A Classification of Vowel Systems and the Influence of Reduction Phenomena

by Jan G.Blom and Florina J.Koopmans - van Beinum

This paper is a combined account of an experiment conducted by Blom and its continuation by Mrs Koopmans.

In 1969 Blom read a paper covering the first part of the experiment for the Nederlands Akoestisch Genootschap[1]. He described at the time the application of principal component analysis and of factor analysis on the data of complete vowel systems. The material used for this experiment had been collected and measured by Mrs Koopmans[2]. Having extensively added to the original material[3] the above analyses were applied once more on all available data. A description of both parts of the experiment and a comparison of the results of them follows below.

When carefully articulated vowels produced by different speakers are presented in random order during a listening experiment a considerable amount of confusion follows. We know from experience that hardly any confusion exists when items of one speaker are presented. Apparently a listener adapts himself to the particular system of a certain speaker which he knows from earlier observation or to which he quickly gets attuned.

The vowel system is not built up out of fixed formant combinations but is the positional relationship of the formant combination to which the various vowels have to conform.

e.g. $F1\,[u] \approx F1\,[ü] \approx F1\,[i]$
$F2\,[u] < F2\,[ü] < F2\,[i]$

When considering the first and second formants only of the

vowel system of a speaker, we can represent this system as a point in a 24-dimensional space, with 12 F1-axes and 12 F2-axes as there are 12 monophthongs in the Dutch vowel system. If we do this for a number of speakers, the result will be a point-swarm in the above mentioned space.
The number of independent parameters which a listener needs as a minimum in order to classify the vowel system of a speaker is equal to the dimensionality of the subspace in which the point-swarm is imbedded.

In order to get some insight into the matter an investigation was made using material previously collected and measured for a comparative dialect study. There was a group of 40 speakers of which 10 male, non-dialect speakers,

10 female, non-dialect speakers,

10 male, Utrecht dialect speakers,

10 female, Utrecht dialect speakers.

As a criterion for dialect speaking, the following point was observed, viz. the speaker's parents and the speaker himself should have lived permanently in the area (surroundings of Maarssen, in the county of Utrecht).
The 40 speakers each produced the twelve Dutch vowels, which were sustained for c. three seconds. The formants of these vowels were measured in the sound curve, the definition of the term formant being a natural frequency of the vocal tract.
A principal component analysis was carried out on the 40 observations of the 24 variables thus obtained.
Only five significant components appear to be present, which account for 80% of the total variance. The remaining 20% of the variance, which cannot be accounted for, must be put down to statistical error. It becomes clear that a listener needs only a slight knowledge of the private system of a speaker, owing to the fact that the systematics within the vowel system occur with great strictness.

Next the common underlying parameters of the vowel systems were searched for.

The results of experiments of artificial vowels composed of two formants in which the subjects were requested to choose an optimal system of vowels by adjusting two dials, demonstrates clearly that an optimal system makes use of the greatest contrast possible within the scope of the producing mechanism.
The supposition that a real speaker when articulating carefully also makes use of the greatest contrasts whithin the possibilities of his vocal tract seems a reasonable one. The theory of the twin-tube model shows that these possibilities are defined by the extreme values which the parameters of this model can assume. Hence it should be possible to reduce a number of the common underlying parameters of well-articulated vowel systems to twin-tube parameters.
In order to test this the data mentioned above in the principal component analysis were subjected to a factor analysis.

In factor analysis we try to find a small number of new parameters, the so-called common factors, which produce the same co-variances as the original variables and a set of unique factors which are related to only one of the original variables, and which account for the uniquenesses being the parts of the variances of the variables that are not explained by the common factors. The explained parts of those variances are called "communalities". The statistical error is part of the uniquenesses.
The common factors span the so-called common factor space.

The common factor space is invariant with respect to rotation (orthogonal as well as oblique) so the solution found is one of an infinite series of equivalent solutions. In practice this solution is rotated into the required shape for simplicity of structure making use of an objective or subjective criterion. In this case a so-called Varimax-rotation was carried out. This is an orthogonal rotation making use of an objective

criterion, so that it can be done by computer.
The analysis into principal components showed up only five significant components. The fifth component shows a strong relation with the F2 of [a] only. For this reason a four-factor model was chosen for the factor analysis with the expectation of a substantial unique factor for F2 of [a] .
The analysis was reiterated four times on communalities. After four iterations the communalities remained stable within 0.01 or less. (Total convergence 99.93 %.) The four common factors together account for 71 % of the total variance. For F2 of [a] , as had been expected, a substantial uniqueness was found, viz. 0.77. As no replication was available it was impossible to estimate which part should be accounted for by error.
The results of Varimax-rotation are reproduced in figures 1 to 4. The mean values for the body investigated are given as points representing the vowels.
Shifting of a vowel is shown by an arrow in the appropriate direction giving the increase of the factors concerned. The first factor can be defined as the ratio of cross-section, or the volume of the posterior tube of the twin-tube model.
The second factor is related to the amount of constriction.
The third is related to the length of the vocal tract.
The fourth factor is related only to vowels which are situated within the vowel system and fixes the position of the line on which the speaker places the vowels [ü] , [ō] and [ə] .
As males, on an average, have a longer vocal tract than females, the score on the third factor is bound to show this.
If we introduce the dichotomy positive-negative in the third-factor score we find:

third factor score	male	female	total
positive	17	2	19
negative	3	18	21
total	20	20	40

with $\chi^2 = 19{,}6$ df. = 1 $p < 10^{-3}$

Utrecht dialect speakers have a system in which the vowels deflect on the left of the vowel diagram. This ought to show in the score on the second factor.
If we introduce the dichotomy positive-negative once more in the score on the second factor we find:

second factor score	Utrecht	Non-Utrecht	total
positive	2	16	18
negative	18	4	22
total	20	20	40

with $\chi^2 = 17$ df. = 1 $p < 10^{-3}$

It should be noted that it became clear after the experiment that one of the Utrecht-speakers lives in surroundings where no dialect is spoken. This subject obtained a positive score on the second factor.

As in practice twin-tube parameters are not quite independent the question arises whether an analysis in factors which can be interpreted as twin-tube parameters and which are mutually independent, constitutes the best solution possible. Therefore the common-factor-space was rotated obliquely, which yields correlated factors, which, once more, have an equivalent solution.
The resulting solution deviates nowhere essentially from the orthogonal solution; the same figures and the same tables remain valid.

In the meantime a second series of formant measurements of the vowels of a comparable group of subjects had been completed. It seemed appropriate, therefore, to add the new material to that already analysed and to carry out, once again, a principal component analysis and a factor analysis on these combined data.

Little notice was taken at first of the fact that the new measurements which had been made, had been carried out on vowels of normal duration lifted from isolated words, and were not done on vowels which had been sustained for a while.

The group of subjects investigated now consisted of 80 speakers:

10 male, non-dialect, sustained vowels,
10 female, " , " " ,
10 male, Utrecht dialect, " " ,
10 female, " " , " " ,
10 male, non-dialect, vowels from isolated words,
10 female, " , " " " " ,
10 male, Utrecht dial., " " " " ,
10 female, " " , " " " " .

A principal component analysis on 80 observations of 24 variables yielded four significant components, which accounted for 74 % of the total variance.

In a factor analysis three significant factors were found. After four iterations the communalities remained stable within 0.01 or less. (Total convergence 99.99 %.)
The three common factors together account for 66 % of the total variance. For F2 of [a] only a considerable uniqueness was found, viz. 0.78.
The results after Varimax-rotation are represented on figures 5 to 7.
Once again the results after oblique rotation did not essentially deviate.

The three factors can be interpreted as follows:
The first is related to the left side of the vowel diagram and the vowels [ü] , [ö] , and [ø] , actually those vowels in which the dialect group and the non-dialect group differ (fig. 8). When we observe the score on the first

factor, we find, on introducing the dichotomy positive-negative:

first factor	Utrecht	Non-Utrecht	total
positive	7	33	40
negative	33	7	40
total	40	40	80

with $\chi^2 = 33.8$ df. = 1 $p < 10^{-3}$.

The second factor is related to the length of the vocal tract (fig. 9). When introducing the positive-negative dichotomy in the score on the second factor we find:

second factor	male	female	total
positive	5	31	36
negative	35	9	40
total	40	40	80

with $\chi^2 = 37.6$ df. = 1 $p < 10^{-3}$.

The third factor is related to the vowel reduction which appears more clearly in accordance with the lack of effort made by a speaker when articulating his vowels. An experiment by Mol and Blom[4], carried out some ten years ago, showed that in current speech speakers tend to cluster their vowels in two groups. The third factor points into the direction of this clustering (fig. 10).
When we introduce the positive-negative dichotomy we find for the score on the third factor:

third factor	sustained	non-sustained	total
positive	10	29	39
negative	30	11	41
total	40	40	80

with $\chi^2 = 18.5$ df. = 1 $p < 10^{-3}$.

The most striking result of this new analysis is the fact that the factors can no longer be interpreted as twin-tube parameters. In non-sustained vowels the possibility of a maximum of contrast being formed, needed to get the twin-tube parameters, is excluded. An influence now is exercised by the degree of care in articulating the vowels.

fig. 1

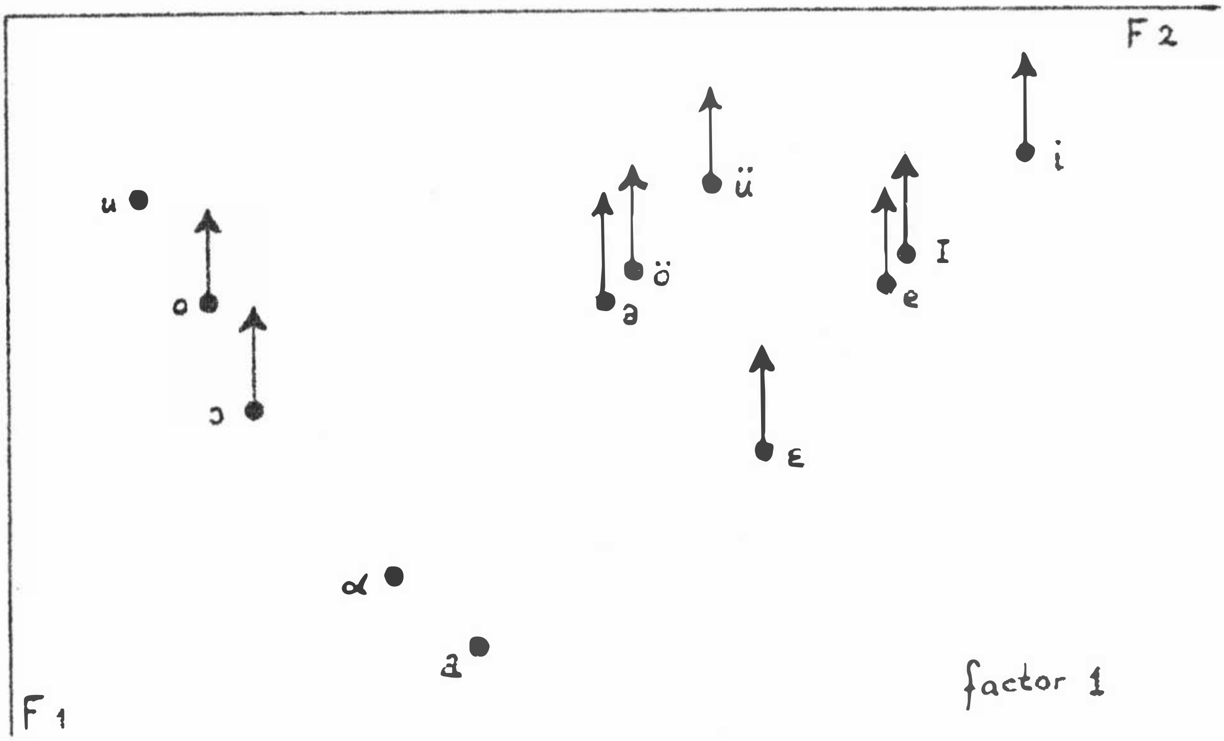


fig. 2

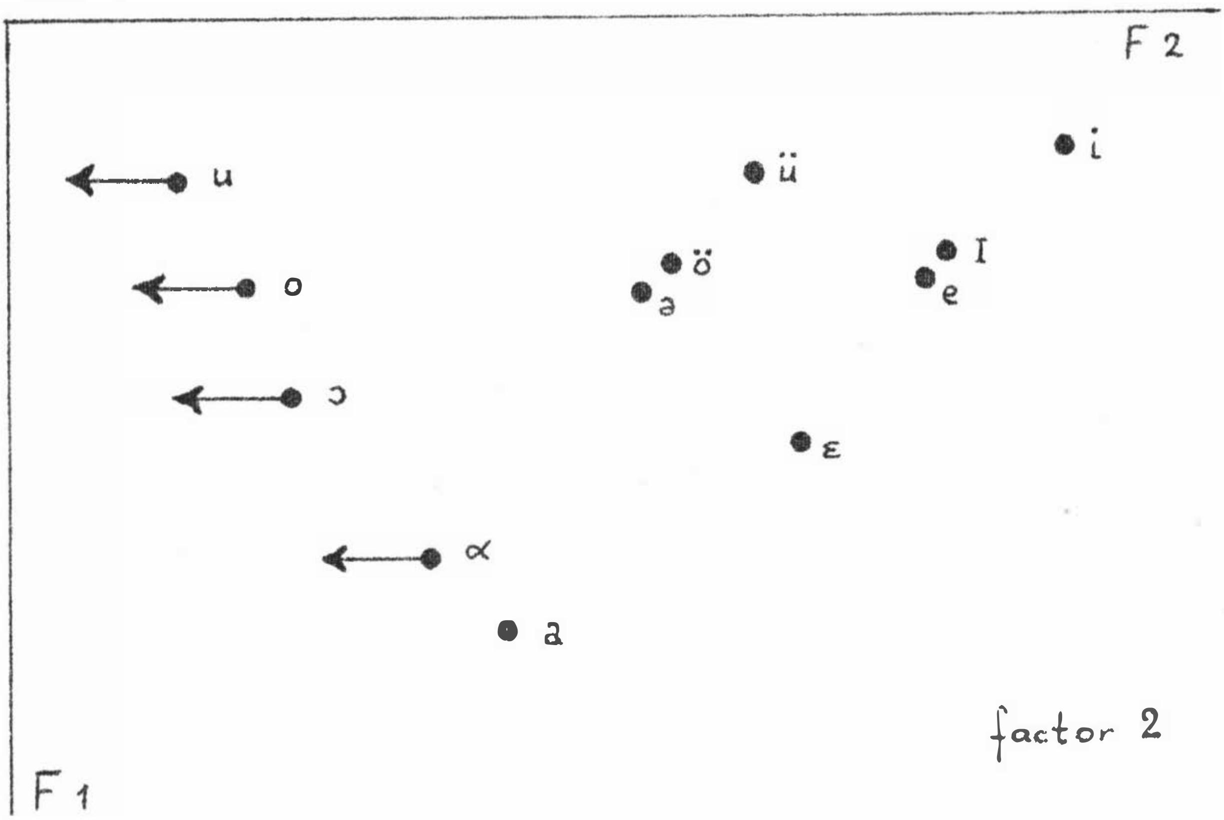


fig. 3

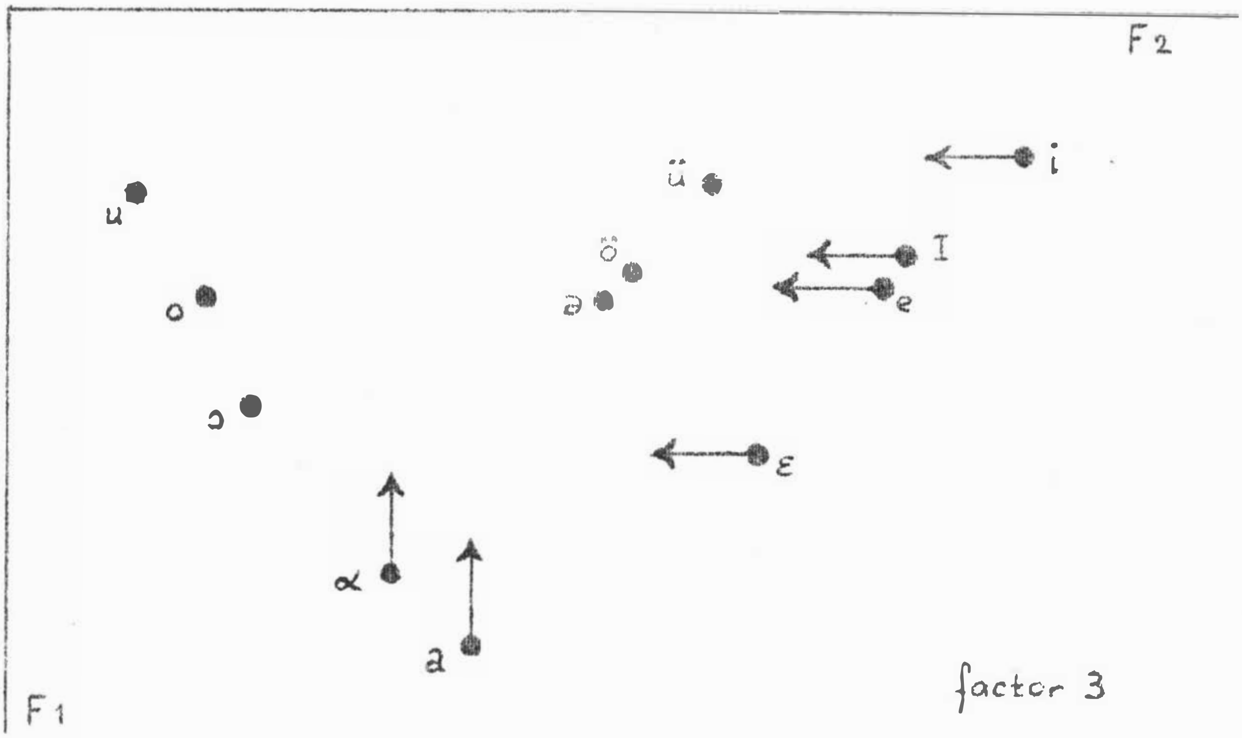


fig. 4

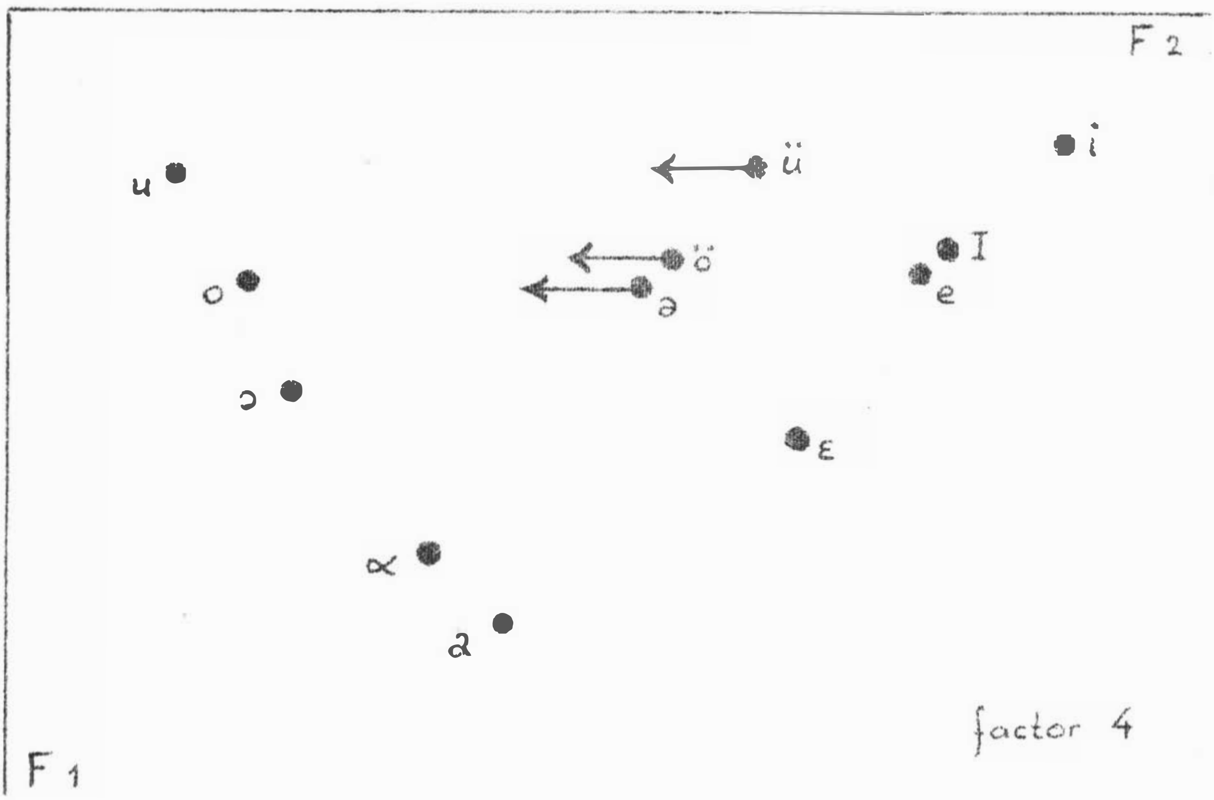


fig. 5

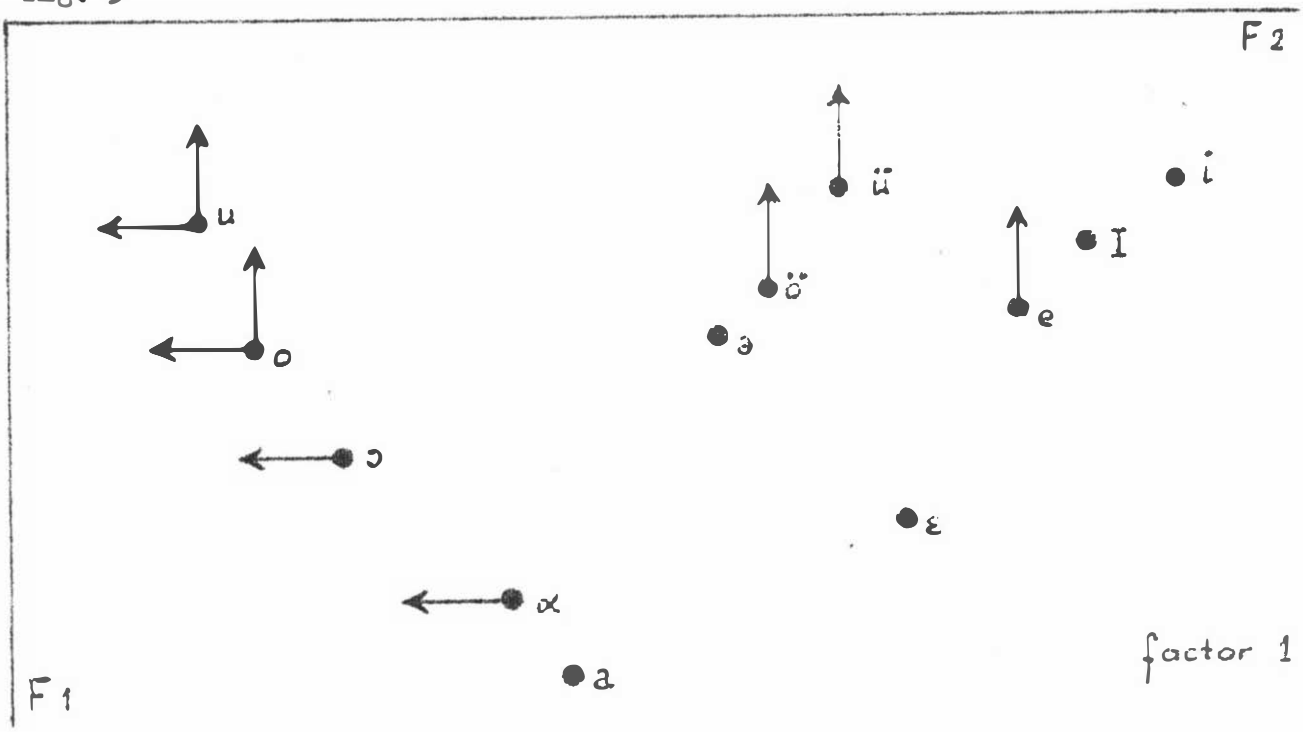


fig. 6

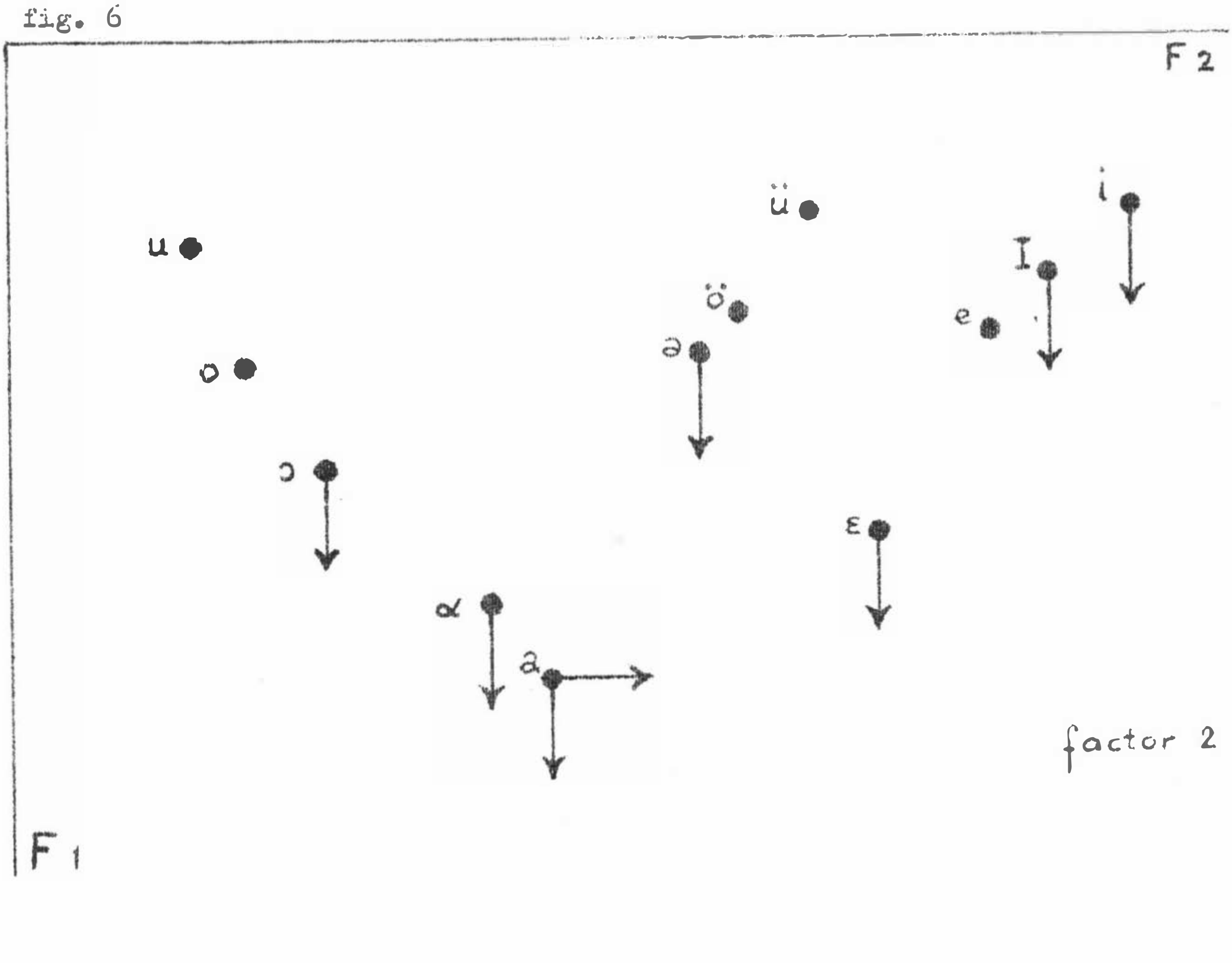


fig. 7

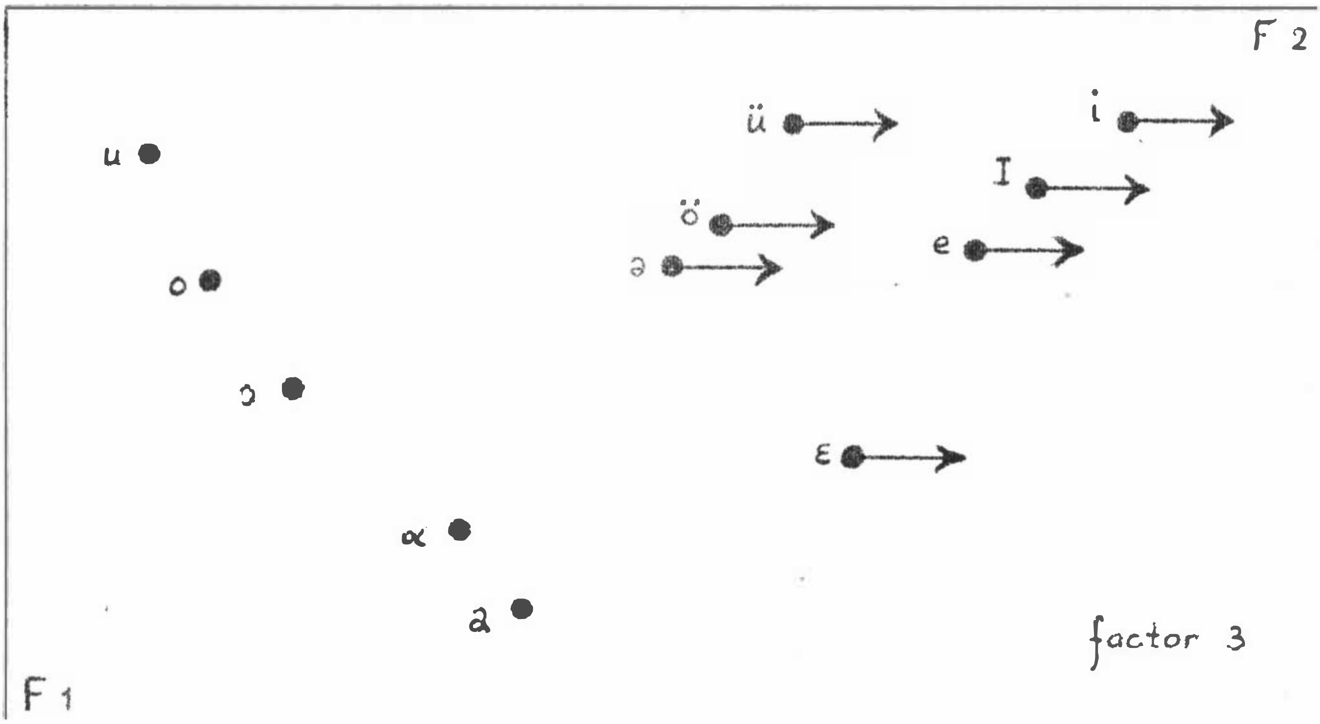
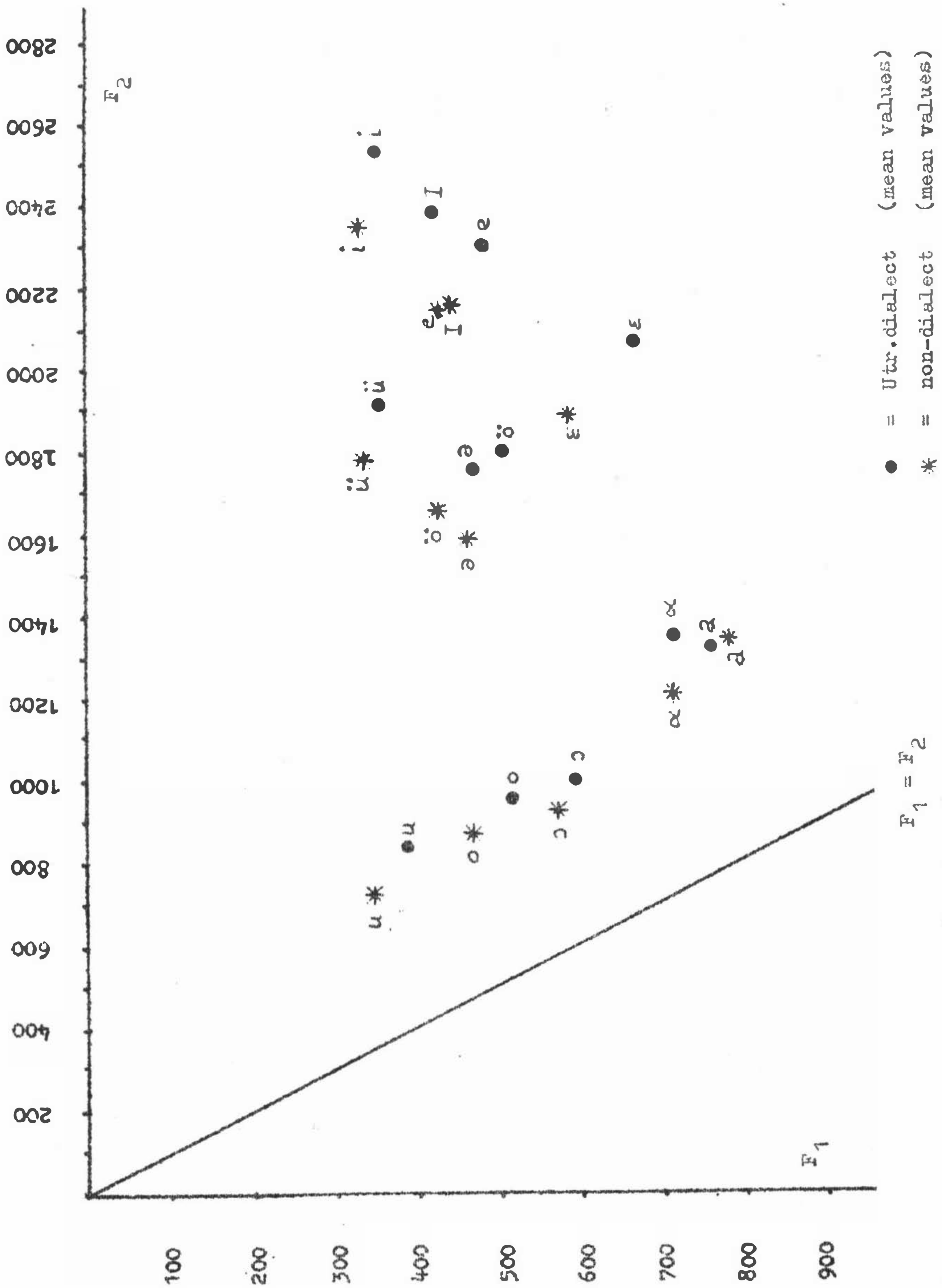
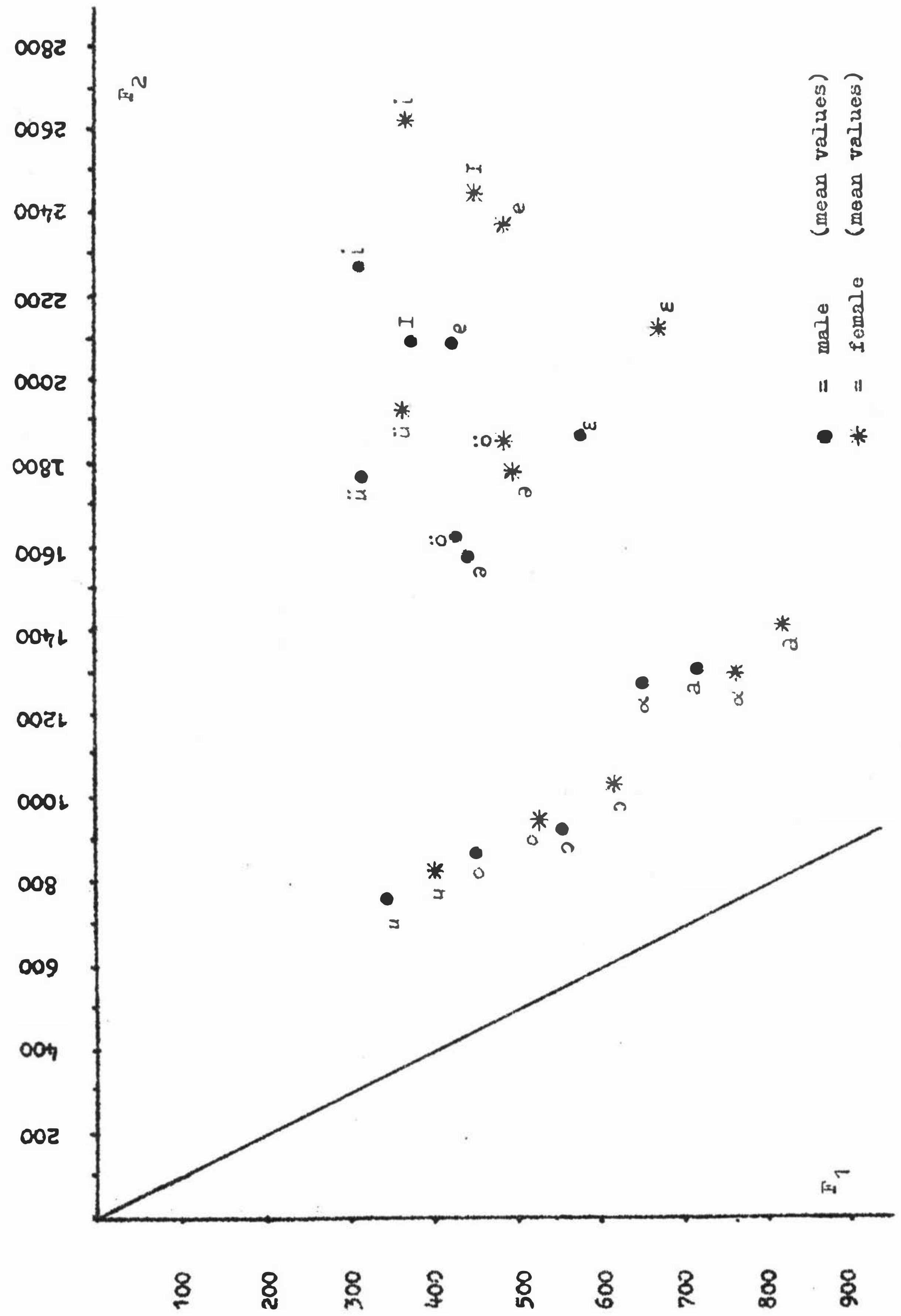


fig. 8

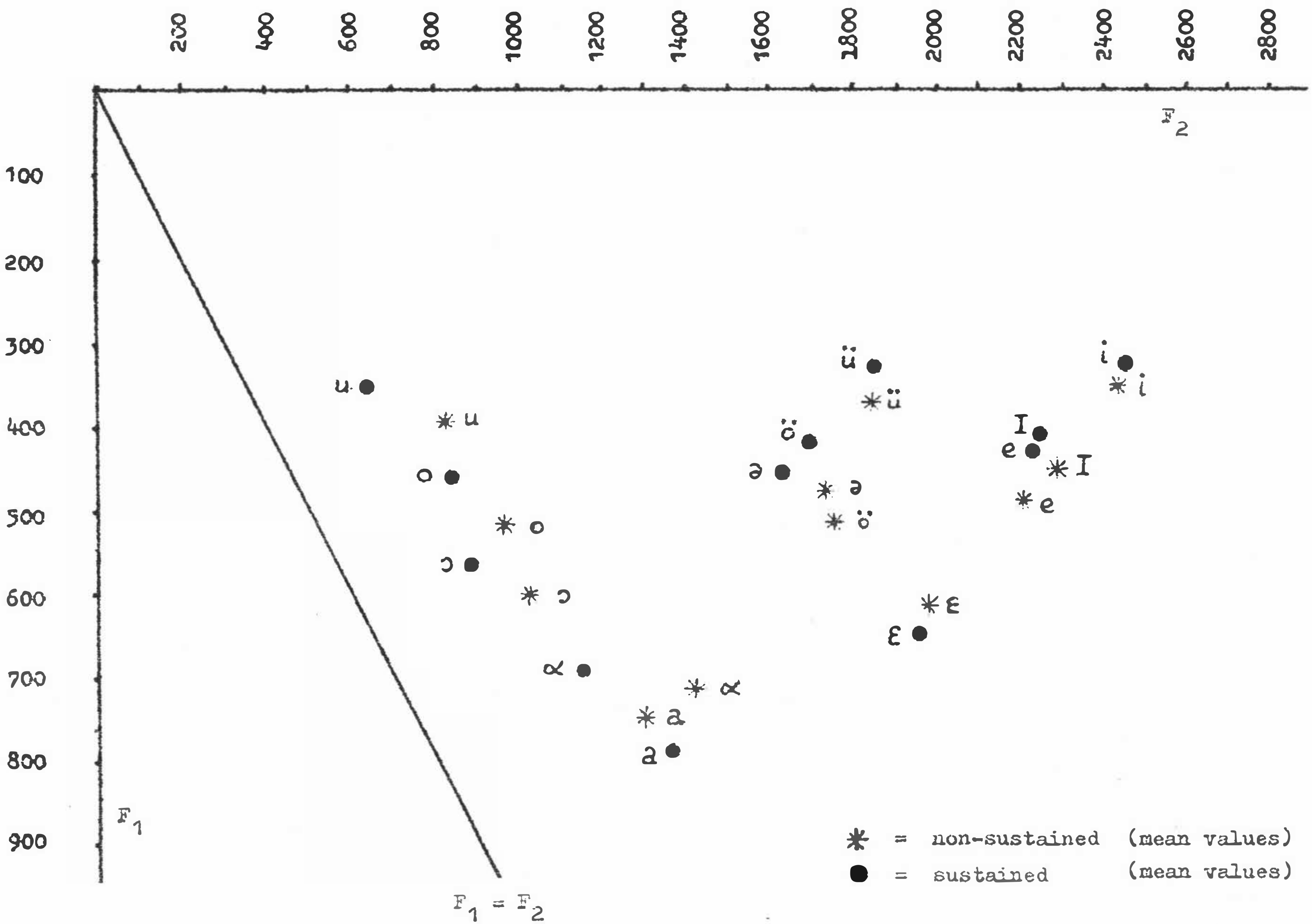


F2
F1
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
100
200
300
400
500
600
700
800
900
F1 = F2
● = Utr.dialect (mean values)
* = non-dialect (mean values)

fig. 9



F2
F1
F1 = F2
● = male (mean values)
✱ = female (mean values)
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2600
2800
100
200
300
400
500
600
700
800
900



200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2600
2800
F2
100
200
300
400
500
600
700
800
900
F1
F1 = F2
u
o
ɔ
α
a
ü
ö
ə
ɛ
i
I
e
✱ = non-sustained (mean values)
● = sustained (mean values)

Fig. 10

Notes

1) J.G.Blom, De klassificatie van klinkersystemen, in: Publication N° 16 of the Nederlands Akoestisch Genootschap, 1969.

2) F.J.Koopmans - van Beinum, Vergelijkend klinkeronderzoek: Utrechts tegenover Nederlands I. Publication N° 26 of the Institute of Phonetic Science. Amsterdam, 1970.

3) F.J.Koopmans - van Beinum, Vergelijkend fonetisch klinkeronderzoek, in: Publication N° 32 of the Institute of Phonetic Science. Amsterdam, 1971.

4) H.Mol en J.G.Blom, Vooronderzoek naar de ligging van de formanten in lopende spraak. Report N° 1 of the Phonetic Laboratory. Amsterdam, 1960.