A METHOD FOR THE DYNAMIC DETERMINATION OF ACOUSTIC VOWEL CONTRAST

Florien J. Koopmans-van Beinum and Rob P. de Saint Aulaire

1. INTRODUCTION

The great variability in the realization of vowel phonemes when produced by the same speaker, but in different speech situations, plays an embarrassing role in speech technology. Vowels in connected speech rarely reach their target position (the intended phoneme) as defined in isolated word and isolated vowel production. In speech synthesis we badly need this variability for the cause of intelligibility as well as for naturalness, but in automatic speech recognition it is an annoying phenomenon with which we do not know how to cope. It is known that the degree of acoustic contrast between the vowels in a speaker's vowel system is dependent on various factors, partly global and partly local, but it is not clear as to how far all these factors are interdependent, if at all.
In literature (for a detailed overview see Koopmans-van Beinum, 1980) we can find a large number of factors that are believed to be responsible for the variability and the reduction of acoustic vowel contrast:
a) acoustic-phonetic factors:
- speech rate: contrast between the vowels within a vowel system decreases when the speech rate increases;
- stress: contrast between vowels within a vowel system decreases in case of unstressed syllables;
- intonation: a 'flat' intonation normally goes with less vowel contrast than a varying intonation;
- local context: neighbouring consonants affect vowels and their mutual contrasts;

b) socio-phonetic factors:
- speech situation: vowel contrast decreases when manner of speaking and choice of words (as in normal conversation) are more free;
- speech training: an untrained speaker, i.e. a non-professional speaker, reveals less vowel contrast when speaking than a trained, professional one;
- sex: men are believed to make less contrast between their vowels than women;

c) linguistic factors:
- grammatical word class: function words show less vowel contrast than content words;
- word frequency: in high frequency words vowel contrast is less than in low frequency words;
- position of the syllable: the occurrence of more or less vowel contrast is dependent on the number of syllables within a word, and on the position of the syllable in the word;
- language structure: reduction of vowel contrast is reported much more in so-called 'stressed-timed' languages than in so-called 'syllable-timed' languages.

Although this enumeration may not be exhaustive, it gives in our opinion the main factors that in speech are responsible for not reaching the intended phonological target positions of the vowels. (See for more details Koopmans-van Beinum, 1980; Koopmans-van Beinum and Harder, 1982/83; De Graaf and Koopmans-van Beinum, 1984; Labov, 1966; De Schutter, 1975; Booij, 1976, 1981, 1982).

As far as the acoustic-phonetic factors are concerned quite a lot of research has been done with respect to the description of vowel contrast in various speech situations. However, the relations between these factors and more specifically their hierarchical structure have been studied only fragmentarily yet. Lindblom (1963) for instance postulates that duration is the main determinant of vowel reduction, whereas Delattre (1969) claims stress and speech rate to be primary determinants with duration as a product of stress and speech rate and therefore a secondary determinant. Gay (1977) and Den Os (1985) both show that an increase of speech rate not necessarily affects the formant frequencies of the vowels. Furthermore Koopmans-van Beinum (1980) indicates a different relation between stress and vowel duration for read texts as compared to texts with a free choice of words (retold story or free conversation). Also from perceptual studies on stress (e.g. Van Katwijk, 1974; Rietveld, 1983; Rietveld and Koopmans-van Beinum, to appear) the relation between loudness, intonation, speech rate, and vowel contrast reduction turns out to be a very complicated one.

In order to reach a better understanding of the relations and the hierarchical structure of the great variability of vowels, it is deemed necessary in our approach to the speech signal to make a distinction between 'global' factors (socio-phonetic aspects as speaker, speech situation, etc.) affecting this variability, and 'local' factors (acoustic-phonetic and linguistic aspects within the neighbouring context).

We therefore decided to start a large project in order to develop and apply strategies to make optimal use of acoustic, socio-phonetic, and if possible linguistic information with respect to the variability in the realization of vowel phonemes. This will be done by means of a semi-automatic method for dynamic vowel analysis and cumulative data processing in three phases:

a) Any speech fragment from any speaker will be subjected to a dynamic acoustic-phonetic analysis to provide information on global aspects as mentioned above about the present vowel system (sex of the speaker, overall speech rate, degree of vowel contrast, etc.). Moreover the acoustic parameter values in the dynamic vowel analysis provide the possibility to define the moment when the global measure for acoustic system contrast (ASC) stabilizes. This indicates the duration of the speech sample needed for defining the ASC-value (and other global measures), and for dynamically adjusting it, if a moving window is used.

b) Subsequently local measures of acoustic vowel contrast or degree of reduction and variability will be developed based on acoustic-phonetic parameters as fundamental frequency, formant frequencies, bandfilter values, vowel duration, amplitude.

c) Finally the results of a) and b) will be used in various applications, as for instance in the labelling of segments as specific vowel phonemes, merely by using the local acoustic parameter values combined with global contrast measures and general information on the present vowel system, and defining the hierarchical structure of factors influencing the variability in vowel phonemes.

This article reports on our very first steps within this project, viz. the development of a method for the dynamic determination of the global measure of acoustic system contrast.

2. DESIGN OF A DYNAMIC ANALYSIS AND DATA PROCESSING METHOD

As the aim of the present subproject is to develop a (semi-) automatic procedure of data processing, two parallel methods had to be compared: a) the traditional method making use of manual segmentation of vowels in the digitized speech fragment by means of a speech editor called SESAM (Buiting, 1981), followed by a dynamic acoustic-phonetic vowel analysis using a spectral analysis program called QQ (Weenink, 1986), and b) a (semi-) automatic method by carrying out a dynamic acoustic-phonetic analysis with QQ, initially on all speech frames, followed by an automatic vowel segmentation realized by a program called KLUIT (De Saint Aulaire, 1986). Finally both methods end up in a data processing program called CORBER (based on formant frequencies, or CORBF if based on bandfilter values) which calculates in a cumulative way the acoustic system contrast measure ASC (Koopmans-van Beinum, 1980; De Saint Aulaire, 1986)). This ASC measure is defined by the total variance of all vowels in the present vowel system, based on frequencies of the first (F1) and second formant (F2), (transformed in 100 * 10log Hz), using the formula:

$$ASC = \frac{1}{N}\sum_{j=1}^{N} (\vec{V}_j - \vec{C})^2$$

in which $\vec{V}_j$ = the 2-dimensional vector of vowel j in the F1/F2-plane,

$\vec{C}$ = the 2-dimensional vector of the centroid C, and

N = the number of vowels in the vowel system.

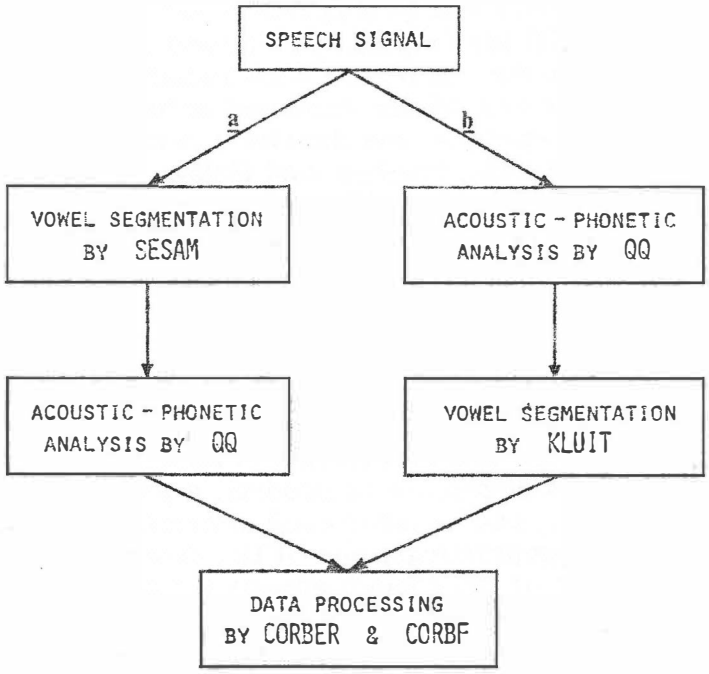

Fig. 1. Schematic display of the two methods of analysis and data processing.

In fig. 1 we display schematically the two procedures a) and b) built up from five different blocks, each of which will be discussed here in brief, referring to the actual speech material used in both procedures.

2.1. Speech fragment

Any speech fragment can be selected from audio recordings and can be stored in digitized form (sample frequency of 10 kHz). In the first phase we used recorded speech material of the same trained male speaker as in Koopmans-van Beinum (1980). This provided us with the possibility to compare the results of the present procedures with previous results. However, an important difference emerges: in the present speech material we used all vowels in the chosen speech fragment, and in the order in which they occurred. Moreover measurements were carried out dynamically with ten millisecond steps. This means that frequency of occurrence of all vowels in normal running speech got the intention it deserves, and that the duration of each occurring vowel weighs proportionally in the calculation of the acoustic system contrast. In the former study ten items of each vowel were used and measured only at one point more or less in the middle of the vowel.
An accidental advantage of this present 'weighing' procedure is the fact that it is no longer necessary to 'label' the vowel segments, i.e. we no longer need to know which vowels the speaker intended to say. The acoustic system contrast ASC of a speech fragment of a specific speaker is defined now by the total variance of all vowels, c.q. of all analyzed 10 ms vowel frames, just as they occur in the speech fragment. The moment at which this ASC stabilizes actually defines the length of the speech fragment needed for the determination of the ASC for that specific speaker in that specific speech situation. As to how far length of fragment depends on speaker, on speach situation, and on language is one of the research questions of the project as a whole. In the present article, however, we will only report on the results concerning one speaker in two speech situations: free conversation (a 30 sec fragment) and read text (a 10 sec fragment). The speech fragments were selected from existing recordings. Our decision to confine ourselves to a 30 sec fragment is based on literature indicating that variables concerning the distribution of spectral energy stabilize within that period of time (Li, Hughes, and House, 1969; Zahorian and Rothenberg, 1981). Our choice of only a 10 sec fragment of read text is defined by the results obtained from the free conversation fragment and the need to confine the material.

2.2. Manual vowel segmentation

Since the present study was meant to compare the traditional method of vowel segmentation and calculating ASC with a (semi-)automatic and cumulative one, it was necessary to produce a phonological transcription of the speech fragments, and to label each occurring vowel item as one of the twelve Dutch monophthongs, one of the three Dutch diphthongs, or as schwa. By means of the speech editing program, all vowel items in the digitized speech fragments were isolated in such a way that the starting-point of the vowel was considered to be the place were the formant pattern of the vowel was clearly visible on the oscillogram for the first time, and the end was taken to be the point were the specific formant pattern disappeared. In case of adjacent voiced consonants only those successive samples were segmented that did not display any auditory nor visually observable consonant information.
Once the vowel segments were isolated, their durations were of course known as well. From the 30 sec fragment of free conversation 121 vowels

could be selected with an average duration of 68.66 ms. From the 10 sec fragment of read text 57 vowels were segmented with an average duration of 71.12 ms (for more durational information see Table 1). A complicating aspect, especially in the free conversation fragment, is the occurrence of 'pause filling' schwa segments (two items with an average duration of 426.1 ms, not included in the 121 vowels). We decided to leave them out, apart from one ASC calculation together with the intended schwa phonemes, in order to state their influence in normal conversational speech.
To test whether the distribution of vowels in this specific fragment of running speech could be considered to be representative for the frequency of occurrence in spoken Dutch, a comparison was made with data from Eggermont (1956). Computation of the Spearman rank correlation coefficient of the vowel occurrence frequencies of both Eggermont's and our free conversation fragment, turns out to be highly significant ($R_s=0.799$, $p<0.01$).

Table 1. Outline of the samples building up the speech fragments of free conversation and read text in parts of silence, consonants, and vowels.

	free conversation			read text		
	n of samples	dur. in sec.	%	n of samples	dur. in sec.	%
silence	89809	8.98	29.94	9572	0.96	9.57
consonant	118595	11.86	39.53	49888	4.99	49.89
vowel	91596	9.16	30.53	40540	4.05	40.54
total	300000	30.00	100.00	100000	10.00	100.00

2.3. Acoustic-phonetic spectral analysis

Each vowel file has been analysed dynamically in 10 ms steps (window size 25.6 ms) by means of the program QQ (Weenink, 1986) using a filter order 12 as a standard.
Apart from a number of other data, not relevant for this study, the program QQ provided us with:
- fundamental frequency (F0) determined on the basis of Duifhuis et al. (1982);
- formant frequencies determined by optional methods; in our case we used Prony's method for LPC-analysis;
- bandfilter values: a bandfilter analysis of the FFT amplitude spectrum is carried out with filter specifications based on Sekey and Hanson (1984).

The resulting data are stored in analysis files consisting of successive records, each of them containing the analysis results of one 10 ms vowel frame. In this way all kinds of selections and calculations can be carried out in the succeeding data processing programs.

2.4. Automatic vowel segmentation

With respect to the development of an automatic procedure of data processing, one of the main problems to overcome is the segmentation of vowels from the speech fragment (cf. Kasuya and Wakita, 1979). We therefore designed procedure b) (see fig. 1) in which the spectral analysis precedes the vowel segmentation. The output records are selected as 'vowel' on the basis of three criteria:

- F0-criterium: each data record including F0 = 0 has to be rejected;
- high/low ratio (H/L): the definition of low and high frequency areas in literature is not uniform: Weinstein et al. (1975) use L = 0-900 Hz and H = 3700-5000 Hz; Kasuya & Wakita (1979) use L = 0-500 Hz and H = 3800-5000 Hz, whereas for Dutch speech material Rietveld (1983) defines L = 262-2230 Hz and H = 5575-11150 Hz. In the present study we used the filters 1-6 for the low frequency area (92-856 Hz) and the filters 13, 14, and 15 for the high frequency area (2549-4239 Hz), since filter 16 turned out not to be reliable in all cases. So if the ratio H/L>1 then the data record is rejected as a vowel record.

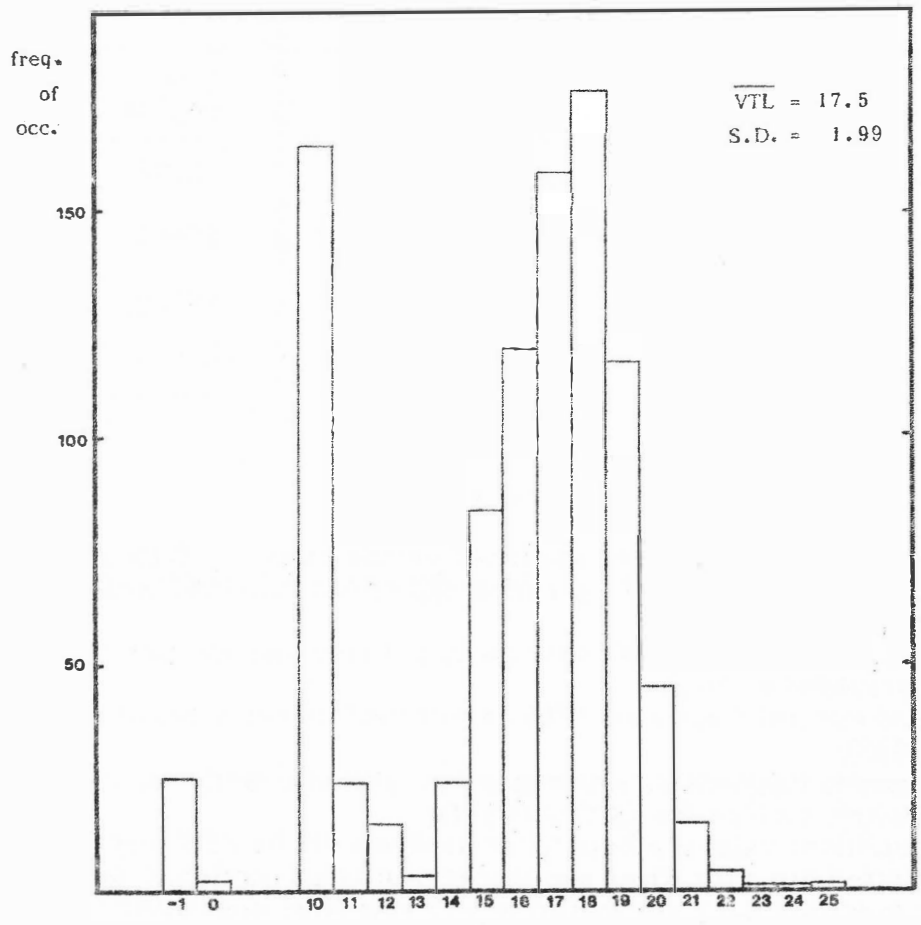

Fig. 2. Distribution of vocal tract length VTL values calculated for the read speech material of one trained Dutch speaker.

- vocal tract length VTL: based on the analysis results of QQ this program calculates also the VTL per record (Wakita, 1977). Considering the formant frequencies and VTL together revealed that in case of low (nasal) F1 the VTL showed very unreal values (0.0 or -1.0 cm), whereas for records with an extreme high F1 value (e.g. F1>1500 Hz) the calculated VLT attained to about 10 cm. All other records display a more or less normal distribution of VTL values (for the read speech sample see fig. 2). Although this VTL criterium needs some more refinement, we obtained satisfactory results in this study by using the criterium that each vowel record had to attain a VTL value of:

$$VTL - 0.5*s.d. \leq VTLx \leq VTL + 1.0*s.d.$$

in which VTLx = the VTL of data record x.

Within the automatic vowel segmentation program the following hierarchy of criteria is used: 1) a first selection is done based on the F0- and the high/low ratio criterium; 2) a second selection is done based on the VTL-criterium, applied to the remaining records.

2.5. Data processing programs

Both procedures a) and b) (see fig. 1) end up in a set of data processing programs.
CORBER calculates cumulatively (in this case record after record) the mean values with variance of the fundamental frequency, of the first four LP-formants, and of the 16 bandfilter values. During the processing the mean F1, F2, mean bandwidths of each formant, mean level of each bandfilter, F0, and the ASC is stored in an output file, together with the deviation of the new ASC compared to the preceding ASC value, each time when a record is closed. By means of a plot program these data can be displayed graphically. At the end of the processing the output of CORBER consists of the final mean values with variance of the parameters mentioned above, and the total number of processed records (=the number of 10 ms vowel frames).
CORBF carries out a principal component analysis on the bandfilter values. The output data consist of the calculated eigenvalues with cumulative percentages of variance accounted for, the eigenvectors, the mean levels of the 16 bandfilters, the position of those mean levels in the various subdimensions, and the total number of processed records. The program CORBER provides the possibility of cumulatively processing the acoustic system contrast (based on formants and therefore henceforth called FASC), and to define the moment when the FASC value stabilizes. The program CORBF provides an acoustic system contrast measure based on bandfilter variance (henceforth called BASC). Since this BASC value has to be comparable to the FASC value in order to state its usefulness and its reliability, we opted for using only the variance accounted for by the first two eigenvalues.

3. RESULTS FROM FREE CONVERSATION AND READ SPEECH MATERIAL

Since the main aim of the subproject described here consisted in the development of data processing methods, we did not apply exactly the same proce-

dure to the speech fragment of free conversation and to the read text fragment. For the free conversation fragment we used only the manual procedure (see fig. 1, a), since our first purpose was to develop cumulative data processing methods for calculating acoustic system contrast, and to compare the results with those from our preceding study (Koopmans-van Beinum, 1980). As for the read text fragment we made use of the manual as well as of an automatic procedure, since the purpose here was to develop an automatic method and to compare the results with those from the manual one applied to the same speech material.

3.1. Results from methods applied to free conversation

Since we wanted to compare the results of our method for dynamic determination of vowel contrast as good as possible with the former data, all vowel segments excised from free conversation and stored with information about their identity had to be divided into stressed and unstressed. Ten colleagues in the institute were asked to listen to the speech fragment carefully, several times if necessary, and to indicate the stressed syllables in the transcribed text in front of them. The vowels of those syllables indicated by seven or more listeners as being stressed, were stored as +stress, all other vowels as -stress.
In this way 16 out of the 123 excised vowels were labeled +stress. Since we believed this number to be too small to calculate cumulatively the ASC value for the stressed vowels separately, we decided to make the distinction between unstressed vowels only (107), and all vowels together (123).

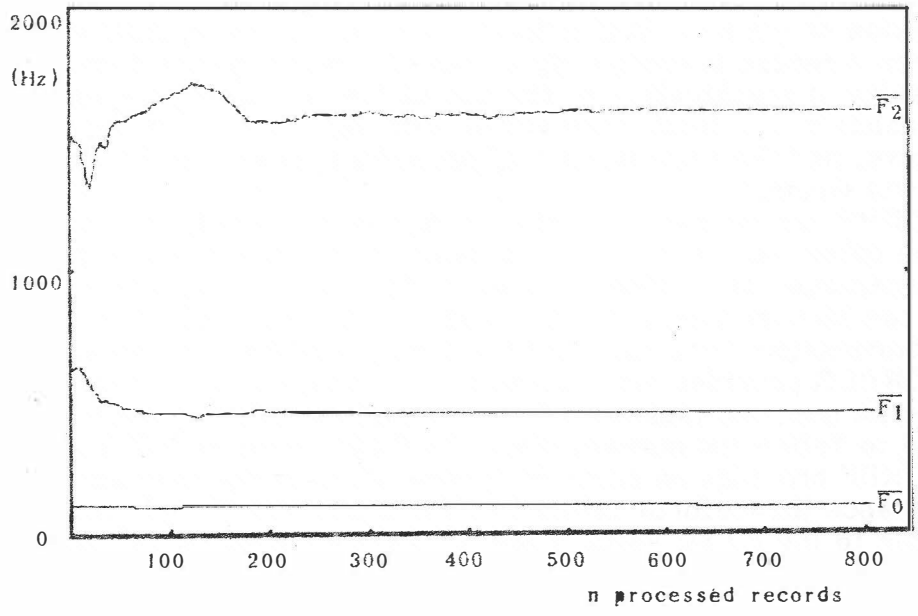

Fig. 3. Cumulatively defined mean values of F0, F1, and F2 within 10ms-records of vowels from a 30 sec speech fragment of free conversation.

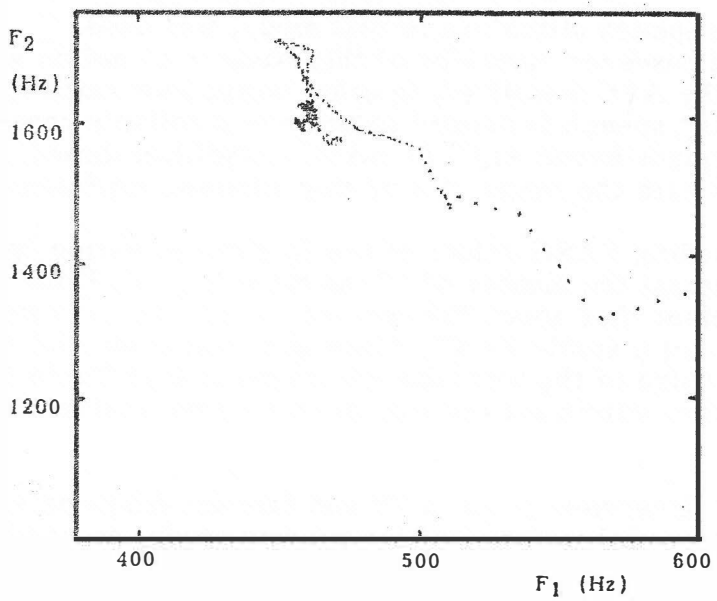

Fig. 4. Progressing centroid position based on cumulatively defined mean F1 and F2 values (see fig. 3).

After the dynamic spectral analysis, the resulting data are used to calculate cumulatively the mean values of F0, F1, and F2 of all vowel records. Fig. 3 displays graphically these parameters related to the number of vowel records processed thus far.
In fig. 3 we can see that the fluctuations in mean F1 values decrease very soon (after 150 records = 1.5 sec of vowel material the changes become less than 10 Hz), whereas the mean F2 attains its stable position after about 250 records. It is clear that the centroid, being the progressively calculated overall mean values of F1 and F2, after some starting excursions rather quickly settles down in a well defined area of the F1/F2 plane (see fig. 4).
With respect to the cumulative calculation of the formant-based acoustic system contrast, we made for comparison purposes various distinctions within the analysed vowel material: all vowels together and unstressed vowels only, and within these two main groups we distinguished three groups:
1) monophthongs + diphthongs
2) monophthongs + diphthongs + intended schwa
3) monophthongs + diphthongs + intended schwa + pause filling schwa
Furthermore for maximum comparison with the former data we added group of monophthongs only, from which the acoustic system contrast is calculated with the same overall speaker centroid value as used in Koopmansvan Beinum (1980). Yet it should be borne in mind that in the former speech material calculations are made over only one measuring point in the stable part of the vowel segment, with probably a maximum of contrast, whereas in the present material we averaged all dynamic analysis data of the entire vowel. Table 2 displays an overview of the results when calculating mean F0, mean F1 and F2 and ASC over the various distinctions mentioned above, together with the results from our previous research. Moreover, when making a further comparison with the previous results, we have to remember that the previous results are based on the twelve Dutch monophthongs, ten items of each, and that the overall speaker centroid, based on

data of all speech situations in that study, was used.
One of the research questions of this study is to define whether, and if so, when the ASC stabilizes, in other words how much speech material from running speech is needed to achieve a reliable result in terms of FASC (formant-based ASC) or BASC (bandfilter-based ASC). This would provide us with the range of a moving window, sufficient for our further research.
The proceeding FASC values of the free conversation speech fragment are plotted against the number of 10-ms steps (fig. 5). From this graph it becomes clear that about 250 records (= 2.5 sec of vowel material) are needed to get a stable FASC. Since the vowel material turned out to be about one third of the total speech fragment (see Table 1), we might conclude that within six seconds of free conversational speech the speaker

Table 2. Overview of mean F0 and formant frequencies, and of ASC values of vowel material of the present study, grouped in various classes. From previous research comparable results have been added. (* = calculated with respect to the overall speaker centroid)

	all			-str.			all	+str.	-str.	all sp. cond.
	1	2	3	1	2	3				
mean F0	108	108	107	107	106	106	-	-	-	-
mean F1	468	461	458	454	449	446	438	446	430	446
mean F2	1585	1593	1612	1603	1611	1633	1531	1540	1522	1554
(F)ASC	141	131	124	128	116	108	272*	264*	174*	-
	De Saint Aulaire, 1986						Koopmans, 1980			

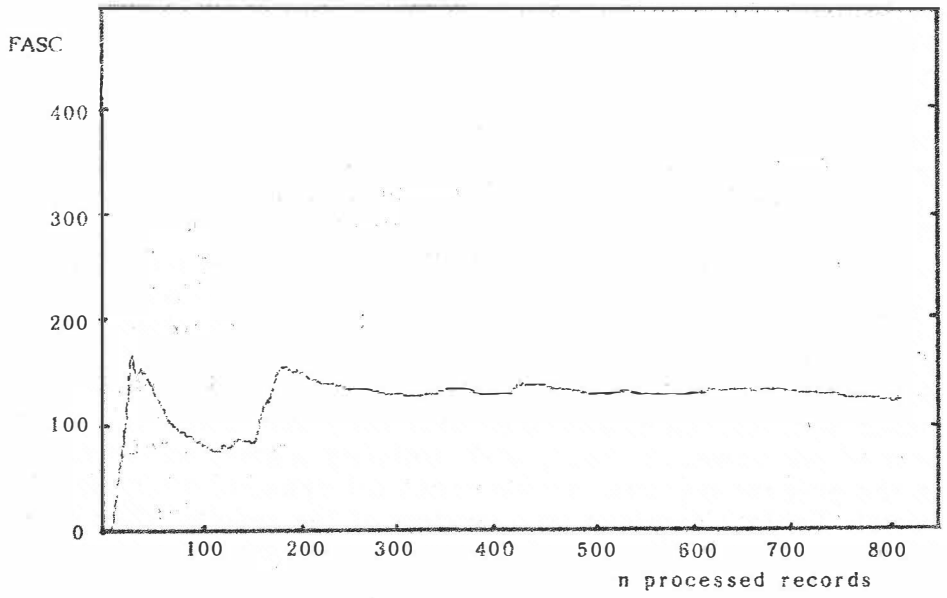

Fig. 5. Cumulatively calculated FASC value processed over the vowels from a fragment of free conversation.

provides us with a global measure for acoustic contrasts within his vowel system.
Another research question of the present study is to explore whether the FASC measure, which is the total variance within a vowel system based on formant frequencies of F1 and F2, can be related to a vowel contrast measure based on the variance of bandfilter values BASC. In order to compare both measures optimally we decided to use only two eigenvalues in our principal component analysis, since for the FASC measure we used the first two formants only.
In table 3 we give the two measures for acoustic system contrast together with the cumulative percentage of variance accounted for by the first two eigenvalues in case of BASC. A further discussion concerning the sense of these values will arise in the next paragraph, where we will compare them with the results of manual and automatic analysis of read speech material.

Table 3. Overview of formant-based and bandfilter-based ASC values, and the cumulative percentage of variance accounted for by the first two eigenvalues for BASC of vowels from free conversation.

	all vowels	unstressed vowels
FASC	124	108
BASC	400	341
cum. % of $e_1 + e_2$	80.2	77.7

3.2. Results from methods applied to read text

For the read text we made use of a manual, as well as of an automatic procedure (see fig. 1), in order to compare the results of both methods on the same speech material. Moreover the results from read speech material (being running speech as well) from the same speaker gives us a good possibility to test our methods, since this speech condition has also been used in the previous study (Koopmans-van Beinum, 1980).
In the first instance we followed the same procedure as for free conversation, described in 3.1. For the read text a fragment of only 10 sec was used, providing 57 vowels of which only seven were considered to be stressed. Again the dynamic spectral analysis and the calculation of formant-based and bandfilter-based acoustic system contrast were processed. Fig. 6 displays the cumulatively defined mean values of F0, F1, and F2 of all vowel records. Here again (see fig. 7) the centroid, based on these mean F1 and F2, after some starting excursions settles down in the same well defined area of the F1-F2 plane. A comparison with fig. 3 reveals, however, that the excursions are larger for read text than for free conversation.
So next we calculated cumulatively the FASC values for read text and plotted them against the processed records in fig. 8. If we compare this figure with the similar one on free conversation (fig. 5), we first have to notice the difference in scaling over the total number of records because of the shorter read speech fragment. We have to conclude that indeed the FASC value for read text is much higher than for conversational speech.

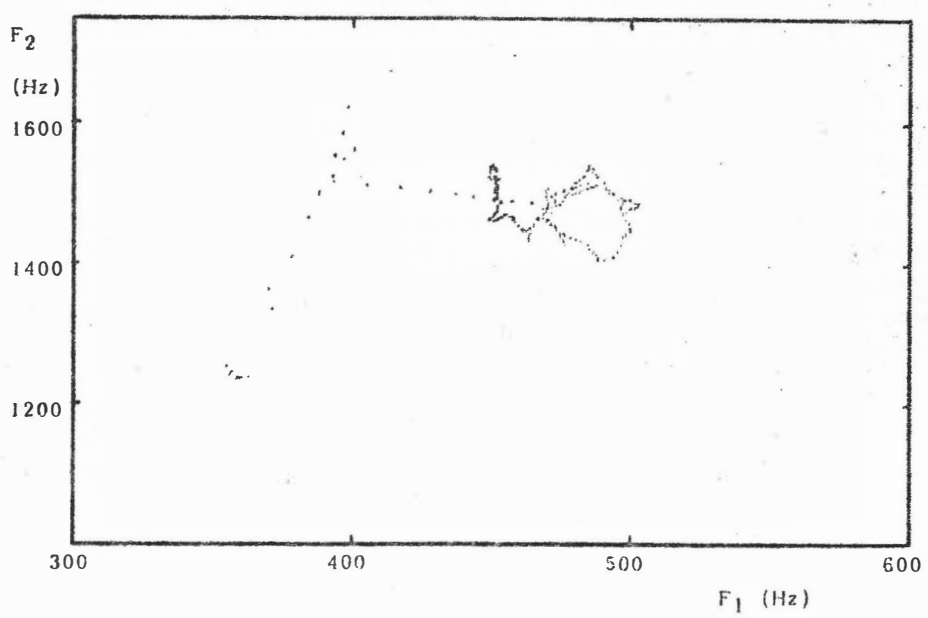

Fig. 6. Cumulatively defined mean values of F0, F1, and F2 within 10ms-records of vowels from a speech fragment consisting of 10 sec read text.

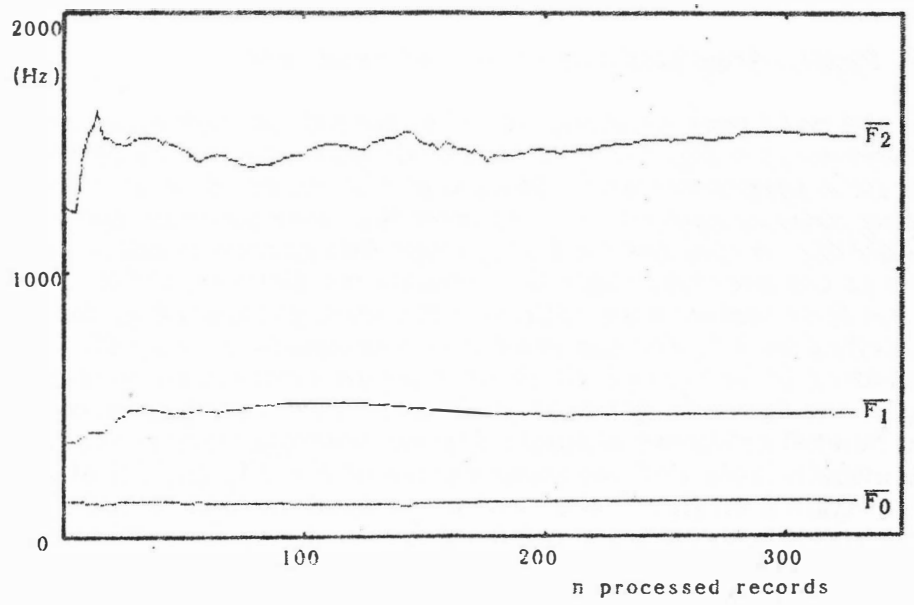

Fig. 7. Progressing centroid position based on cumulatively defined mean F1 and F2 values (see fig. 6) for vowels from a 10 sec fragment of read text.

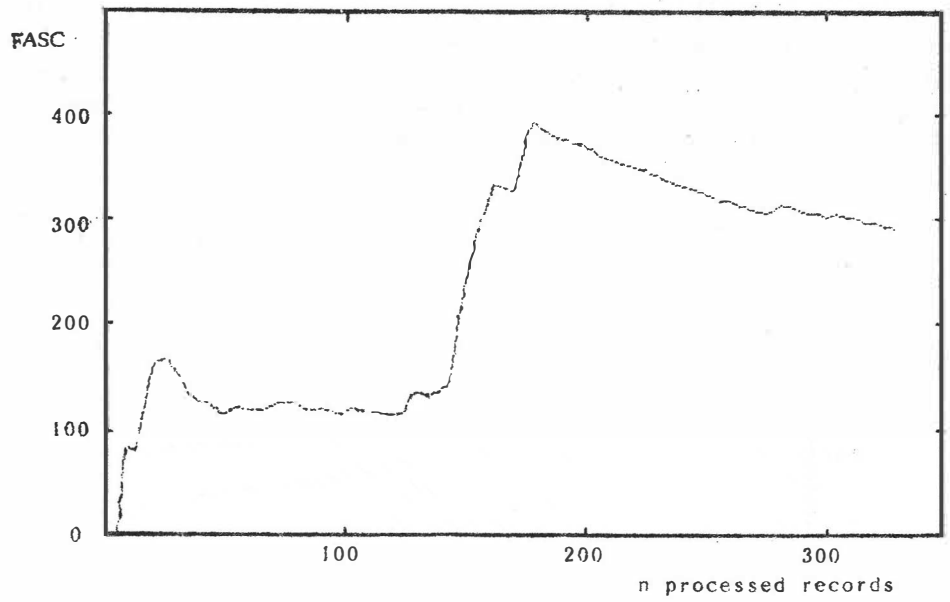

Fig. 8. Cumulatively calculated FASC value processed over the vowels from a fragment of read text.

Actually the speech fragment should have been a little bit longer to secure a sufficient range of the moving window. A striking point in the plot is the high jump between record 150 and record 200. Inspection of the read text, however, shows us that the fragment starts with a rather long subordinate clause, and that exactly at this point the main clause starts. Before that point the proceeding FASC value is more similar to that of free conversation. Table 4 gives the FASC (formant-based) and the BASC (bandfilter-based) values over the whole vowel material of the read text, divided in unstressed vowels and all vowels.

Table 4. Overview of formant-based and bandfilter-based ASC values, and the cumulative percentage of variance accounted for by the first two eigenvalues for BASC of vowels from read text.

	all vowels	unstressed vowels
FASC	292	276
BASC	562	485
cum. % of $e_1 + e_2$	82.4	80.7

On the same read speech material we carried out the automatic segmentation procedure as well, so that we can compare the results optimally. In the manual methods we excised 57 vowels, 39 of which (= 68.4%) were indicated directly as a vowel, whereas 14 (= 24.6%) were indicated partly, i.e. a few records within a whole series were not indicated as a vowel. Besides 5.9% of the whole 10 sec. read speech fragment is indicated automatically as being a vowel whereas manual segmentation considered them to be consonantal.

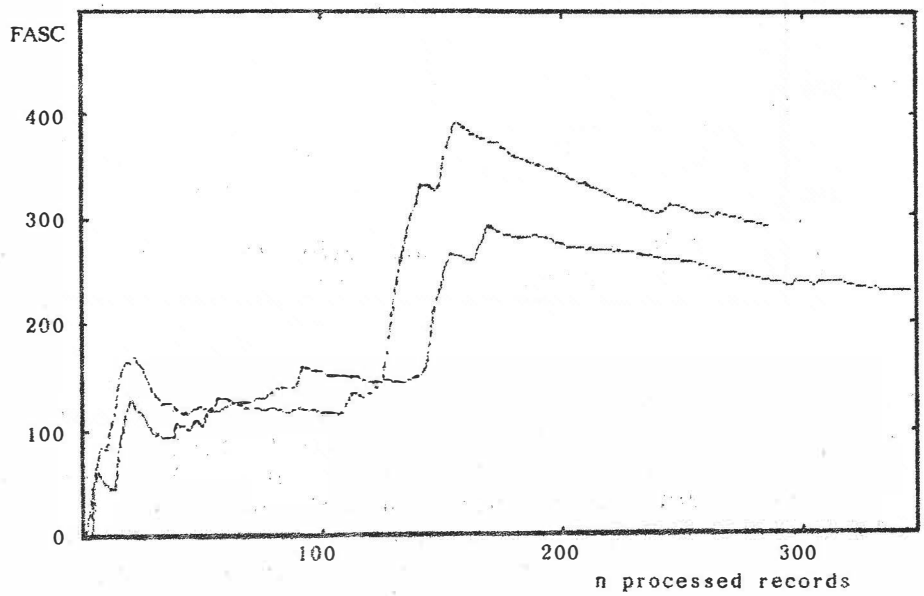

Fig. 9. Cumulatively calculated FASC values for manually and automatically segmented vowels from read text.

Table 5. Overview of the resulting output of the FASC and BASC calculations for manually and automatically segmented vowels from read text.

	automatically	manually
mean F0	120	123
mean F1	417	453
mean F2	1468	1520
FASC	230	292
BASC	466	562
cum. % $e_1 + e_2$	79.4	82.4
n records	398	332

For further calculations of formant-based and bandfilter-based acoustic system contrast we used all records of excised vowels by means of the automatic segmentation program. In fig. 9 we plotted the cumulatively calculated FASC for the manual and the automatic procedure, whereas in table 5 we displayed the calculated values from both methods together. From fig. 9 and table 5 it is clear that the results from the two procedures are very similar, although the automatic procedure processed 66 more records than the manual one. Moreover all values from the automatic method are somewhat lower than those from the manual procedure. In fig. 9 we can see a great similarity between both curves, albeit a little bit shifted. In the automatic FASC curve we also find that high jump, probably caused by the different types of clauses, only less high since the preceding part is longer. Which parts of the speech fragment are exactly causing the shift has not yet been investigated.

4. CONCLUSIONS

In table 6 an overview of the results on acoustic system contrast in the two speech situations is given, together with the results from our earlier research involving the same speaker. As for the earlier data, especially for the unstressed vowels, it should be kept in mind that at that time no schwa sounds and no diphthongs were included. Especially the high number of schwa sounds in the Dutch language will cause a much lower ASC when we are processing the programs automatically. In fact this tendency is clear from table 6.

Table 6. Overview of formant-based (FASC) and bandfilter-based (BASC) values of acoustic system contrast together with the ASC values as given in Koopmans-van Beinum (1980)in various speech situations for one male Dutch speaker.
Correlation coefficients between FASC and BASC are calculated over the marked (* or **) values. For more details see text.

condition		ASC	FASC		BASC	
		KvB '80	manual segm.	autom. segm.	manual segm.	autom. segm.
convers.	unstr.	174	108*		341*	
	all		124*		400*	
	stressed	264				
retold	unstr.	166				
	stressed	262				
read	unstr.	273	276*		485*	
	all		292*	230*	562*	466*
	stressed	343				
words	stressed	406**			715** (Pols '77)	
vowels		433				

As for the manual and the automatic segmentation method it turns out that the correlation coefficients between the comparable FASC and BASC values are .95 (* = the word-condition excluded) and .97 (** = the word-condition included), being both significant ($p < 0.01$).

Based on the results so far we can draw the following conclusions.

- The formant-based FASC, cumulatively processed on the output data of a dynamic acoustic-phonetic vowel analysis, compares favourably with the acoustic system contrast (ASC) values as processed on output data from static vowel analysis.
- The bandfilter-based BASC turns out to be a good alternative for the formant-based acoustic system contrast, since the results are very similar in the distinct speech situations. Our results are even confirmed by data from Pols (1977) who gives a total variance of 715 dB^2 between the mean spectra of the twelve Dutch vowels pronounced in monosyllabic words by 50 male speakers. The rank order of FASC and of BASC values in the distinct speech situations is exactly the same, as far as data are available, and correlation coefficients are significant.
- The in this study developed automatic procedure for processing running speech provides the possibility to define quickly and for extended speech material global reduction data in terms of acoustic system contrast.

In the near future especially the automatic procedure will be enhanced. Moreover it will be applied in various speech situations of more speakers and of other languages as well.

5. ACKNOWLEDGEMENT

The authors wish to thank Louis Pols, Tjeerd de Graaf, and David Weenink for their valuable help and comments with respect to this study.

6. REFERENCES

Booij, G.E. (1976). Klinker-reduktie in het Nederlands. Leuvense Bijdragen 65, 461-469.

Booij, G.E. (1981). Generatieve fonologie van het Nederlands. Utrecht: Het Spectrum.

Booij, G.E. (1981/82). Fonologische en fonetische aspecten van klinker-reductie. Spektator 11-4, 295-301.

Buiting, H.J.A.G. (1981). SESAM, Speech Editing System Amsterdam, IFA-report nr. 70.

Delattre, P. (1969). An acoustic and articulatory study of vowel reduction in four languages. International Review of Applied Linguistics 7, 295-325.

Duifhuis, H., Willems, L.F. & Sluyter, R.J. (1982). Measurement of pitch in speech: an implementation of Goldstein's theory of pitch perception. J. Acoust. Soc. Am. 71, 1568-1580.

Eggermont, J.P.M. (1956). De klankfrequentie in het hedendaagse gesproken Nederlands. De Nieuwe Taalgids 49, 221-223.

Gay, T. (1977). Effect of speaking rate on vowel formant movements. Haskins Lab. Status Report on Speech Research SR-51/52, 101-117.

Graaf, T. de & Koopmans-van Beinum, F.J. (1984). Vowel contrast reduction in terms of acoustic system contrast in various languages. Proc. Inst. of Phonetic Sciences Amsterdam 8, 41-53.

Kasuya, H. & Wakita, H. (1979). An approach to segmenting speech into vowel- and nonvowel-like intervals. IEEE, ASSP-27, 4, 319-327.
Katwijk, A.F.V. van (1974). Accentuation in Dutch. Diss. Utrecht University.
Koopmans-van Beinum, F.J. (1980). Vowel contrast reduction. An acoustic and perceptual study of Dutch vowels in various speech conditions. Diss. Universiteit van Amsterdam.
Koopmans-van Beinum, F.J. & Harder, J.H. (1982/83). Word classification, word frequency, and vowel reduction. Proc. Inst. of Phonetic Sciences Amsterdam 7, 61-69.
Labov, W. (1966). The social stratification of English in New York City. Center for Applied Linguistics, Washington, D.C.
Li, K.P., Hughes, G.W. & House, A.S. (1969). Correlation characteristics and dimensionality of speech spectra. J. Acoust. Soc. Am. 46, 1019-1025.
Lindblom, B.E.F. (1963). Spectrographic study of vowel reduction. J. Acoust. Soc. Am. 35, 1773-1781.
Os, E.A. den (1985). Vowel reduction in Italian and Dutch. PRIPU 10, 2, 3-12
Pols, L.C.W. (1977). Spectral analysis and identification of Dutch vowels in monosyllabic words, Diss. Vrije Universiteit Amsterdam.
Rietveld, A.C.M. (1983). Syllaben, klemtonen en de automatische detectie van beklemtoonde syllaben in het Nederlands. Diss. Nijmegen University.
Rietveld, A.C.M. & Koopmans-van Beinum, F.J. (to appear). Vowel reduction and stress. Speech Communication (submitted for publication).
Schutter, G. de (1975). De plaats van de [ə] in een fonologische beschrijving van het Nederlands. Leuvense Bijdragen 64, 173-202.
Sekey, A. & Hanson, B.A. (1984). Improved 1-Bark bandwidth auditory filter. J. Acoust. Soc. Am. 75, 1902-1904.
Wakita, H. (1977). Normalisation of voweis by vocal tract length and its application to vowel identification. IEEE Trans. ASSP-25, 183-192.
Weenink, D.J.M. (1986). QQ: Een programma voor analyse, resynthese en herkenning van klinkersegmenten, IFA-report nr. 82.
Weinstein, C.J., McCandless, S.S., Mondshein, L.F. & Zue, V.W. (1975). A system for acoustic-phonetic analysis of continuous speech. IEEE ASSP-23, 1, 54-67.
Zahorian, S.A. & Rothenberg, W. (1981). Principal-components analysis for low-redundancy encoding of speech spectra, J. Acoust. Soc. Am. 69, 3, 832-845.