ON THE INFLUENCE OF DURATION AND PITCH ON THE IDENTIFICATION OF ARTIFICIAL TWO-FORMANT VOWELS

by Jan G. Blom and Hendrik Mol

introducing and discussing an investigation carried out by Mr Willem Pieters as part of his examination for his MA degree.

Though, in Dutch, one traditionally distinguishes between long and short vowels, the exact role of vowel duration in speech recognition has yet to be determined.

In the course of the years we have formed the following general picture which, after all, needs elaborate experimental verification involving advanced statistical methods.

The formants are the main cues leading to vowel recognition followed by auxiliary cues like duration, vocal pitch, situation, context etc. As soon as, for one reason or another, the formants lose their discriminative power, the other cues come into play instead of merely being reinforcing.

In carefully pronounced, isolated words, produced by one single talker, he realizes optimal mutual contrasts between the formant positions of his 12 (Dutch) vowels. It can easily be measured that between two vowels with adjacent formant positions there may be a contrast in duration, sometimes amounting to a factor 2 , that probably supports the formant contrast. One may speculate that, in case the formant contrast between this pair is reduced for some reason, the duration contrast might take over the discriminative function, in that way furnishing one bit of information, leaving less to guess-work.

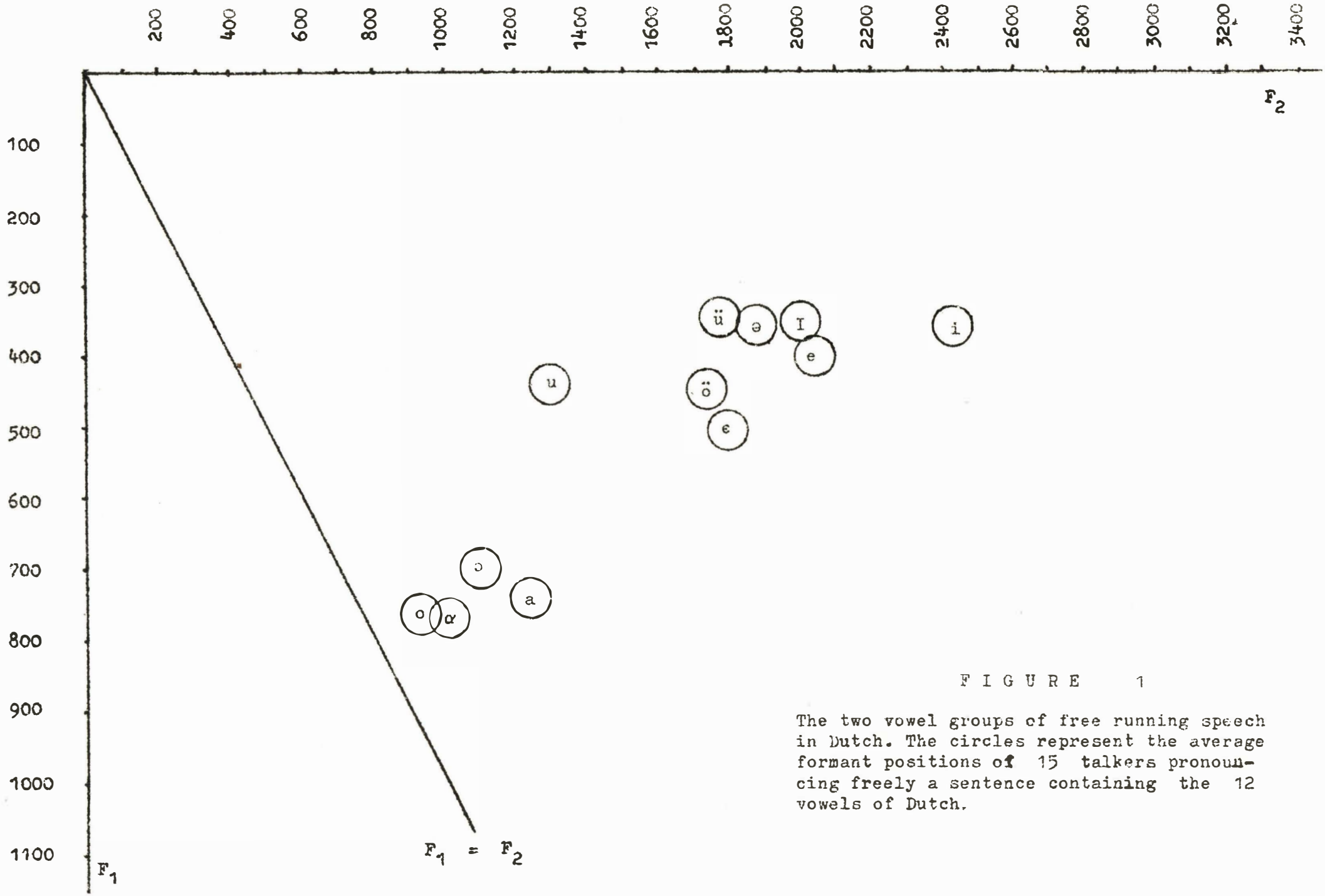
We like to stress here that the well-known successful identification of the vowels of one single talker, carefully pronounced in key-words or even in isolation, must be attributed to the recognition of the systematic contrasts (differences) between the formants of the vowels rather than to the recognition of the absolute formant positions of the talkers vowels.

This is as it should be because of the vast anatomical differences in length of the vocal tract between human talkers. Therefore, in listening experiments, we should not mix the vowels pronounced by different talkers. Also, in identification experiments involving artificial vowels with random formant positions that do not cohere systematically, we must expect considerable disagreement and overlap between listeners.

We are highly interested in the possible role of duration in the recognition of free running speech (connected speech) as opposed to carefully pronounced isolated words. We know that in free running speech there is a vast reduction in the formant contrasts, reducing the 12 vowel system to a mere 2 vowel system, one vowel group embracing the vowels [i I ü e ə u ɛ ö], the other group containing the vowels [o ɔ ɑ a] , see fig. 1.[1)]
From the beginning, we were baffled by the burden this 2 vowelgroup system places on the listener who has to derive the missing information from context, situation, knowledge of the language in question etc. If, however, the reduction in free running speech is limited to the formant contrasts and does not affect the duration contrasts,things would be looking up. For instance , in the group [o ɔ ɑ a], where the reduced formant contrast would compel the listener to make a choice of 1 out of 4 , the benefit of the duration contrast short-long would allow him to gain 1 bit and to make a choice of 1 out of 2 , backed by the context, the situation etc. Likewise , the other group, containing the kernel [I ü e ɛ ö ə] with highly reduced mutual contrasts, flanked by the obviously resistant big contrast [u i], may be expected to profit from the duration contrast, especially in the kernel.

Before setting up laborious experiments on a large scale, involving objective and subjective measurements on real speech, we thought it advisable to run a preliminary series of experiments using artificial two-formant vowels.

Therefore we invited Mr Pieters to repeat in the first place the Blom and Uys[2)] experiment pertaining to the identification of artificial two-formant vowels of constant duration, at the same time extending the frequency range the formants were allowed to cover.



F I G U R E 1

The two vowel groups of free running speech in Dutch. The circles represent the average formant positions of 15 talkers pronouncing freely a sentence containing the 12 vowels of Dutch.

The results again indicated that in certain regions of the F_1F_2 plane (also called the formant field) the agreement among listeners was decidedly better than in other parts where there was severe overlap.

It was decided to extend the experiment in the following way.

A new collection of artificial vowels was made by giving each formant pair 9 combinations of 3 different durations (100 , 250 and 400 msec) and 3 different pitches (80, 125 and 180 c/s). The order of magnitude of these values was adapted to those met in real speech.

The formant positions of these ' enriched' stimuli were restricted to the areas of severe overlap indicated by the first series of experiments with equally long and equally high formant pairs. The reason for this restriction was twofold: the number of stimuli presented to the listeners could not possibly be multiplied by a factor 9 whereas the possible influence of duration and pitch on the identification of the formant pairs could be expected in the zones of overlap where the formants alone could not inspire the listeners with unanimity.

As a measure for overlap we used the Discrimination Index D as defined by Mr J.G. Blom in the following way

$$D = \frac{k}{k-1} \frac{1}{N^2} \sum_{i=1}^{i=k} \left(n_i - \frac{N}{k} \right)^2 . 100 \%$$

where:

k is the number of possible answers the listeners are allowed to give

$n_1, n_2, \ldots\ldots, n_i, \ldots\ldots, n_k$ are the frequencies of the scores on these answers

$N = \sum_{i=1}^{i=k} n_i$ is the total number of scores

In the case that there is no preference for a special score we have

$$n_i = \frac{N}{k}$$

which leads to $D = 0$, the mimimum.

When there is an outspoken preference for only one score, namely score j , then

$$n_j = N \quad \text{and} \quad n_i = 0 \quad \text{for} \quad i \neq j \ .$$

In this special case we arrive at

$$D = \frac{k}{k-1} \frac{1}{N^2} \left[\left(N - \frac{N}{k} \right)^2 + (k-1) \left(\frac{N}{k} \right)^2 \right] 100 \ \%$$

or $D = 1$, which represents the maximum.

By writing D as follows

$$D = \frac{1}{N(k-1)} \sum_{i=1}^{i=k} \frac{\left(n_i - \frac{N}{k} \right)^2}{\frac{N}{k}} \ 100 \ \%$$

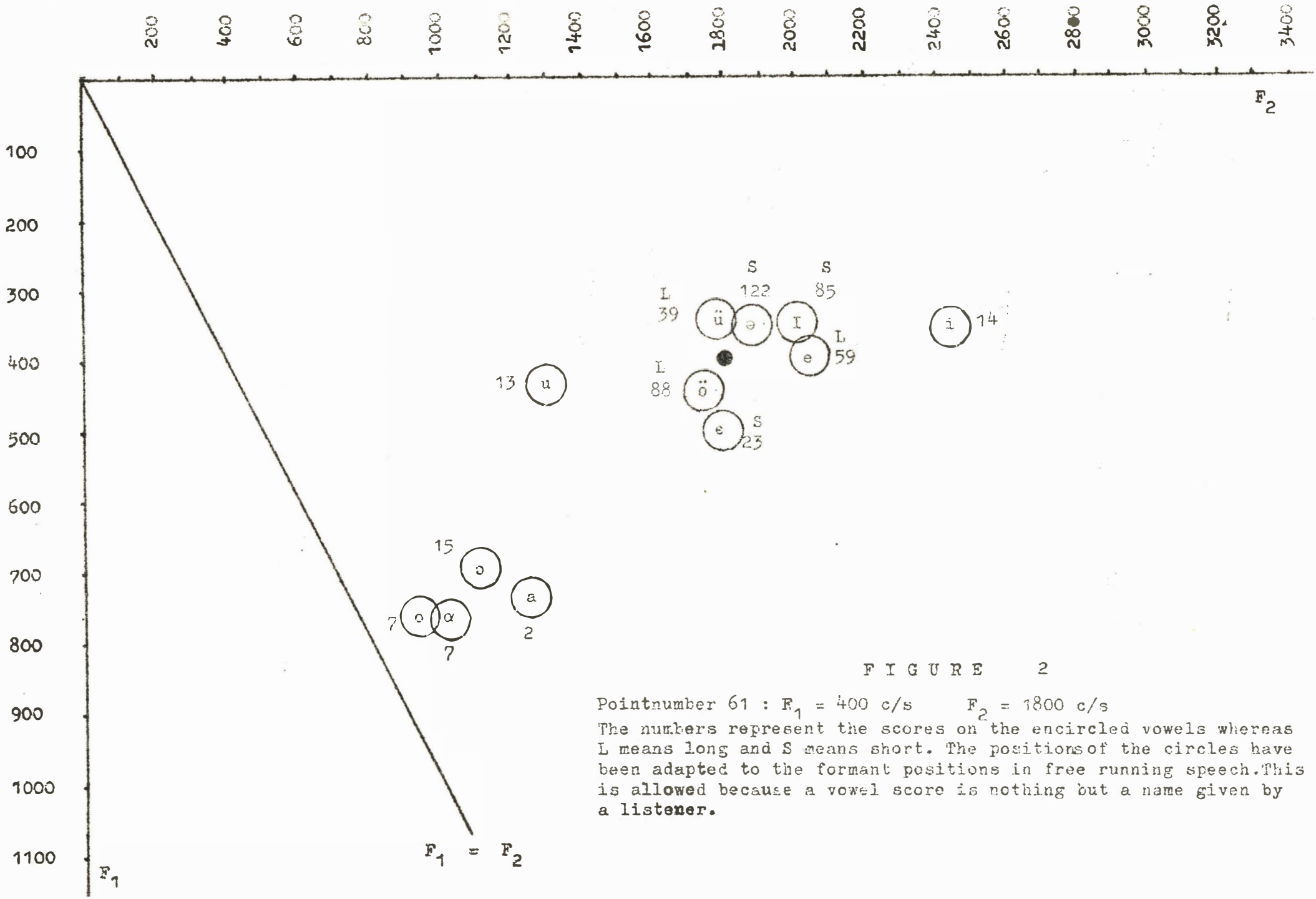
Blom was able to attach an interesting meaning to his index D, because

$$\chi^2 = \sum_{i=1}^{i=k} \frac{\left(n_i - \frac{N}{k} \right)^2}{\frac{N}{k}}$$

represents the deviation of the frequency distribution from the zero hypothesis (H_0) that there is no preference at all for a certain, special score. For that reason we are able to test D .

Obviously , $\frac{N(k-1)D}{100}$ has a χ^2 distribution with k-1 degrees of freedom.

Though the results of Mr Pieters' experiments with both 'poor' and 'enriched' artificial two-formant vowels will be treated statistically and after that published in a separate publication of our institute, we can already now present some interesting results with respect to the problems of free running speech.



F I G U R E 2

Pointnumber 61 : F_1 = 400 c/s F_2 = 1800 c/s

The numbers represent the scores on the encircled vowels whereas L means long and S means short. The positions of the circles have been adapted to the formant positions in free running speech. This is allowed because a vowel score is nothing but a name given by a listener.

For that reason we select the score pertaining to the enriched artificial vowel F_1 = 400 c/s, F_2 = 1800 c/s, depicted in fig. 2 and TABLE I .

In fig. 2 the total number of scores on a certain vowel (as defined here as the name a listener gives to something he hears) has been indicated. It strikes the eye that the scores on the vowels in the kernel are much higher than those on the other vowels outside the kernel, so that the kernel, pertaining to free running speech, comes to the fore here as a group with its own identity.

T A B L E I

Pointnumber 61: F_1 = 400 c/s , F_2 = 1800 c/s

100 msec	61	7	6	10	8	51
250 msec	47	22	13	9	22	19
400 msec	14	59	20	4	29	15
total	122	88	39	23	59	85
	ə	ö	ü	ɛ	e	I
	S	L	L	S	L	S

This, however, is not the only interesting property of the kernel: inside the kernel there is a duration discrimination as Table I clearly shows. When we call a vowel long when its score increases with duration and short when its score decreases with duration we get the following classification:

long vowels	ü	ö	e
short vowels	ə	I	ɛ

which certainly is not a variance with the traditional approach. The scores of the vowels outside the kernel, though for reasons of simplicity not shown in the Table, do not show a clear-cut dependence on duration.

The next intriguing example is shown in fig. 3 and Table II. This is the enriched vowel F_1 = 700 c/s, F_2 = 1200 c/s. Fig. 3 shows that now the group [o ɔ ɑ a] rakes in most scores, the other vowels being poorly endowed. So also the second group of free running speech vowels shows its identity in this identification experiment featuring artificial two-formant vowels.

Inspection of Table II , pertaining to the same enriched formant pair shows that we have the following classification:

long vowels	o a
short vowel	ɑ
?	ɔ

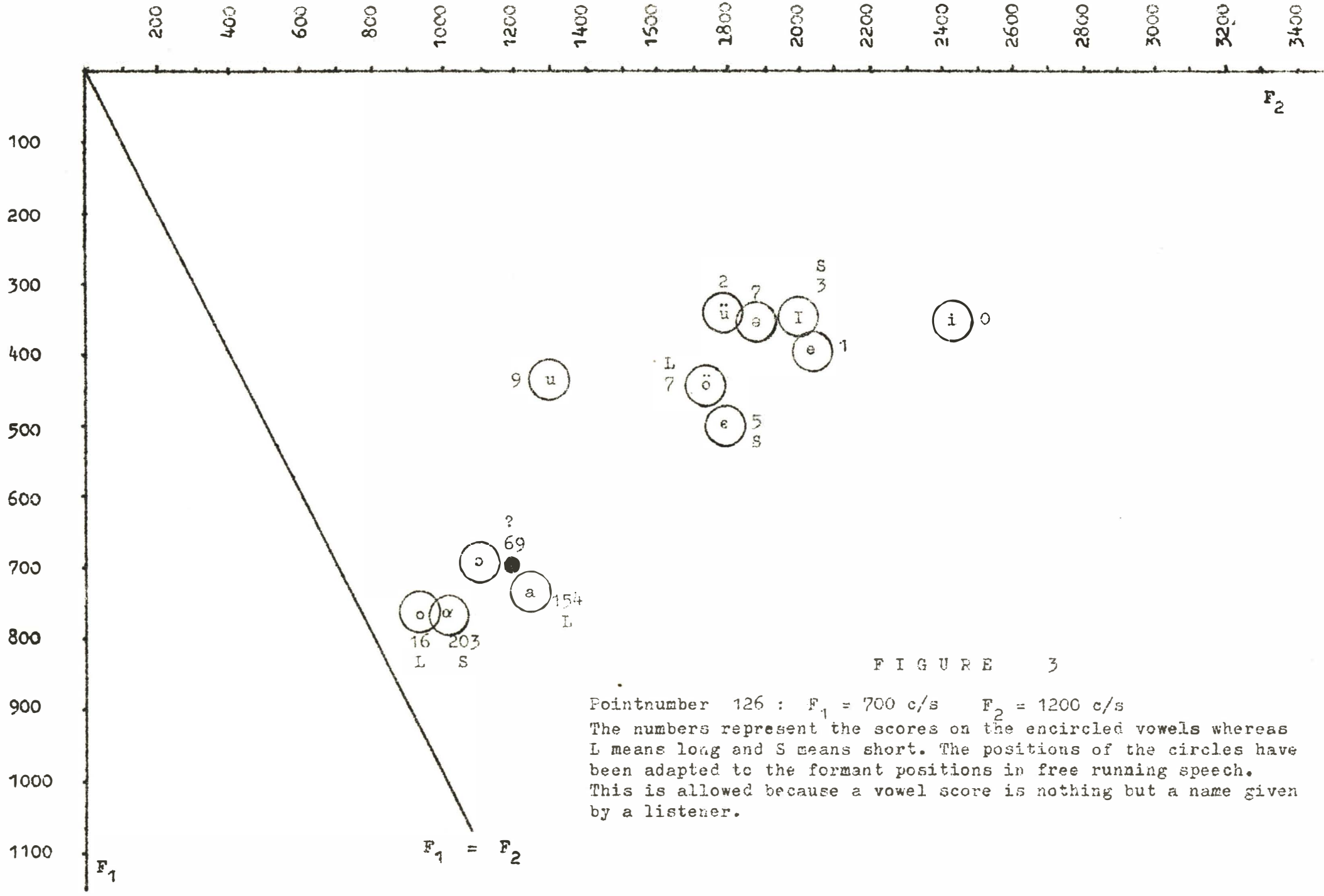
T A B L E II

Pointnumber 126 : F_1 = 700 c/s , F_2 = 1200 c/s

100 msec	0	43	104	1
250 msec	1	7	72	72
400 msec	15	19	27	81
total	16	69	203	154
	o	ɔ	ɑ	a
	L	?	S	L

The classification of the vowels o, ɑ and a is in agreement with traditional nomenclature. The vowel ɔ is unwilling to have itself classified in this experiment.

It is interesting to note here that, in this case, even in the group of outcasts depicted in Table III , there is a glimmering of a short-long distinction.



F I G U R E 3

Pointnumber 126 : F_1 = 700 c/s F_2 = 1200 c/s
The numbers represent the scores on the encircled vowels whereas L means long and S means short. The positions of the circles have been adapted tc the formant positions in free running speech. This is allowed because a vowel score is nothing but a name given by a listener.

T A B L E III

Pointnumber 126 : F_1 = 700 c/s , F_2 = 1200 c/s

100 msec	3	1	0	0	4	0	3	0
250 msec	2	1	2	0	1	0	0	0
400 msec	4	5	5	2	0	1	0	0
total	9	7	7	2	5	1	3	0
	u	ə	ö	ü	ɛ	e	I	i
	?	?	L	?	S	?	S	?

Three vowels may be classified in accordance with tradition:

long vowel	ö	
short vowels	I	ɛ

Vowel u is more or less indifferent , in agreement with the role it quite often plays. Vowels ü and e , as it were reluctantly, declare themselves long for very long durations whereas, ə , for long durations, is willing to have itself classified as long, but against tradition. Vowel i is not in the picture at all and should be considered as an also-ran in this case.

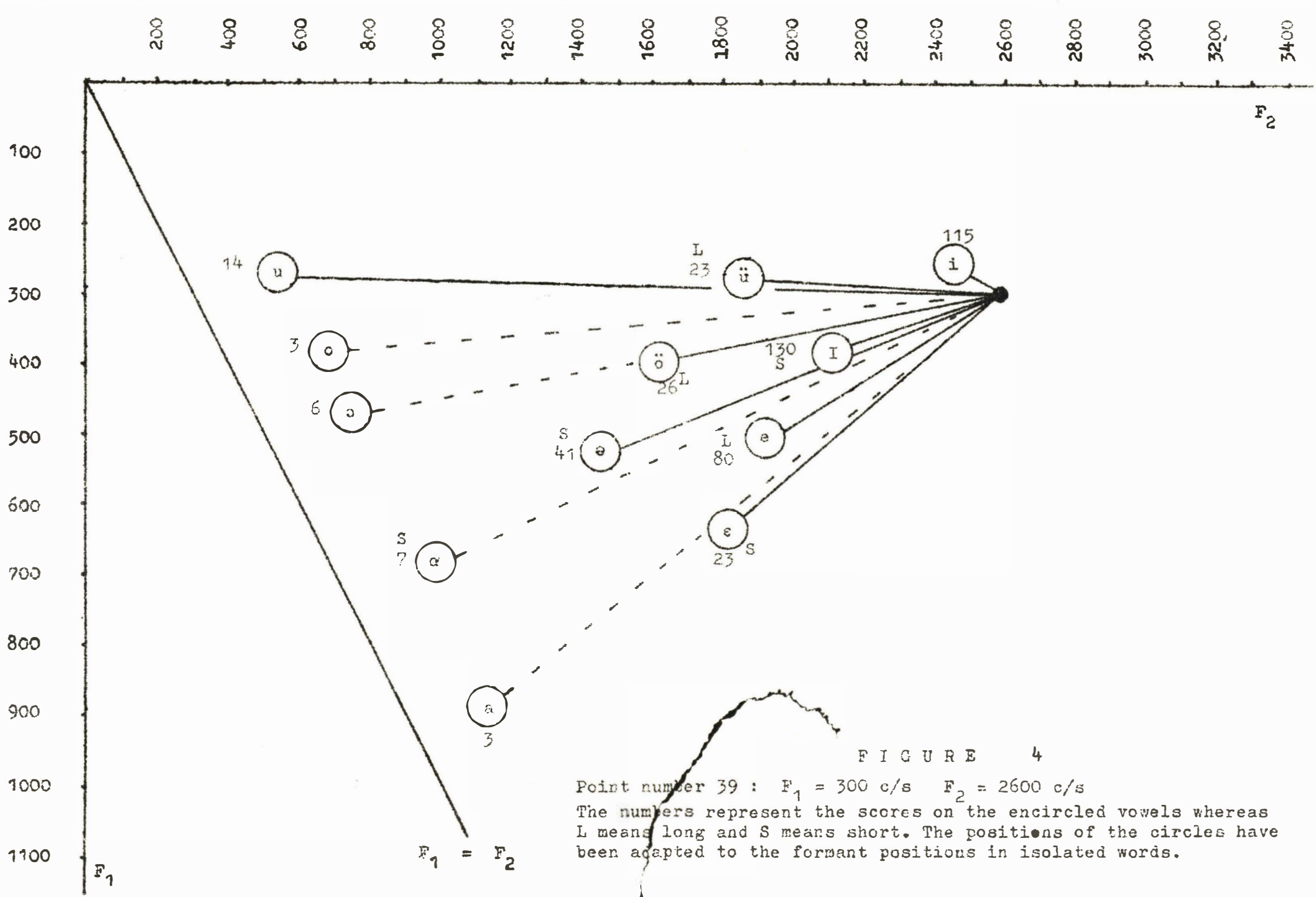
It is very rewarding to compare the scores on the following two enriched pointnumbers:

pointnumber 39 : F_1 = 300 c/s F_2 = 2600 c/s

pointnumber 126 : F_1 = 700 c/s F_2 = 1200 c/s

They represent the pointnumbers with the highest and the lowest second formants.

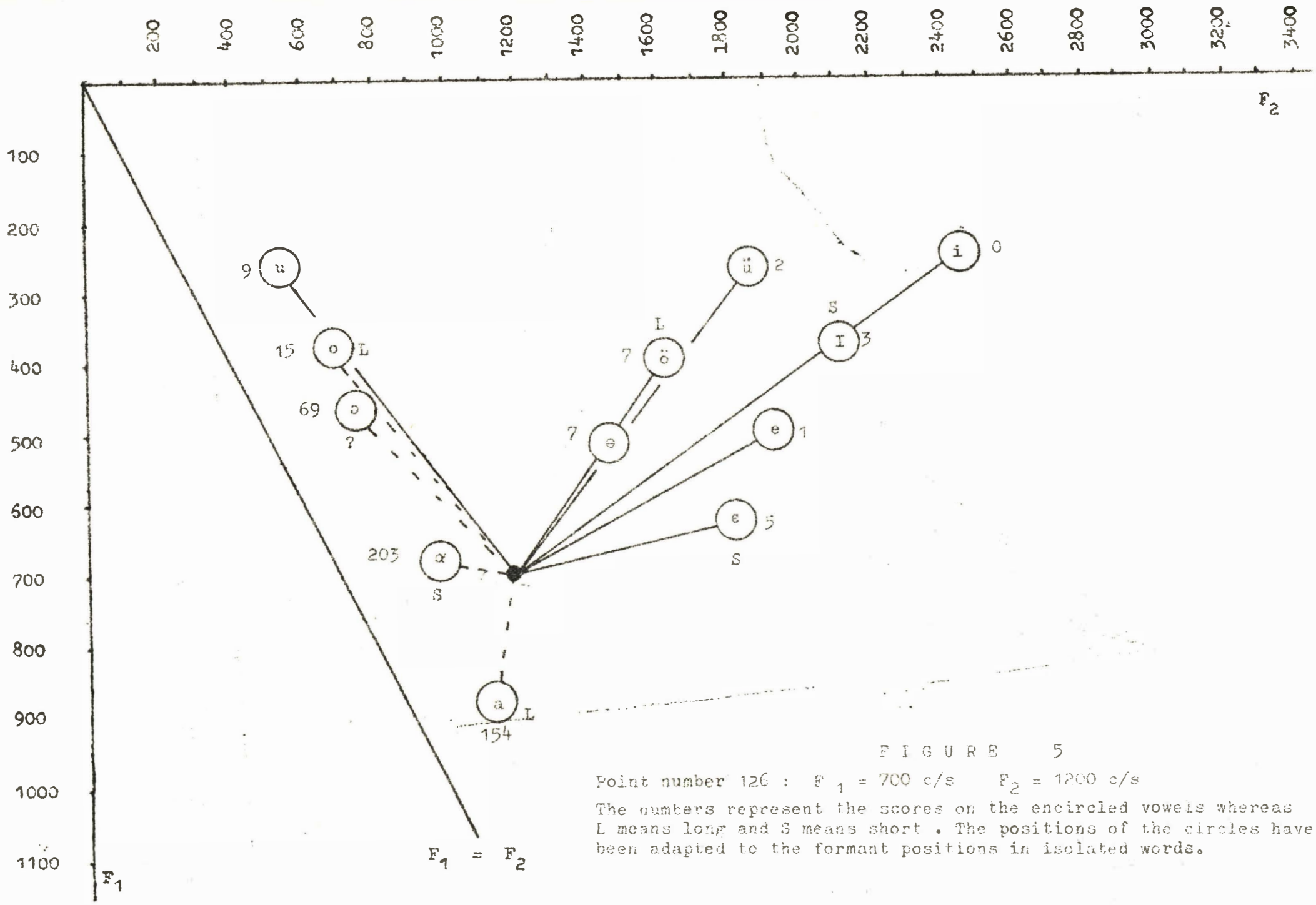
As shown in figures 4 and 5 we again indicate the scores on each of the twelve vowels as well as the short-long distinction, if any, but now we have at the same time indicated the formant postions as achieved by an average talker in carefully pronounced isolated words. The indications short and long have only been given when the scores show a clear-cut tendency to be influenced by duration.



F I G U R E 4

Point number 39 : F_1 = 300 c/s F_2 = 2600 c/s

The numbers represent the scores on the encircled vowels whereas L means long and S means short. The positions of the circles have been adapted to the formant positions in isolated words.



F I G U R E 5

Point number 126 : F_1 = 700 c/s F_2 = 1200 c/s

The numbers represent the scores on the encircled vowels whereas L means long and S means short . The positions of the circles have been adapted to the formant positions in isolated words.

It is interesting to see how in both figures the two groups of free running speech are clearly separated by the score numbers. Remark. Figures 3 and 5 display the scores of the same point-number 126 but in a different manner: fig. 3 alludes to free running speech whereas fig 5 projects the scores on isolated words.

Though the pitch of artificial vowels is not completely without influence on their identification, the results indicate that, in Dutch, this influence is of a much lower hierarchy than that of formant positions and duration. We shall not discuss it in this preliminary appraisal.

Conclusions

Though the statistical treatment of the results of the experiments of Mr Pieters has not yet been completed, the mere tabulation of the scores of the enriched formant pairs (meaning that the duration and the pitch of a formant pair have been introduced as additional variables) shows interesting phenomena.

In the first place the two vowel(group)s , found by formant measurements on real vowels in free running speech, also come to the fore in the scores of the listeners identifying artificial (and therefore unintended !) vowels in the overlap areas of the perceptive formant field. It seems improbable that this phenomenon is due to chance.

In the second place, in these overlap areas, the short-long distinction clearly is at work in the perceptive domain.

These two experimentally derived facts are not at variance with our hypothesis that in isolated words the formant positions are more important (have a higher hierarchy) than the durational cues and that in free running speech the durational cues partly take over the discriminatory task of the formants, leaving the rest to context , situation etc.

As the phonemes are defined in isolated words, we may suspect that, in Dutch, vowel duration does not directly enter the picture on the phonemic level.

We hopefully suggest that further experiments along simular lines as indicated in this publication may deepen our insight in the relation between free running speech and isolated words.

References.

1) H. Mol, Fundamentals of Phonetics, I: The Organ of Hearing, The Hague 1963, p. 50.
2) J.G.Blom and J.Z.Uys, Some Notes on the Existence of a ‚Universal Concept' of Vowels, Phonetica 15: 65-85 (1966)