefficient values were plotted, F2 against F1, for both speaking rates (compare figure 1 of Van Son and Pols, 1990). In figure 2.b the second order coefficients were presented. For both orders, the mean coefficient values of the individual vowels are ordered in the familiar vowel triangle. For the zero 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 0.01%.

Figure 2.b suggests that the second order coefficient values could be interpreted as a measure of openness in the F1 direction: closed has value zero, e.g. the vowels /u, y, i/. In the F2 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 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/, /ɛ, ɑ, a/ and /ə/ could be distinguished from each other with statistical significance ($p \leq 0.1\%$), by only using the value of the second order

2.a



2.b

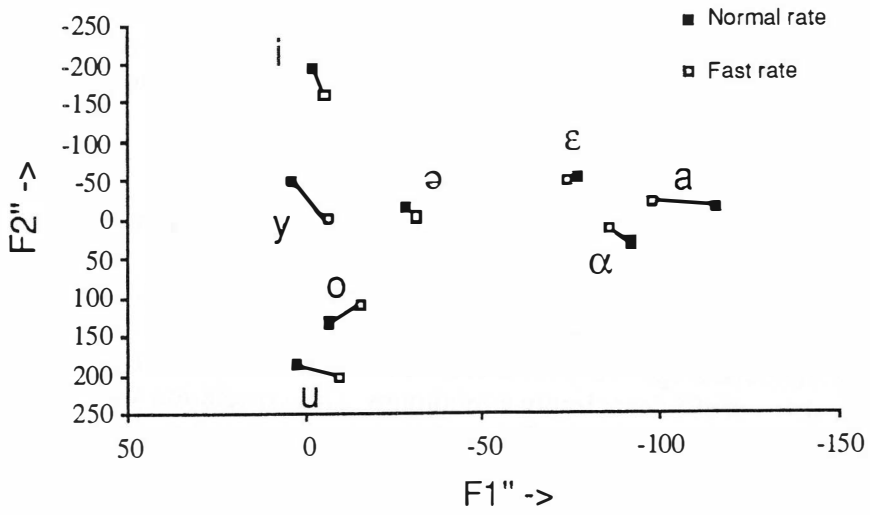


Figure 2: Vowel space (F1/F2 space) constructed by plotting mean Legendre polynomial coefficient values for the second formant frequency against the mean coefficient values for the first formant frequency for all vowels used. Filled squares: normal rate tokens, open squares: fast rate tokens.
a. Zëroth order Legendre polynomial coefficients (i.e. mean formant frequency within the realization). This plot results in the normal vowel triangle.
b. Second order Legendre polynomial coefficients, note reverse axes.

TABLE V. 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 II).
Correlation coefficients that are statistically different from zero are underlined (coefficients recalculated for Student's t-test, $p \leq 0.1\%$).

vowel		0	1	2	SL 1/4	SL 3/4
ɛ	F1	0.62	0.47	0.47	0.46	0.41
	F2	0.87	0.76	0.54	0.69	0.44
ɑ	F1	0.86	0.67	0.46	0.64	0.49
	F2	0.91	0.86	0.68	0.81	0.61
a	F1	0.71	0.59	0.55	0.47	0.52
	F2	0.85	0.85	0.67	0.85	0.56
i	F1	0.57	0.69	0.46	0.42	0.51
	F2	0.32	0.50	0.25	0.29	0.04
o	F1	0.85	0.69	0.70	0.66	0.60
	F2	0.87	0.78	0.76	0.68	0.75
ə	F1	0.55	0.36	0.40	0.74	0.28
	F2	0.95	0.83	0.19	0.66	0.55
u	F1	0.04	0.75	0.26	0.06	0.58
	F2	0.73	0.86	0.73	0.83	0.75
y	F1	0.73	0.62	0.54	0.39	0.19
	F2	0.84	0.88	0.72	0.32	0.81

coefficient of individual vowel realizations. This fact and the large contribution to the overall shape of the formant tracks (especially F1, see section II.C.1) suggested that the second order coefficient could be an important cue of the relation between vowel identity and vowel track shape.

The correlation between zeroth and second order Legendre coefficients was not statistically significant for the tokens of any single vowel ($|r| \leq 0.15$ none significant, not shown). 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 seven per cent was statistically significant ($p \leq 0.01\%$ 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.

II.D Effects of speaking rate

The zeroth order component (i.e. mean formant value) of F1 from the vowels /ɛ, ɑ, a, o/ (table IV) 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 F1 of the vowel /ɛ/, and second order F1 of the vowel /a/ and F2 of the vowel /ɑ/ (table IV). From this we could conclude that the F1 frequency of fast spoken vowels is higher than the F1 frequency of tokens spoken at a normal rate. The difference is uniform and irrespective of vowel identity.

Correlations between speaking rates of the the zero order (mean value) component were high and statistically significant ($p \leq 0.1\%$, table V). 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 F2 (table V, only second order is shown). The correlation coefficients of F2 were higher than those of F1 in most vowels. The correlation coefficients decreased with increasing order but still remained quite high (up to $r=0.74$ for /o/, third order F2, 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.

II.E Relation between polynomial coefficients and vowel duration

The polynomial coefficient values found for the formant tracks were correlated to 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 F1 of the vowels /ɛ, ɑ, a/ ($r \approx 0.33$-0.52, $p \leq 0.1\%$). These coefficient values were almost as high as the between-speaking-rate correlation coefficients (cf. table V). The correlations between duration and second order components of F1 implied a decrease in 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 both speaking rates.

II.F Effects of context

A subset of the tokens of the most numerous vowels /ɛ, ɑ, 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 (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.

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 IV). The second order Legendre coefficients of the F1 tracks of the vowels /ɛ, ɑ, a/ might be an exception. The tokens of these three high F1 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 uniform higher F1 than the normal rate tokens ($p \leq 0.1\%$ for /ɑ, o/, zeroth order). The vowel /o/ also showed a slightly higher F2 in the fast rate tokens (42Hz $p \leq 0.1\%$, zeroth order). The between-speaking-rate correlation coefficients of the Legendre coefficients 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 (table V).

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 II.C.4.). Still, only few correlation coefficients were statistically significant ($p \leq 0.1\%$, fast rate F1: second order coefficient of /ɛ, ɑ, a/ and slope at 1/4 of /ɛ/) or larger than the corresponding correlation between speaking rates (c.f. table V). An exception was the second order Legendre coefficients of the F1 tracks of the fast rate tokens of the vowels /ɛ, ɑ, a/. These correlation coefficients were higher ($|r| \approx 0.60$-0.75, $p \leq 0.1\%$) 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.

II.G Effects of stress

The previous analyses were repeated on token pairs of the vowels /ɛ, ɑ, 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%).

For the F1, zeroth and (negative) second order Legendre coefficient values of the stressed tokens of the high F1-target vowels /ɛ, ɑ, a/ were higher than those of the unstressed tokens at both rates ($p \leq 1\%$ for vowels pooled). 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, both for zeroth order (5%) and second order coefficients (25%). The slopes of the F1 tracks of stressed tokens were generally steeper than those of unstressed tokens. For the F2, the second order coefficient values and track slopes were often not statistically significant for the stressed 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 (zeroth order, $p \leq 0.1\%$, stressed /ɛ, a/, unstressed all individual vowels).

Generally, correlation coefficients were higher in stressed tokens than in unstressed tokens, both for vowel duration and formants between speaking rates and between formants and vowel duration. 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.

III Discussion

The results found here 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. As these will be published elsewhere (Van Son and Pols, submitted), we will not discuss them any further.

II.A 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 F1 frequency is higher in vowels spoken at a fast rate than in vowels spoken at a normal rate. This rate-dependent rise in F1 frequency was found irrespective of vowel identity. It was also uniform, that is, limited to the zeroth order Legendre polynomial (i.e. mean formant value). This means that the equivalent results found by Van Son and Pols (1990) for "static" measurements, in which method Average is identical to using the zeroth order coefficient, cannot be attributed to a change in formant track shape due to speaking rate.

III.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 (/ɛ, ɑ, 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 second order Legendre coefficients ($0.33 \leq |r| \leq 0.52$). 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 track parameters ($|r| \leq 0.91$, table V). This difference in correlation indicated that duration is not a major determinant of overall vowel formant track shape in read speech. However, the corresponding correlation coefficients between speaking rates for the second order Legendre polynomial were not larger (i.e. $0.46 \leq |r| \leq 0.55$, table V) than those with duration. This indicates that the formant track excursion size of the F1, as measured by the second order Legendre coefficient, is indeed (cor-) related to vowel duration in a way predicted by the target-undershoot model (i.e. shorter duration combined with more level formant tracks). The size of the correlation coefficients were comparable to those resulting from textual factors (i.e. the between speaking rate correlation). But again, the absolute size of the effect of duration on track shape is minimal, generally explaining less than a quarter of the variance observed.

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

III.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 from 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 F1). 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 F1 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 /ɛ, ɑ, a/ for the F1 tracks and on the back vowel /o/ for the F2 tracks (see section I.B.). For fast rate tokens we found an increase in the correlation between the second order Legendre coefficient of the F1 tracks of the vowels /ɛ, ɑ, a/ and vowel duration. This suggests that the constraints on F1 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 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, except that in this context the F2 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 F2 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 puts an upper limit to the importance of sentence stress and context in determining the result of speaking rate increases.

III.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, which predicts more reduction (both static and dynamic) in vowel articulation with a faster speaking rate. 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 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 did find evidence that they 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 powers are 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.

Acknowledgments

The authors wish to thank Dr. A.C.M. Rietveld of the Catholic University of Nijmegen, the Netherlands, for performing the sentence accent labelling 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. This research project is part of the Dutch national program "Analysis and synthesis of speech" funded by the Dutch program for the Advancement of Information Technology (SPIN).

References

Abramowitz, M., and Stegun, I.A., (1965). *Handbook of mathematical functions* (Dover publications, Inc., New York NY, 9th printing).

Benguerel, A-P., and McFadden, T.U. (1989). "The effect of coarticulation on the role of transitions in vowel perception", *Phonetica* **46**, 880-96.

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.

Churchhouse, R.F. ed. (1981). *Handbook of applicable mathematics*, vol.III: *Numerical methods* (John Wiley & Sons), pp. 194-201.

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.

Ferguson, G.A. (1981). *Statistical analysis in psychology and education, International student edition* (McGraw-Hill, 2nd printing), pp. 381-406.

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.

Koopmans-van Beinum, F.J. (1990). "Spectro-temporal reduction and expansion in spontaneous speech and read text: the role of focus words", *ICSLP 90 proceedings* 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 N.Y.), 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.* **88** Suppl. 1, paper 6SP13 (A).

Nord, L. (1987). "Vowel reduction in Swedish" in *Papers from the Swedish phonetics conference* edited by Olle Engstrand (Upsala), pp. 16–21.

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* **2**, 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.

Van Son, R.J.J.H., and Pols, L.W.C. (submitted). "Formant movements of Dutch vowels in a text, read at normal and fast rate", (*J. Acoust. Soc. Am.*).

Vogten, L.L.M. (1986). "LVS speech processing programs on IPO-VAX11/780", *Manual* 67, Institute for Perception Research, Eindhoven,The Netherlands.
Weismer, G., Kent, R.D., Hodge, M., and Martin, R. (1988). "The acoustic signature for intelligibility test words", *J.Acoust.Soc.Am.* **84**, 1281-1291.
Willems, L.F. (1986). "Robust formant analysis", *Annual progress report* 21, Institute for Perception Research, Eindhoven, The Netherlands, 34-40.