# SPECTROGRAPHIC ANALYSIS OF ENGLISH VOWELS[1]

Peter Holtse

## 1. Introduction

The British English system of stressed vowels is generally agreed to consist of eleven phonologically distinctive elements of relatively pure quality. These eleven monophthongs are traditionally divided into two classes: one comprising five relatively long vowels /i:, a:, o:, u:, ə:/, and one consisting of six short vowels /I, e, æ, ʌ, ɔ, U/.

It was, however, recognised quite early that the phonological opposition between any two vowel phonemes is never dependent on a difference in duration alone. Thus the difference in duration is always accompanied by a certain difference in quality. And it was in fact shown by Gimson (1945-49) that of the two factors: quality and duration, the former is probably the more important for the distinction between pairs of vowels such as /i:/-/I/ and /u:/-/U/ which had otherwise been considered to differ mainly in duration.

The quantitative differences between the English vowels were measured as early as 1903 by E.A. Meyer. Meyer described in details how the exact duration of a given vowel is not only determined by the class to which it belongs (long or short) but also by its quality (open vowels are longer than close vowels), and by the nature of the following consonant (vowels are longer before voiced than before voiceless consonants and longer before e.g. fricatives than before stop consonants).

---

1) The contents of the article is part of a thesis work for the degree of cand.phil.

Similar principles were later shown to be operating in American English, first by R-M. S. Heffner in a number of articles and later with improved experimental technique by House and Fairbanks (1953), Peterson and Lehiste (1960), and House (1961). It is true that Heffner found only slight indications of a difference between the so-called "long" and "short" vowels. But Peterson and Lehiste found the vowels to be clearly divided in two groups corresponding to the traditional division. The only exception was the vowel /æ/ which according to Peterson and Lehiste, but contrary to traditional descriptions, should be classified as "long".

The number of investigations dealing with the same problems in British English are surprisingly small. One finds references to various kymographic experiments but no major study before Wiik (1965). On the whole Wiik's results from five informants are in good agreement with the findings of E. A. Meyer except that Wiik classifies /æ/ as neutral with respect to the "long" "short" opposition.

As to the qualitative differences between the vowels American English is again quite well provided with instrumental investigations (e.g. Peterson and Barney (1952), and Fairbanks and Grubb (1961)). The British English vowels, however, have only been described in a short article by Wells (1963) and in the more detailed study by Wiik (1965).

The aim of the present paper is to provide additional analytical data on both the formant frequencies and the duration of British English vowels. This material should further provide a useful basis for a discussion of how quality and quantity may interact as factors distinguishing different vowel phonemes.

## 2. Procedure

### 2.1. Material

The material of the investigation consisted of the following list of CVC-words:

| | | | | | | |
|---|---|---|---|---|---|---|
| [i:] | heat | heed | [ɔ] | pot | pod | cod |
| [I] | hit | hid | [o:] | port | pawed | cord |
| [e] | set | head | [U] | put | could | hook |
| [æ] | hat | had | [u:] | coot | cooed | hoop |
| [ʌ] | cut | cud | [ə:] | hurt | heard | |
| [a:] | heart | hard | | | | |

This list contains examples of each of the eleven monophthongs before a voiced and a voiceless consonant: the environmental factor which has the greatest influence on the duration of the vowel.

The effect of the initial consonant on the duration of the vowel was considered to be negligible. Therefore only the possible influence from the initial consonants on the vowel formant frequencies was taken into consideration, and the test words were chosen so as to minimize this influence as far as possible.

### 2.2. Informants

Six male native speakers of Standard English[2] acted as informants. All six of them were judged by competent observers to speak a very close approximation to RP.[2] Three of them were university students, two were university lecturers, and one was a civil servant. One informant was 71 years of age, the remaining five were between the age of 20 and 30.

---

2) Cf. Abercrombie (1965) p. 10-16.

## 2.3. Recordings

The test words were arranged in six randomized lists of isolated words which were read aloud by each of the informants. Two of these read two of the lists twice. After mispronounced words had been discarded the number of recorded words totalled 906.

The recordings of two informants were made in the recording studio of the Department of Linguistics, University of Edinburgh, using a Revox tape recorder and EMI Recording Tape. The recordings of the other four informants were made in the studio of the Institute of Phonetics, University of Copenhagen, using a Lyrec Professional Recorder and Scotch Magnetic Tape.

The recordings were analyzed on a Kay-Electric Sona-Graph. One wide band, one narrow band, and at least one cross section spectrogram were made of each of the recorded words. At least once every time the sound spectrograph had been used a calibration was made using a 100 Hz and a 500 Hz tone.

# 3. Formant Frequencies

## 3.1. Formant frequency measurements

The vowel formants were identified as the regions of relatively high intensity and the centre frequencies of these regions were measured. Wherever possible four formants were measured. But as could be expected, the back vowels proved rather difficult in this respect, and in a number of cases only formants one and two were in fact visible on the spectrograms.

The formant frequencies were measured with an accuracy of approximately 25 Hz. In view of the importance of very small differences in the low frequency regions a not altogether successful attempt was made to raise the accuracy of the F1-measurements to 10-15 Hz. This was done by using the individual harmonics of the narrow band spectrogram as a scale after the fundamental frequency at the point had been determined.

In all cases the formants were measured in the middle of a period of relatively steady state.

The results were fed into a computer and run through a standard program (XFON) which for each of the eleven vowels calculated the arithmetic mean, standard deviation, standard error of mean, and 95 and 99 pct. confidence limits.[3]

## 3.2. Vowel diagrams

A general idea of the auditory quality of a vowel is provided by showing the two most important formants, F1 and F2, in a two dimensional diagram. This is actually a very crude representation, and we might wish to improve it by somehow taking at least F3 into account as well. In the present study this has been done in a few cases in the form of a double-diagram with F1 as a common vertical axis to the two horizontal axes F3 and F2 (see Figs. 2 and 3).

The other possible solution: computing a weighted average of F2 and F3, seemed less attractive since we have at the time being no satisfactory way of taking both formant frequency and formant intensity into account.

In all the diagrams the frequencies of the formants have been given in Mels since this scale provides the best approximation so far to the way frequency is perceived by the ear. It should, however, be borne in mind that the Mel scale has been

---

3) The computer programs were written by cand.scient. J. E. Knudsen whose help is gratefully acknowledged.

found from experiments with simple sounds or narrow band noise and we only assume that it is applicable to complex sounds as well.

The vowel diagrams show the mean formant frequencies marked with a cross. The dispersion of the individual measurements of a given vowel is indicated by an ellipse drawn through four points each lying at a distance of two standard deviations from the mean, and measured along the F1- and F2-dimensions. Fig. 1 shows an example of the dispersion of a set of individual measurements within their 2s-ellipses.

Since the scale of the diagrams was drawn in Mels the standard deviations used to construct the 2s-ellipses had to be expressed in Mels as well. Otherwise the ellipses would have been skewed. Therefore all the individual measurements were converted from Hz into the corresponding Mel-values and run through the standard XFON-program once more. These calculations were used in drawing the vowel diagrams.

Traditionally the so-called "phoneme area" of a given vowel has been represented in the vowel diagram simply by a line drawn round all the individual measurements. It has, however, always been a problem how far one or two extreme values could be excluded from the envelope. Using the 2s-ellipses appears to be a satisfactory way of solving this problem. On the other hand the effect on the average of extreme values is still quite marked as far as this is calculated as the arithmetic mean. And it has been suggested to me by Eli Fischer-Jørgensen that a possible solution to this problem might be to use either the geometric mean or the median and quartiles as a measure of central tendency instead of the arithmetic mean.

One serious drawback in the use of 2s-ellipses is that any tendency to correlation between the two formants is completely obscured since it will only appear as an increase of dispersion. Obviously this difficulty ought to be taken

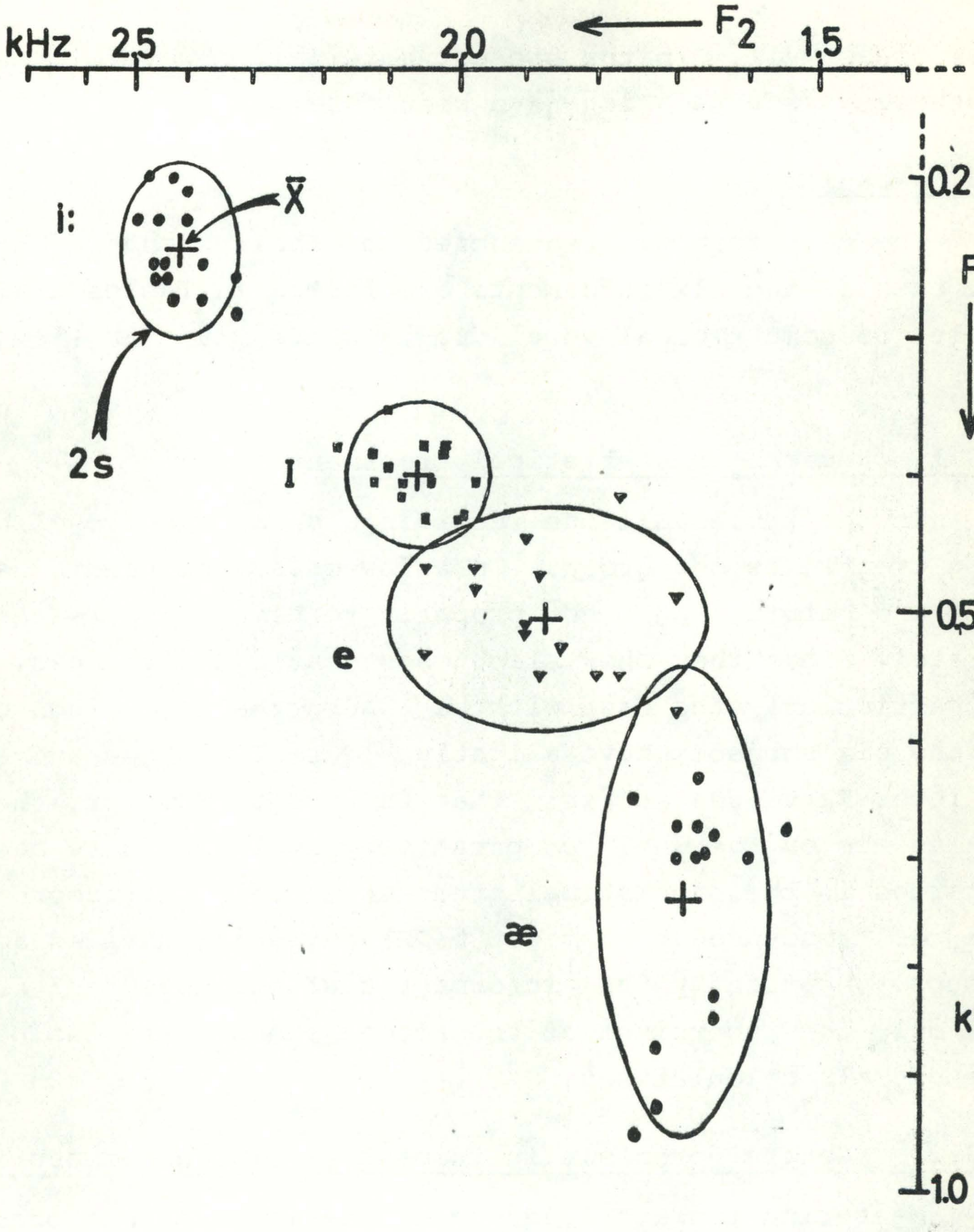



Fig. 1. Example of the variation of individual measurements. The axes of the ellipses correspond to a distance of two standard deviations from the mean.

care of as well. In the present material, however, no such instances of correlation have been observed.

### 3.3. Results

The mean formant frequencies and their standard deviations for each of the six informants are listed in Tables I-VI. Examples of some typical vowel diagrams are given in Figs. 2 and 3.

#### 3.3.1. Remarks on statistical treatment

In the tables all the recordings of a given vowel have been treated as one group. In a few cases the vowels before [-t] and before [-d] ought properly to have been treated separately since they show slight systematic differences. This is particularly the case with [e] where the recordings of four of the six subjects have slightly higher F1 values and slightly lower F2 values in [set] than in [hed]. However, the differences are on the whole comparatively small and have been disregarded in the statistical treatment. Only the recordings of [o:] as pronounced by subject 5(DH) have been divided into two groups. Apparently this informant distinguishes /ɔe/ from /o:/, and only the recordings of the second phoneme are included in the overall calculations.

#### 3.3.2. General problems in formant finding and measuring

Measuring formant frequencies presents various practical problems. Ladefoged (1962) describes the general ones like e.g. the difficulty in measuring very low first formants or the problems caused by a high fundamental frequency etc. Another kind of difficulty, also mentioned by Ladefoged, is the appearance of so-called spurious formants at frequencies where no formants should be or the possible absence of normal formants at other frequencies.

The present material shows many examples of these phenomena. Thus subject 1(AW) consistently has what looks like two first formants in the vowels [æ] and [ʌ], one at about 900 Hz and another at 700 Hz. And in the same vowels subject 2(GG) has an extra F2: Besides the normal F2 at 1700 Hz in [æ] there is an extra formant at 13-1400 Hz. Similarly in [ʌ], where F2 is normally found at 1250 Hz, an extra formant appears at 1700 Hz. The same tendency although less pronounced is found in the vowels of the other four informants.

In view of the consistency with which these spurious formants appear, it seems doubtful if they should be completely disregarded. If, as appears to be the case, vowel quality is perceived by some sort of weighting of frequency areas rather than by identification of specific formant frequencies, these very strong extra formants must certainly have som influence on the auditory quality of the vowels. Therefore the traditional vowel diagrams, which take into account only the "official" formants, may give rather a distorted impression of the auditory distances between the vowels. In the present case [æ] and [ʌ] may in fact be perceived as much closer in quality than is indicated by the vowel diagrams.

Another example of the same problem is found in [ʌ] as pronounced by informant 6(RD). In all his recordings of [ʌ] before [-d] an extra formant is found at about 500 Hz. If this formant has any effect it must give the vowel a much more centralised quality than we would expect from the normal diagrams.

Furthermore, the apparently "split formants" provide some interesting problems. Many recordings of subject 4(DC) have two distinct formants in the F3 area. Both of them take their origin from the same frequency position and are clearly distinct from F4 and F2. In other cases, especially among the back vowels, F3 is readily identified but a rather strong extra formant may then appear at about 1500 Hz.

TABLE I

Results of formant frequency measurements. Informant 1(AW). The table shows mean values ($\overline{X}$), standard deviation (s), and number of tokens (N) for each of the first four formants.

| | | F1 | F2 | F3 | F4 | N |
|---|---|---|---|---|---|---|
| [i:] | $\overline{X}$: | 243 | 2419 | 2953 | 3666 | 16 |
| | s: | (26.8) | (49.6) | (61.8) | (52.3) | |
| [ɪ] | $\overline{X}$: | 394 | 2063 | 2745 | 3783 | 15 |
| | s: | (21.9) | (55.0) | (48.3) | (105.5) | |
| [e] | $\overline{X}$: | 503 | 1873 | 2712 | 3975 | 15 |
| | s: | (47.5) | (110.4) | (50.8) | (126.8) | |
| [æ] | $\overline{X}$: | 740 | 1688 | 2542 | 3767 | 16 |
| | s: | (107.1) | (59.8) | (69.9) | (66.9) | |
| [ʌ] | $\overline{X}$: | 646 | 1234 | 2409 | 3545 | 16 |
| | s: | (33.4) | (32.7) | (54.7) | (70.8) | |
| [a:] | $\overline{X}$: | 654 | 1070 | 2480 | 3506 | 16 |
| | s: | (31.6) | (64.7) | (95.4) | (84.9) | |
| [ɔ] | $\overline{X}$: | 574 | 927 | 2520 | 3457 | 16[4] |
| | s: | (45.7) | (34.7) | (50.2) | (96.1) | |
| [o:] | $\overline{X}$: | 450 | 788 | 2368 | 3183 | 15[5] |
| | s: | (23.3) | (62.6) | (87.9) | (78.6) | |
| [U] | $\overline{X}$: | 432 | 1042 | 2281 | 3316 | 16 |
| | s: | (34.4) | (84.5) | (60.2) | (80.0) | |
| [u:] | $\overline{X}$: | 285 | 1131 | 2312 | 3383 | 16[6] |
| | s: | (26.7) | (68.0) | (56.6) | (50.6) | |
| [ə:] | $\overline{X}$: | 503 | 1464 | 2539 | 3667 | 16 |
| | s: | (39.9) | (38.7) | (37.6) | (48.9) | |

4) F4 only 15 ex.

5) F3 only 14 and F4 only 12 ex.

6) F3 and F4 only 15 ex.

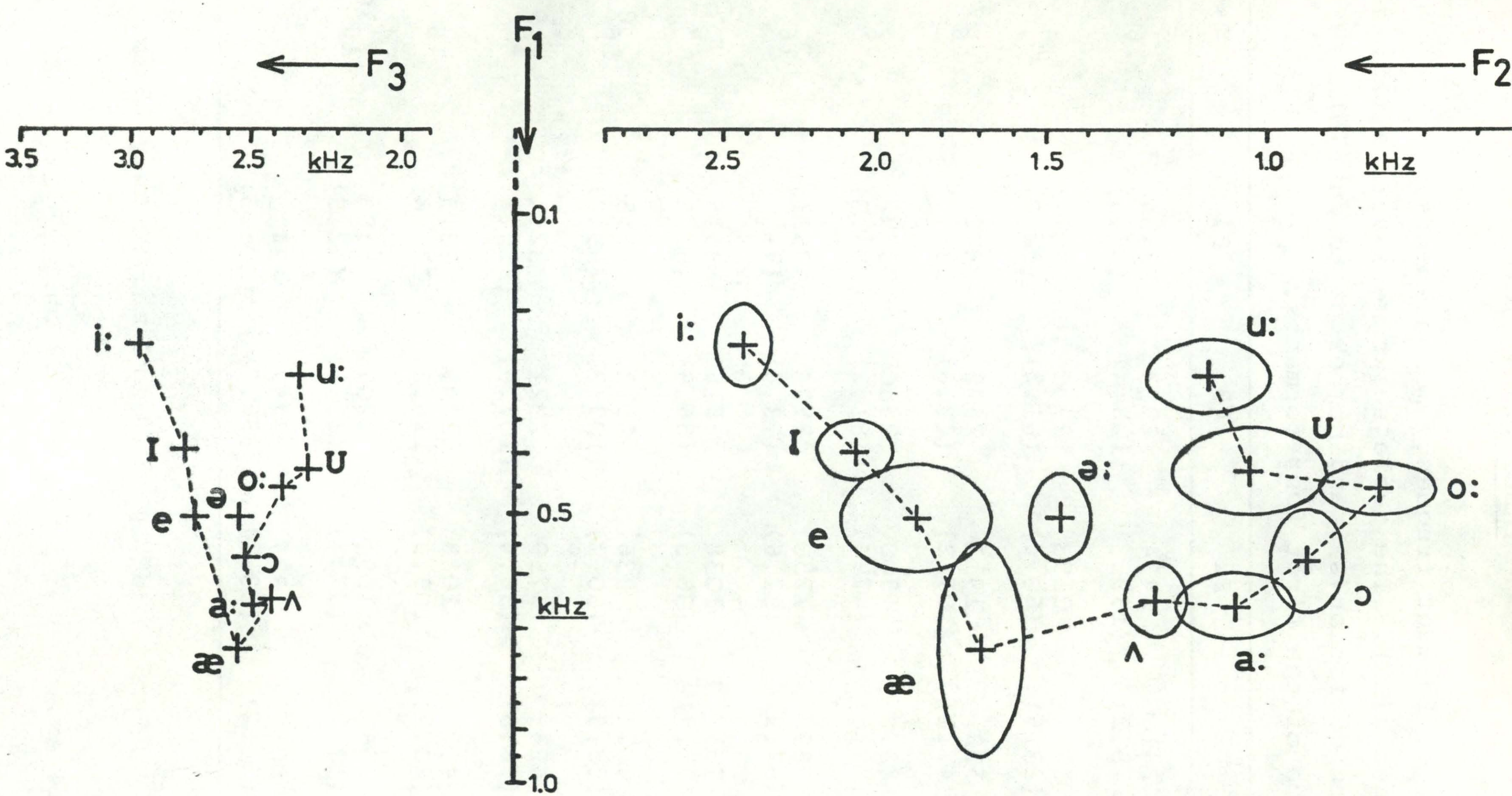



Fig. 2. Vowel diagram Informant 1(AW)

TABLE II

Results of formant frequency measurements. Informant 2(GG). The table shows mean values ($\overline{X}$), standard deviation(s), and number of tokens (N) for each of the first four formants.

| | | F1 | F2 | F3 | F4 | N |
|---|---|---|---|---|---|---|
| [i:] - | $\overline{X}$: | 241 | 2434 | 2933 | 3534 | 16 |
| | s: | (18.2) | (139.3) | (128.4) | (39.6) | |
| [ɪ] - | $\overline{X}$: | 377 | 2153 | 2700 | 3697 | 16 |
| | s: | (23.5) | (60.4) | (61.2) | (82.0) | |
| [e] - | $\overline{X}$: | 532 | 1911 | 2622 | 3764 | 16 |
| | s: | (47.1) | (107.2) | (43.6) | (134.5) | |
| [æ] - | $\overline{X}$: | 887 | 1700 | 2680 | 3792 | 16 |
| | s: | (35.0) | (47.4) | (81.2) | (106.3) | |
| [ʌ] - | $\overline{X}$: | 789 | 1250 | 2533 | 3531 | 16 |
| | s: | (67.9) | (31.6) | (73.4) | (91.5) | |
| [a:] - | $\overline{X}$: | 761 | 1038 | 2616 | 3515 | 16[7] |
| | s: | (73.7) | (36.5) | (56.9) | (87.5) | |
| [ɔ] - | $\overline{X}$: | 617 | 936 | 2513 | 3515 | 16[8] |
| | s: | (55.6) | (30.2) | (109.3) | (69.3) | |
| [o:] - | $\overline{X}$: | 424 | 759 | 2321 | 3254 | 16[9] |
| | | (28.8) | (36.4) | (107.6) | (119.2) | |
| [U] - | $\overline{X}$: | 402 | 1088 | 2299 | 3313 | 16 |
| | s: | (20.7) | (109.9) | (116.2) | (124.8) | |
| [u:] - | $\overline{X}$: | 256 | 1091 | 2207 | 3483 | 16[10] |
| | s: | (17.7) | (77.4) | (85.8) | (42.5) | |
| [ə:] - | $\overline{X}$: | 508 | 1377 | 2452 | 3420 | 16 |
| | s: | (31.2) | (23.2) | (86.3) | (101.3) | |

7) F4 only 15 ex.

8) F3 and F4 only 15 ex.

9) F3 only 12 ex., F4 only 14 ex.

10) F3 only 15 ex.

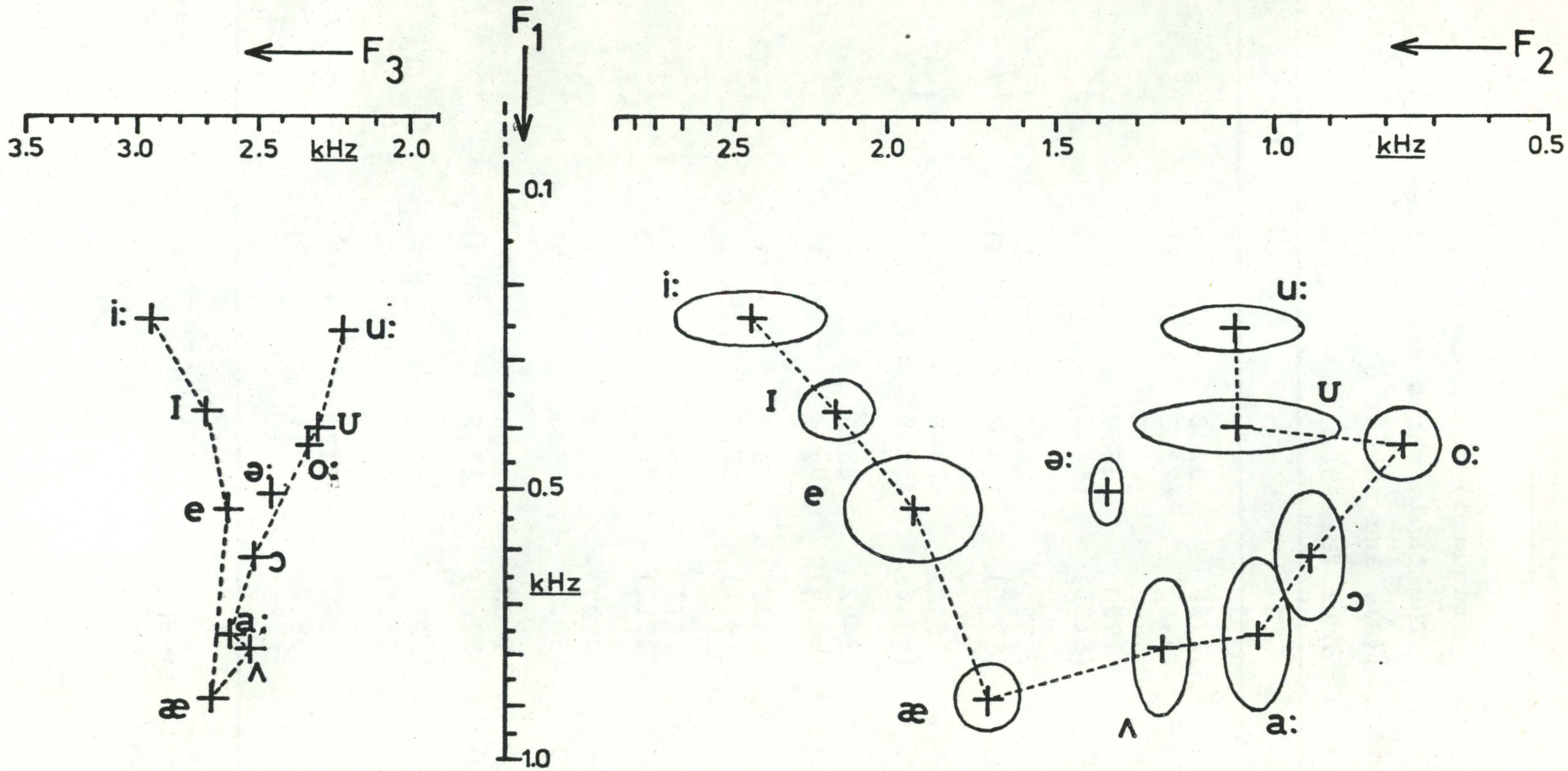



Fig. 3. Vowel diagram Informant 2(GG)

TABLE III

Results of formant frequency measurements. Informant 3(CLB). The table shows mean values ($\overline{X}$), standard deviation(s), and number of tokens (N) for each of the first four formants.

| | | F1 | F2 | F3 | F4 | N |
|---|---|---|---|---|---|---|
| [i:] - | $\overline{X}$: | 238 | 2258 | 3208 | 3652 | 12 |
| | s: | (26.8) | (54.7) | (73.3) | (94.8) | |
| [ɪ] - | $\overline{X}$: | 368 | 1983 | 2650 | 3690 | 12 |
| | s: | (27.6) | (61.5) | (54.4) | (58.8) | |
| [e] - | $\overline{X}$: | 493 | 1881 | 2617 | 3835 | 12 |
| | s: | (29.4) | (57.5) | (34.2) | (185.4) | |
| [æ] - | $\overline{X}$: | 648 | 1842 | 2490 | 3807 | 12[11] |
| | s: | (46.3) | (62.6) | (55.9) | (137.4) | |
| [ʌ] - | $\overline{X}$: | 727 | 1402 | 2502 | 3865 | 12 |
| | s: | (62.1) | (62.6) | (96.8) | (97.4) | |
| [a:] - | $\overline{X}$: | 672 | 1027 | 2607 | 3573 | 12[12] |
| | s: | (62.7) | (29.1) | (152.9) | (252.1) | |
| [ɔ] - | $\overline{X}$: | 624 | 977 | 2573 | 3388 | 12[13] |
| | s: | (33.9) | (27.1) | (82.2) | (195.5) | |
| [o:] - | $\overline{X}$: | 489 | 727 | 2642 | 3458 | 12 |
| | s: | (36.5) | (60.7) | (69.3) | (92.5) | |
| [U] - | $\overline{X}$: | 411 | 1000 | 2204 | 3416 | 12 |
| | s: | (35.0) | (50.0) | (85.8) | (61.3) | |
| [u:] - | $\overline{X}$: | 275 | 835 | 2158 | 3409 | 12[11] |
| | s: | (30.5) | (75.0) | (49.2) | (76.0) | |
| [ə:] - | $\overline{X}$ | 548 | 1360 | 2473 | 3579 | 12 |
| | s: | (21.8) | (48.2) | (37.6) | (152.9) | |

11) F4 only 11 ex.
12) F3 only 11 ex., F4 only 10 ex.
13) F4 only 10 ex.

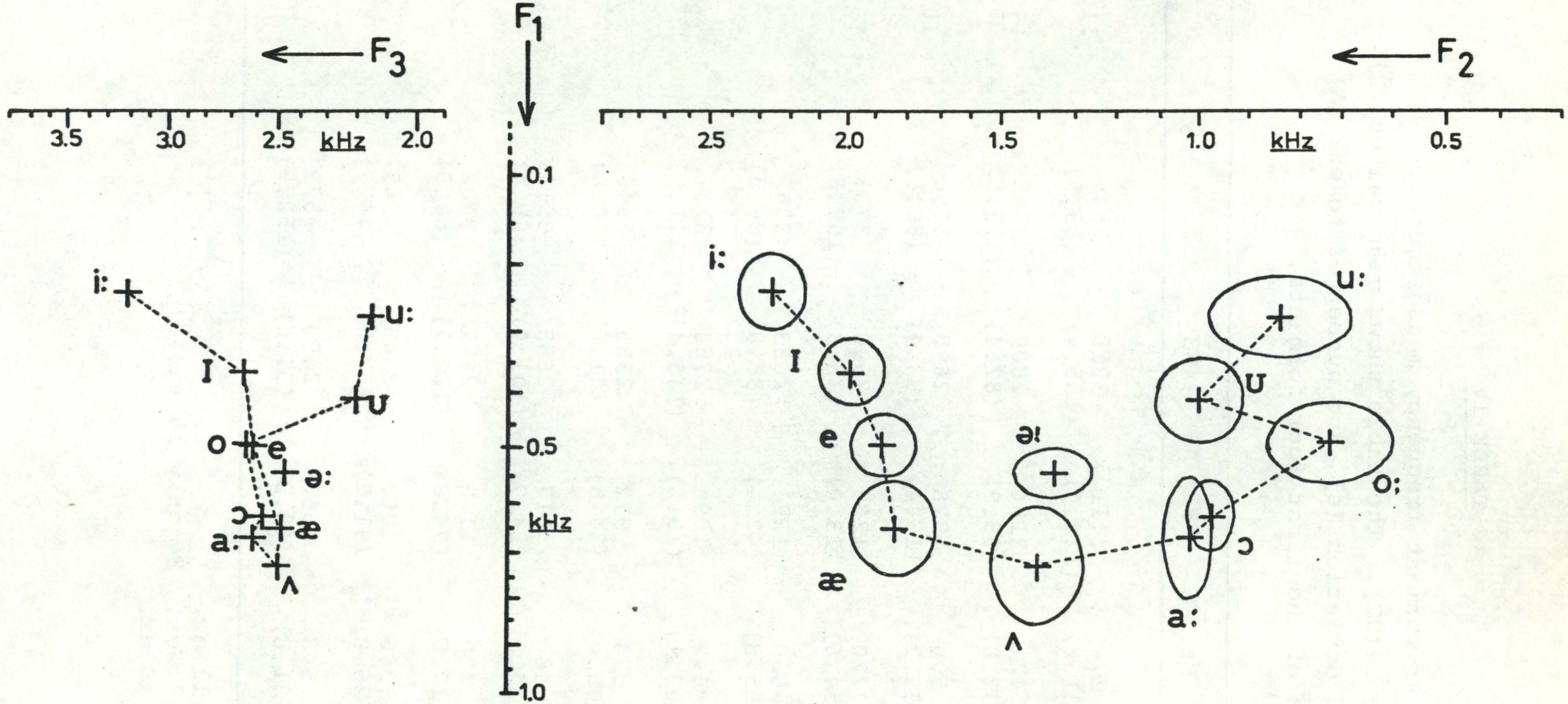



Fig. 4. <u>Vowel diagram</u> Informant 3(CLB)

TABLE IV

Results of formant frequency measurements. Informant 4(DC). The table shows mean values ($\overline{X}$), standard deviation (s), and number of tokens (N) for each of the first four formants.

| | | F1 | F2 | F3 | F4 | N |
|---|---|---|---|---|---|---|
| [i:] - | $\overline{X}$ | 274 | 2411 | 3218 | 3846 | 11 |
| | s: | (21.2) | (124.7) | (155.4) | (83.5) | |
| [ɪ] - | $\overline{X}$ | 421 | 2184 | 2809 | 3789 | 11 |
| | s: | (33.8) | (88.9) | (82.4) | (117.5) | |
| [e] - | $\overline{X}$: | 500 | 1928 | 2695 | 3758 | 10 |
| | s: | (32.2) | (116.4) | (93.4) | (80.8) | |
| [æ] - | $\overline{X}$: | 720 | 1739 | 2590 | 3729 | 13[14] |
| | s: | (54.6) | (113.0) | (78.8) | (169.8) | |
| [ʌ] - | $\overline{X}$: | 660 | 1271 | 2633 | 3802 | 12 |
| | s: | (55.8) | (42.4) | (88.8) | (88.8) | |
| [a:] - | $\overline{X}$: | 642 | 1094 | 2756 | 3731 | 12[15] |
| | s: | (33.1) | (15.5) | (145.4) | (111.9) | |
| [ɔ] - | $\overline{X}$: | 574 | 972 | 2588 | 3646 | 12[16] |
| | s: | (45.0) | (34.5) | (81.5) | (70.6) | |
| [o:] - | $\overline{X}$: | 480 | 817 | 2635 | 3519 | 12 |
| | s: | (25.8) | (54.7) | (101.9) | (60.4) | |
| [U] - | $\overline{X}$: | 470 | 1142 | 2671 | 3602 | 12 |
| | s: | (17.5) | (74.9) | (113.2) | (80.8) | |
| [u:] - | $\overline{X}$: | 306 | 1094 | 2248 | 3269 | 12[14] |
| | s: | (26.8) | (91.2) | (57.8) | (74.7) | |
| [ə:] - | $\overline{X}$: | 548 | 1360 | 2473 | 3579 | 12 |
| | s: | (21.8) | (48.2) | (37.6) | (152.9) | |

14) F4 only 11 ex.
15) F3 only 11 ex., F4 only 10 ex.
16) F4 only 10 ex.

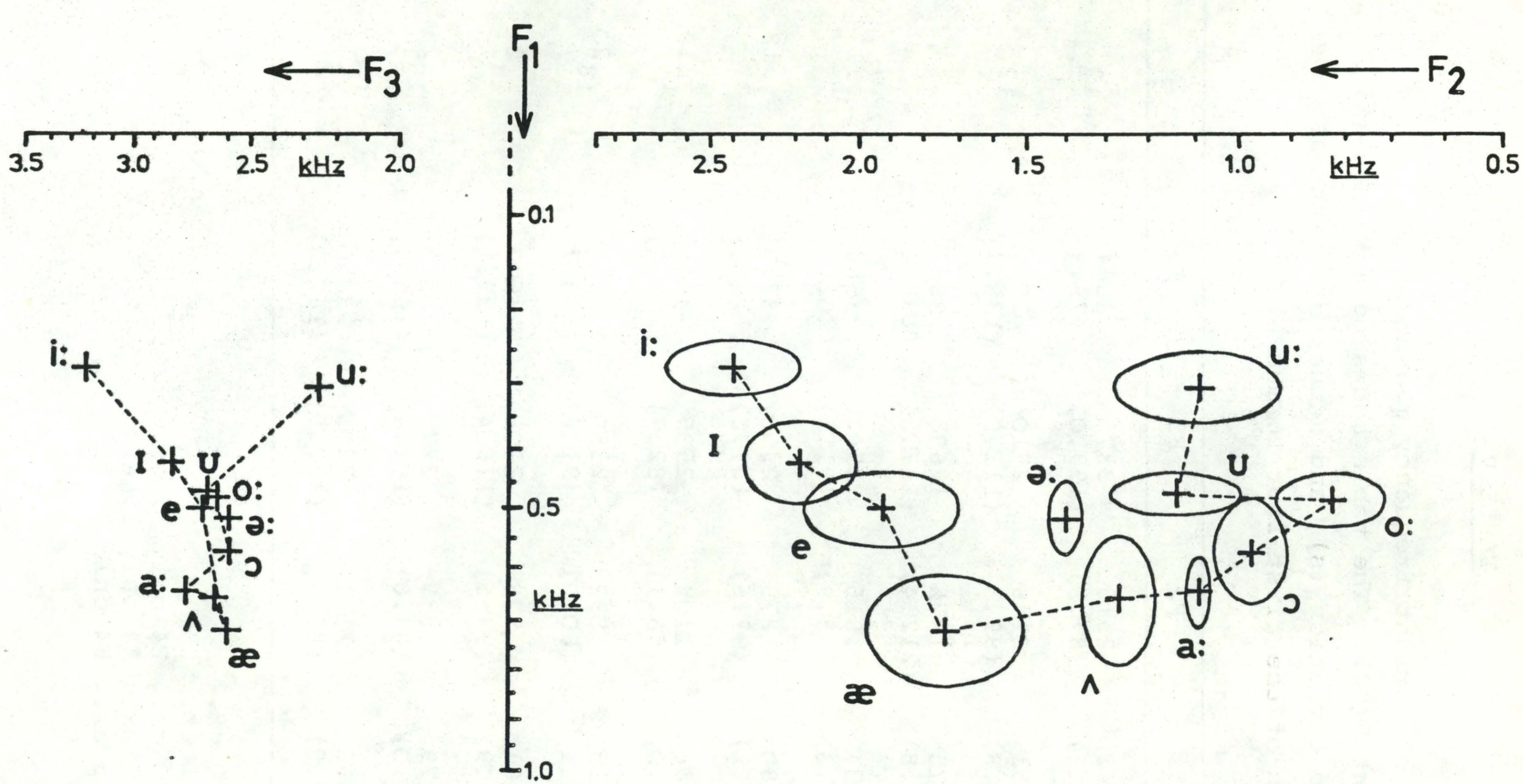



Fig. 5. Vowel diagram Informant 4(DC)

TABLE V

Results of formant frequency measurements. Informant 5(DH). The table shows mean values ($\overline{X}$), standard deviation (s), and number of tokens (N) for each of the first four formants.

| | | F1 | F2 | F3 | F4 | N |
|---|---|---|---|---|---|---|
| [i:] - | $\overline{X}$: | 243 | 2515 | 3373 | 3844 | 12 |
| | s: | (17.1) | (87.5) | (199.0) | (79.9) | |
| [ɪ] - | $\overline{X}$: | 351 | 2146 | 2710 | 3815 | 12 |
| | s: | (52.8) | (99.3) | (111.0) | (138.4) | |
| [e] - | $\overline{X}$: | 475 | 1975 | 2635 | 3804 | 12 |
| | s: | (39.8) | (212.4) | (147.5) | (138.4) | |
| [æ] - | $\overline{X}$: | 634 | 1815 | 2542 | 3688 | 12[17] |
| | s: | (38.3) | (55.9) | (91.3) | (202.5) | |
| [ʌ] - | $\overline{X}$: | 680 | 1279 | 2450 | 3723 | 12[18] |
| | s: | (74.2) | (97.6) | (157.8) | (179.0) | |
| [a:] - | $\overline{X}$: | 686 | 1104 | 2535 | 3741 | 12[19] |
| | s: | (88.0) | (53.1) | (82.7) | (200.7) | |
| [ɔ] - | $\overline{X}$: | 615 | 892 | 2350 | 3815 | 18[20] |
| | s: | (48.4) | (60.0) | (91.4) | (113.7) | |
| [o:] - | $\overline{X}$: | 361 | 538 | 2306 | 3607 | 18[21] |
| | s: | (75.9) | (94.5) | (116.4) | (137.1) | |
| [ʊ] - | $\overline{X}$: | 379 | 889 | 2227 | 3713 | 18[21] |
| | s: | (35.0) | (134.8) | (101.4) | (166.9) | |
| [u:] - | $\overline{X}$: | 263 | 1101 | 2235 | 3750 | 18[19] |
| | s: | (16.9) | (83.4) | (102.9) | (162.8) | |

17) F4 only 10 ex.
18) F4 only 11 ex.
19) F3 only 10 ex., F4 only 11 ex.
20) F3 only 17 ex., F4 only 15 ex.
21) F3 only 16 ex., F4 only 14 ex.

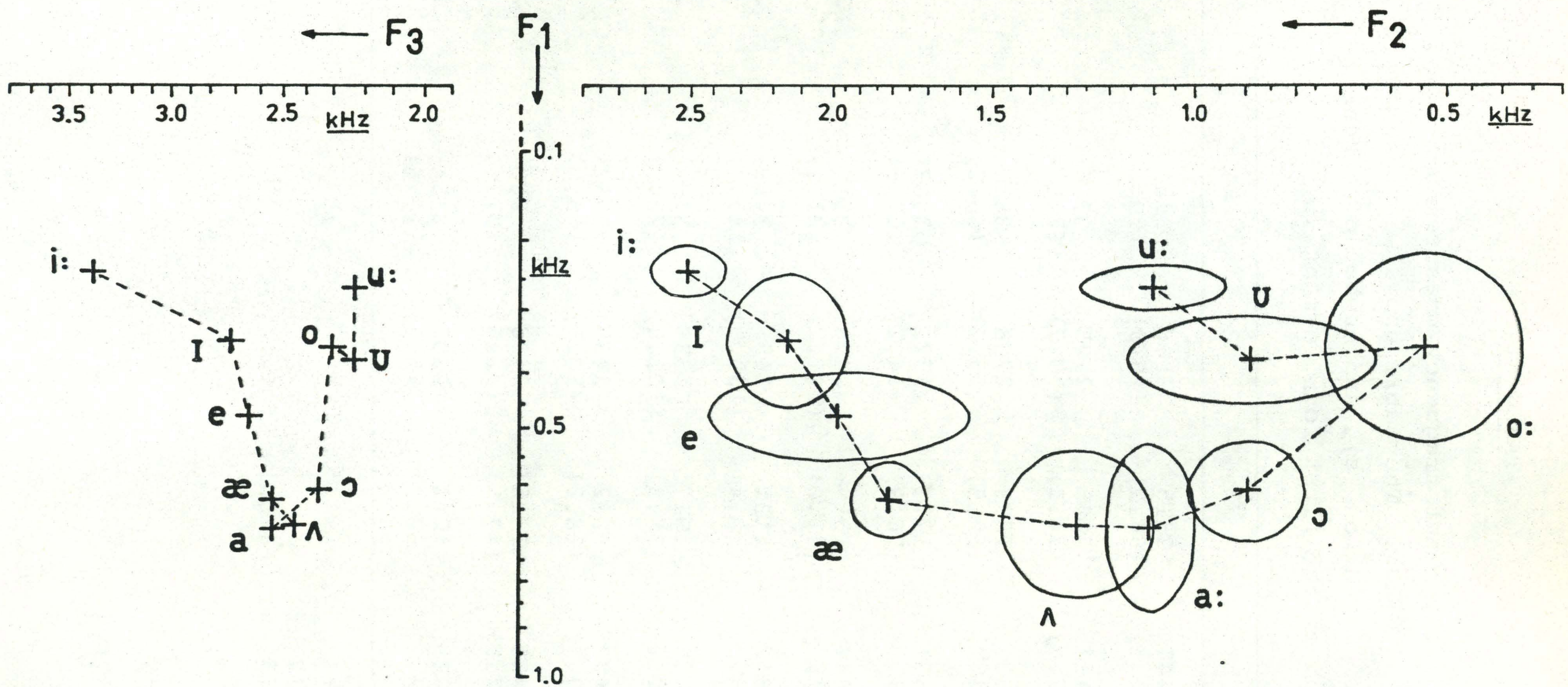



Fig. 6. Vowel diagram Informant 5 (DH)

TABLE VI

Results of formant frequency measurements. Informant 6(RD). The table shows mean values ($\overline{X}$), standard deviation (s), and number of tokens (N) for each of the first four formants.

| | | F1 | F2 | F3 | F4 | N |
|---|---|---|---|---|---|---|
| [i:] - | $\overline{X}$: | 245 | 2515 | 2808 | 3988 | 12 |
| | s: | (11.1) | (66.9) | (46.9) | (90.1) | |
| [ɪ] - | $\overline{X}$: | 414 | 2085 | 2725 | 3706 | 12 |
| | s: | (37.4) | (74.2) | (70.7) | (70.0) | |
| [e] - | $\overline{X}$: | 526 | 1898 | 2708 | 3810 | 12 |
| | s: | (24.7) | (69.5) | (56.7) | (65.2) | |
| [æ] - | $\overline{X}$: | 831 | 1773 | 2721 | 3802 | 12[22] |
| | s: | (66.8) | (56.9) | (54.2) | (67.5) | |
| [ʌ] - | $\overline{X}$: | 842 | 1329 | 2498 | 3514 | 12[22] |
| | s: | (51.0) | (41.0) | (141.2) | (121.1) | |
| [a:] | $\overline{X}$: | 840 | 1131 | 2456 | 3468 | 12[22] |
| | s: | (65.3) | (30.4) | (139.0) | (145.4) | |
| [ɔ̇] - | $\overline{X}$: | 565 | 1017 | 2333 | 3407 | 18[23] |
| | s: | (20.4) | (38.3) | (101.8) | (128.0) | |
| [o:] - | $\overline{X}$: | 489 | 892 | 2435 | 3552 | 12 |
| | s: | (19.4) | (28.9) | (50.5) | (95.0) | |
| [U] - | $\overline{X}$: | 438 | 1018 | 2438 | 3280 | 18[24] |
| | s: | (32.5) | (47.6) | (92.6) | (107.4) | |
| [u:] | $\overline{X}$: | 269 | 1244 | 2303 | 3354 | 18[23] |
| | s: | (37.6) | (82.9) | (75.7) | (112.6) | |

22) F4 only 11 ex.
23) F4 only 17 ex.
24) F3 only 16 ex., F4 only 15 ex.

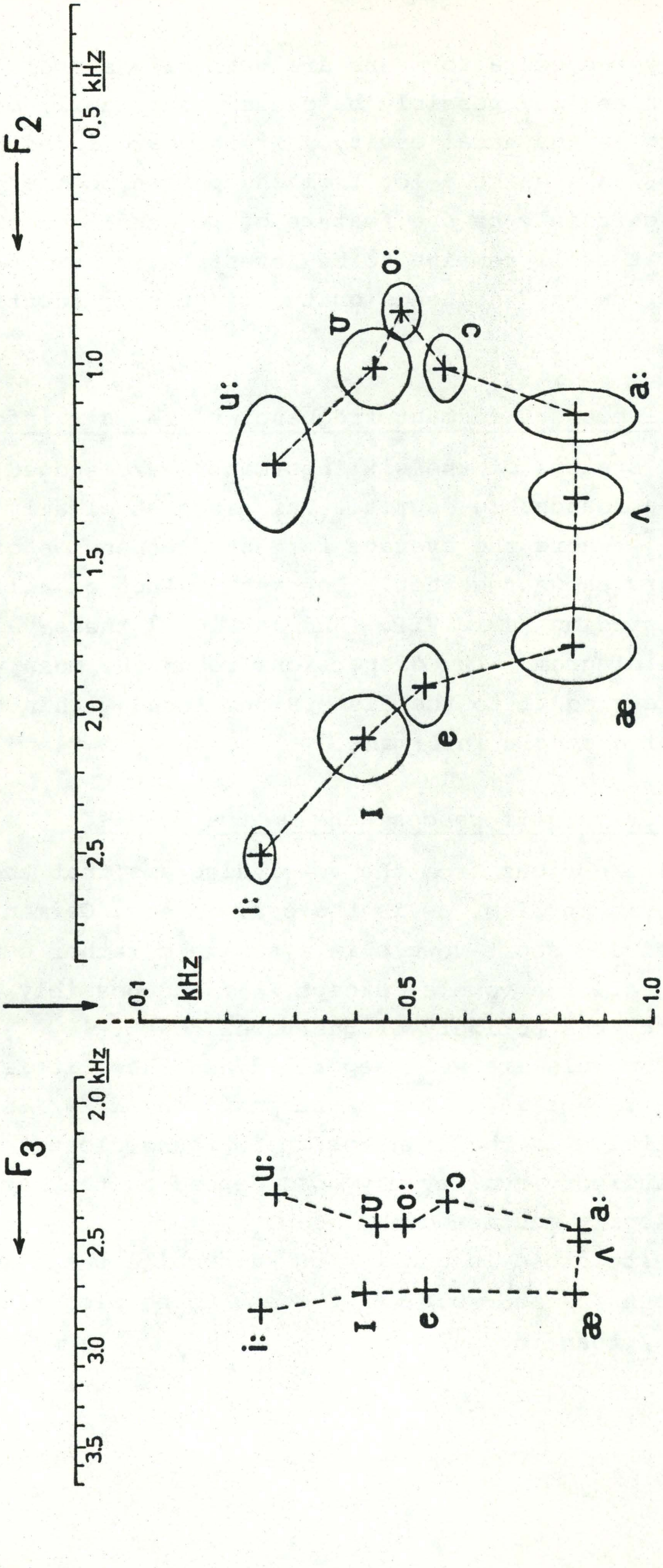



Fig. 7. Vowel diagram Informant 6 (RD)

Apparently the extra formants are primarily associated with the open vowels and may possibly be caused genetically by a slight coupling to the nasal cavity. If they are found at all their positions vary quite a lot from one person to the other, and they must certainly be one feature of personal voice quality. But it still remains to be investigated how far their presence has any influence on the linguistic identity of the vowels.

### 3.3.3. Overall mean of formant frequencies from six informants

The vowel systems of the six informants were judged to be comparable to a reasonable degree. This is most clearly revealed by Fig. 8 where the average formant frequencies of all six subjects are shown together. The mean values of all six systems are listed in Table VII. And in Fig. 9 these values are shown in a diagram. The dispersions round the mean values are approximately equal to the dispersions found within the vowel system of a single informant.

### 3.3.4. Qualitative differences between the vowels

It is quite obvious from the vowel diagrams that there is no tendency in English, as is the case in e.g. German, to keep the relatively short vowels in a separate rather centralised system. All the vowels, except /ə:/ and possibly /U/, are placed along the perifery of the vowel chart.

The front vowels are well separated and show little or no overlapping in quality. The exact position of /ʌ/ is somewhat doubtful. In the diagrams it is found closest to /a:/ but the auditory quality may also depend on the extra formants previously mentioned.

/ɔ/ is quite close in quality to /a:/ while there is some distance between /ɔ/ and /o:/. /U/ seems to be closer in quality to /o:/ than to /u:/.

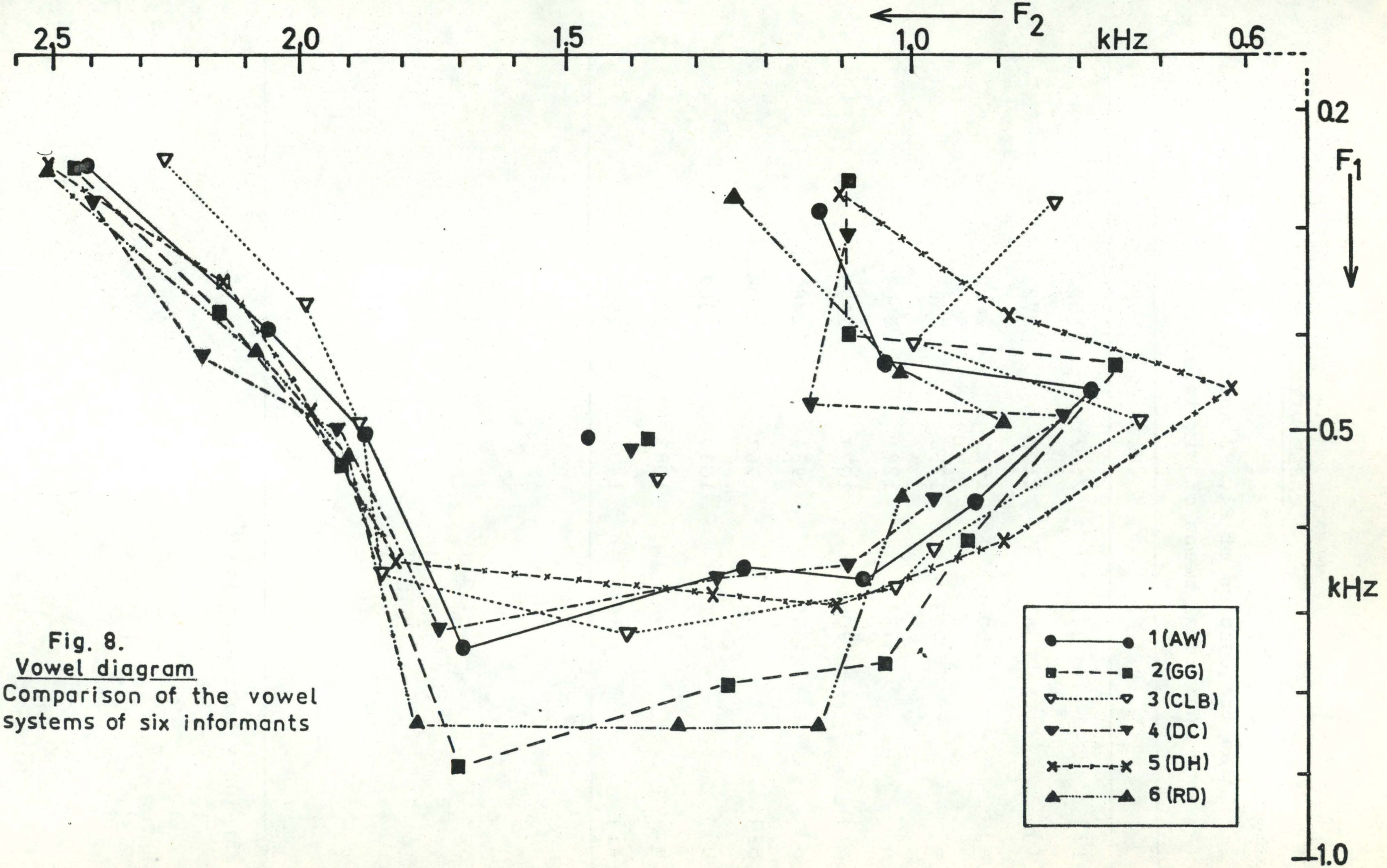



Fig. 8.
<u>Vowel diagram</u>
Comparison of the vowel systems of six informants

TABLE VII

Mean frequencies of F1, F2 and F3. Based on the individual means of six informants.[25]

| | F1 | F2 | F3 |
|---|---|---|---|
| [i:] | 247 | 2422 | 3082 |
| [ɪ] | 387 | 2102 | 2723 |
| [e] | 505 | 1911 | 2682 |
| [æ] | 743 | 1756 | 2561 |
| [ʌ] | 724 | 1294 | 2504 |
| [a:] | 709 | 1079 | 2573 |
| [ɔ] | 595 | 954 | 2480 |
| [o:] | 465 | 767 | 2409 |
| [ʊ] | 421 | 1030 | 2352 |
| [u:] | 276 | 1083 | 2243 |
| [ə:] | 514 | 1401 | 2511 |

25) Only four examples of [ə:].

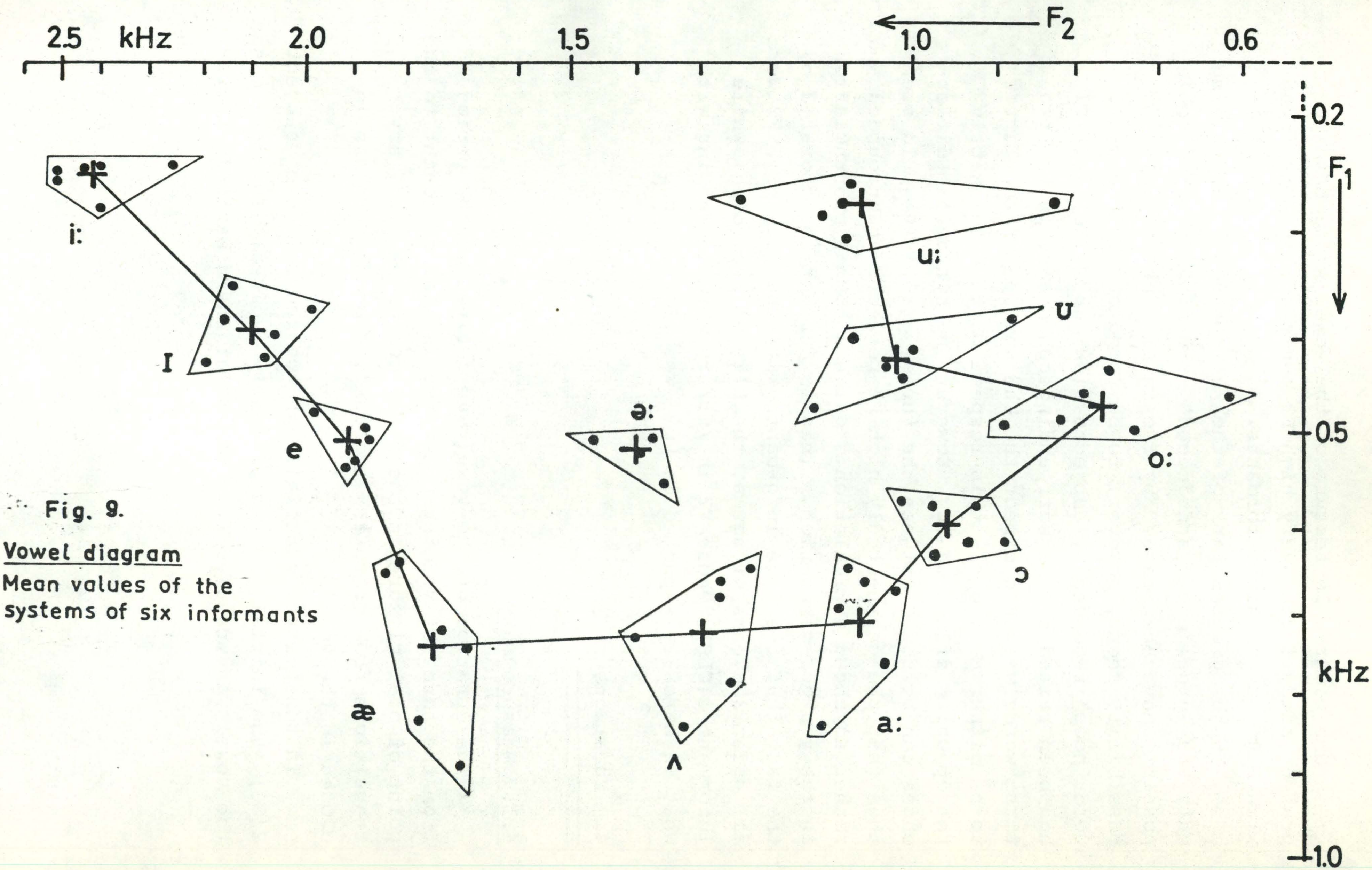



Fig. 9.

Vowel diagram

Mean values of the systems of six informants

One remarkable feature of the vowel systems in the present study is the very high F2 positions of /u:/ found with five out of six informants.

In the greater part of the /u:/-material, containing only the words "coot" and "cooed", the high position could possibly have been caused by the initial [k-]. This possibility seems, however, to be ruled out by the extra word "hoop" recorded by subjects 5(DH) and 6(RD). If F2 had been raised by the surrounding consonants in the two first words we would expect the [h] and [p] of the latter to be either neutral in this respect or to have exactly the opposite effect on F2. However, no systematic differences can be found between the three words. It would seem then that the unexpectedly high F2 reflects an unrounded and rather advanced articulation: quite a common pronunciation in the younger generation. This is further corroborated by the fact that the five informants with the very high F2 were all under 30 years of age at the time of recording, while informant 3(CLB), whose F2 in /u:/ is more in keeping with what is normally expected in [u], was 71 years old.

## 4. Durations

### 4.1. Measurements

The duration of the vowel was defined as the period from the onset of voicing after the initial consonant to the point of oral closure in the final consonant, i.e. any transitions from or to the surrounding consonants are included in the vowel.

In some cases, especially in the recordings of subject 2(GG), the initial [h] was partially voiced. In these cases the vowel was measured from the point where periodic vibrations appeared in the higher formants since the energy of the [h] appeared to be concentrated in the low frequency regions.

Before final voiceless consonants the oral closure was frequently obscured in the spectrograms by a rather strong glottal stop. Segmenting these words proved quite difficult since the glottal closure was often preceded by a period of creaky voice. Generally the vowel was then measured to the point when energy disappeared from the upper formants. But in a few cases the result was not wholly satisfactory.

The durations were measured with an accuracy of about 5 msec. And the results were run through the standard XFON-program, each vowel before [t] and before [d] taken separately.

### 4.2. Results

The results of the measurements are listed in Tables VIII-XIII for each of the six informants separately. The overall averages are listed in Table XIV.

#### 4.2.1. Division into long and short vowels

It is generally observed that there is no clearcut difference between English so-called long and short vowels. This is also true of the present material. Thus if all the vowels before [-t] are pooled (Fig.10a),the curve comes very close to a normal distribution. Dividing the vowels into traditionally long and short vowels produces two distinct distributions but still leaves a considerable overlapping of the two groups (Fig. 10b).

#### 4.2.2. Normalising duration with respect to quality

Part of this overlapping is caused by the tendency of close vowels to be shorter than open vowels. Therefore

TABLE VIII

Vowel durations in cs. Informant 1(AW).
The table shows mean value, standard deviation and number of tokens.

| | Vowel before [t] | | | Vowel before [d] | | |
|---|---|---|---|---|---|---|
| | $\overline{X}$ | s | N | $\overline{X}$ | s | N |
| [i:] | 12.2 | 1.41 | 8 | 27.4 | 1.02 | 8 |
| [ɪ] | 7.6 | 1.28 | 7 | 11.2 | 1.53 | 8 |
| [e] | 9.8 | 0.39 | 7 | 13.1 | 1.35 | 8 |
| [æ] | 10.8 | 0.88 | 8 | 15.3 | 1.79 | 8 |
| [ʌ] | 8.9 | 1.36 | 8 | 11.5 | 2.00 | 8 |
| [a:] | 17.5 | 1.04 | 8 | 29.0 | 1.91 | 8 |
| [ɔ] | 9.5 | 1.49 | 8 | 14.9 | 2.23 | 8 |
| [o:] | 15.7 | 1.19 | 8 | 28.6 | 2.54 | 8 |
| [U] | 8.4 | 1.66 | 8 | 11.3 | 1.46 | 8 |
| [u:] | 12.1 | 1.16 | 8 | 28.3 | 2.36 | 8 |
| [ə:] | 15.8 | 1.00 | 8 | 27.0 | 0.96 | 8 |

TABLE IX

Vowel durations in cs. Informant 2(GG)
The table shows mean value, standard deviation and number of tokens.

| | Vowel before [t] | | | Vowel before [d] | | |
|---|---|---|---|---|---|---|
| | $\overline{X}$ | s | N | $\overline{X}$ | s | N |
| [i:] | 12.7 | 1.19 | 8 | 21.3 | 1.39 | 8 |
| [ɪ] | 10.2 | 0.84 | 8 | 12.8 | 1.69 | 8 |
| [e] | 11.7 | 0.53 | 8 | 14.4 | 1.33 | 8 |
| [æ] | 13.1 | 1.46 | 8 | 17.1 | 1.36 | 8 |
| [ʌ] | 10.9 | 0.98 | 8 | 14.2 | 1.25 | 8 |
| [a:] | 17.4 | 0.35 | 8 | 23.5 | 1.25 | 8 |
| [ɔ] | 12.3 | 1.60 | 8 | 17.6 | 1.62 | 7 |
| [o:] | 16.5 | 1.32 | 7 | 23.1 | 1.35 | 8 |
| [ʊ] | 11.1 | 0.78 | 8 | 13.2 | 0.92 | 8 |
| [u:] | 12.7 | 1.44 | 8 | 21.4 | 1.57 | 7 |
| [ə:] | 17.3 | 1.39 | 8 | 22.0 | 1.41 | 8 |

TABLE X

Vowel durations in cs. Informant 3(CLB). The table shows mean value, standard deviation and number of tokens

| | Vowel before [t] | | | Vowel before [d] | | |
|---|---|---|---|---|---|---|
| | $\bar{X}$ | s | N | $\bar{X}$ | s | N |
| [i:] | 13.3 | 1.73 | 6 | 33.8 | 2.04 | 6 |
| [ɪ] | 7.8 | 0.82 | 6 | 12.9 | 1.59 | 6 |
| [e] | 10.2 | 1.03 | 6 | 15.6 | 0.59 | 6 |
| [æ] | 12.6 | 0.59 | 6 | 20.9 | 1.83 | 6 |
| [ʌ] | 9.2 | 1.63 | 6 | 14.9 | 1.20 | 6 |
| [a:] | 19.3 | 1.41 | 6 | 34.3 | 2.66 | 6 |
| [ɔ] | 11.8 | 0.94 | 6 | 18.3 | 0.93 | 6 |
| [o:] | 18.0 | 1.52 | 6 | 35.3 | 3.31 | 6 |
| [ʊ] | 8.3 | 1.25 | 6 | 14.9 | 1.53 | 6 |
| [u:] | 15.2 | 1.47 | 6 | 35.6 | 1.39 | 6 |
| [ə] | 17.3 | 1.70 | 6 | 32.7 | 1.78 | 6 |

## TABLE XI

Vowel durations in cs. Informant 4(DC). The table shows mean value, standard deviation and number of tokens

| | Vowel before [t] | | | Vowel before [d] | | |
|---|---|---|---|---|---|---|
| | $\overline{X}$ | s | N | $\overline{X}$ | s | N |
| [i:] | 16.0 | 2.21 | 6 | 32.8 | 1.68 | 5 |
| [I] | 12.4 | 0.74 | 5 | 20.8 | 1.99 | 6 |
| [e] | 17.3 | 1.86 | 5 | 24.5 | 5.08 | 4 |
| [æ] | 19.3 | 2.28 | 5 | 25.0 | 2.88 | 6 |
| [ʌ] | 13.0 | 2.55 | 6 | 21.1 | 1.28 | 6 |
| [a:] | 25.4 | 3.26 | 6 | 37.9 | 6.28 | 6 |
| [ɔ] | 16.3 | 1.21 | 6 | 21.6 | 1.80 | 6 |
| [o:] | 23.0 | 1.87 | 6 | 38.1 | 2.20 | 6 |
| [U] | 13.4 | 1.77 | 6 | 22.5 | 3.62 | 6 |
| [u:] | 15.5 | 1.61 | 6 | 35.7 | 2.81 | 6 |
| [ə:] | 22.3 | 2.21 | 6 | 39.1 | 2.08 | 6 |

TABLE XII

Vowel durations in cs. Informant 5(DH)
The table shows mean value, standard deviation and number of tokens

| | Vowel before [t] | | | Vowel before [d] | | |
|---|---|---|---|---|---|---|
| | $\overline{X}$ | s | N | $\overline{X}$ | s | N |
| [i:] | 9.8 | 1.72 | 6 | 32.8 | 4.80 | 6 |
| [ɪ] | 6.1 | 1.16 | 6 | 10.8 | 0.82 | 6 |
| [e] | 7.6 | 0.81 | 6 | 17.0 | 1.70 | 6 |
| [æ] | 9.6 | 0.86 | 6 | 15.7 | 2.07 | 6 |
| [ʌ] | 8.0 | 0.89 | 6 | 14.3 | 1.61 | 6 |
| [a:] | 17.3 | 1.78 | 6 | 34.8 | 2.75 | 6 |
| [ɔ] | 7.9 | 0.67 | 6 | 15.0 | 0.84 | 6 |
| [o:] | 15.9 | 1.53 | 6 | 38.0 | 2.17 | 6 |
| [ʊ] | 5.9 | 0.67 | 6 | 11.8 | 2.12 | 6 |
| [u:] | 10.6 | 1.36 | 6 | 38.4 | 8.21 | 6 |

TABLE XIII

Vowel durations in cs. Informant 6(RD).
The table shows mean value, standard
deviation and number of tokens

| | Vowel before [t] | | | Vowel before [d] | | |
|---|---|---|---|---|---|---|
| | $\overline{X}$ | s | N | $\overline{X}$ | s | N |
| [i:] | 12.9 | 1.53 | 6 | 24.7 | 3.68 | 6 |
| [I] | 9.3 | 0.98 | 6 | 16.1 | 1.43 | 6 |
| [e] | 10.5 | 0.89 | 6 | 15.4 | 1.93 | 6 |
| [æ] | 14.2 | 1.21 | 6 | 20.1 | 1.07 | 6 |
| [ʌ] | 11.0 | 1.48 | 6 | 18.5 | 2.32 | 6 |
| [a:] | 22.2 | 1.51 | 6 | 29.9 | 3.34 | 6 |
| [ɔ] | 13.9 | 1.72 | 6 | 20.1 | 1.72 | 6 |
| [o:] | 21.8 | 1.51 | 6 | - | - | 0 |
| [U] | 11.2 | 1.33 | 6 | 17.0 | 1.79 | 6 |
| [u:] | 13.3 | 1.54 | 6 | 25.0 | 2.03 | 6 |

TABLE XIV

Mean vowel durations before [t] and [d]. Based on the individual mean values from six informants

| | Duration before [t] | Duration before [d] |
|---|---|---|
| [i:] | 12.8 | 28.8 |
| [ɪ] | 8.9 | 14.1 |
| [e] | 11.2 | 16.7 |
| [æ] | 13.3 | 19.0 |
| [ʌ] | 10.2 | 15.7 |
| [a:] | 19.8 | 31.6 |
| [ɔ] | 11.9 | 17.9 |
| [o:] | 18.5 | 32.6 |
| [U] | 9.7 | 15.1 |
| [u:] | 13.2 | 30.7 |
| [ə:] | 18.2 | 30.2[26] |

26) Only four examples of [ə:].

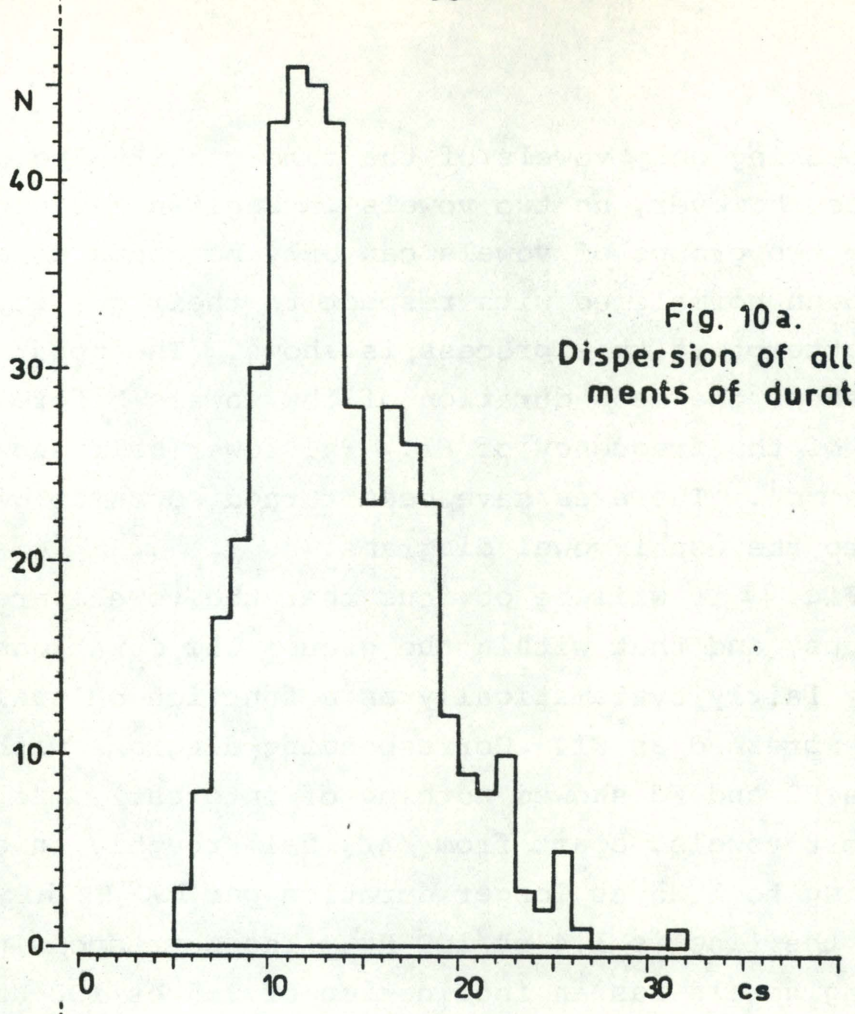



Fig. 10a.
Dispersion of all measurements of duration

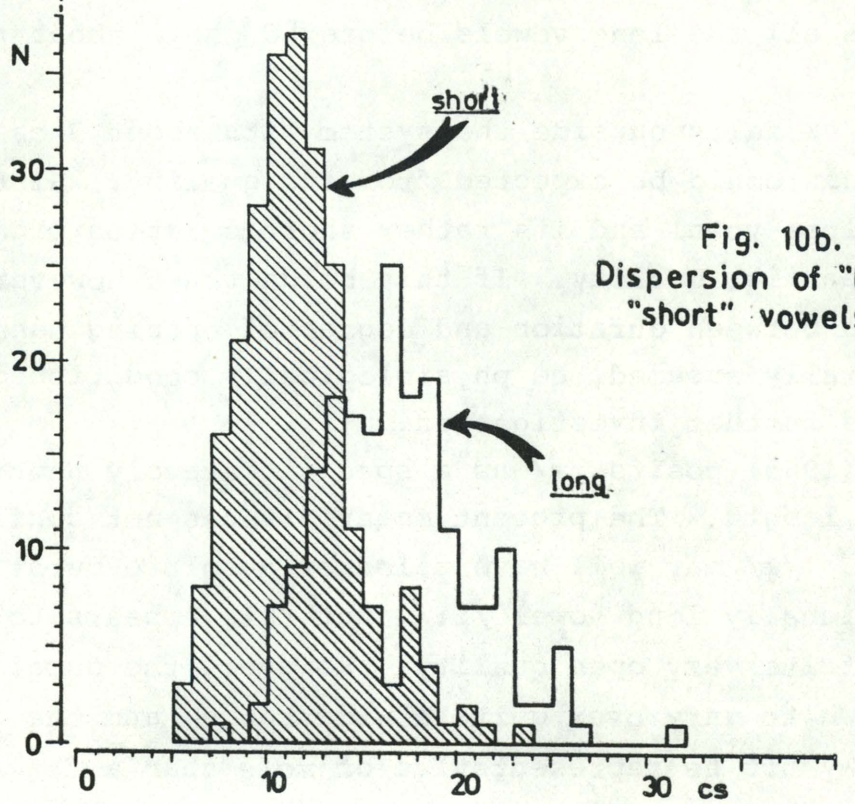



Fig. 10b.
Dispersion of "long" and "short" vowels

strictly speaking only vowels of the same quality are comparable. Since, however, no two vowels in English have the same quality the two groups of vowels can only be compared after they have been normalised with respect to their quality. In Fig.11 an attempt at this process is shown. The upper half of Fig.11 shows the mean duration of the vowels before [t] as a function of the frequency of F1. The lower half shows the same before [d]. The axes have been turned so as to give easy reference to the usual vowel diagrams.

From Fig. 11 it will be obvious that the vowels are divided in two groups, and that within the groups the durations of the vowels vary fairly systematically as a function of quality (in this case expressed as F1. Corresponding diagrams with duration versus F2 and F3 showed nothing of interest). Before [t] all the short vowels, apart from /ʌ/, fall roughly on a line corresponding to 1.25 cs longer duration per 100 Hz higher F1. Before [d] the line is 1.4 cs/100 Hz. The corresponding line for the long vowels has an inclination of 1.5 cs/100 Hz before [t] whereas all the long vowels before [d] have about the same duration.

Only /ʌ/ falls outside the system with about 3 cs shorter duration than could be expected from its quality. Historically /ʌ/ is a close vowel and its rather short duration might reflect its earlier quality. If this holds true, however, the correlation between duration and degree of opening cannot, as it is generally assumed, be physiologically conditioned. This point needs further investigation.

Wiik (1965) posits /æ/ as a special category neutral with respect to length. The present material does not confirm this hypothesis. /æ/ may well have a longer absolute duration than the traditionally long vowel /i:/, but this appears to be a function of its very open quality. However, the duration of /æ/ is known to vary over quite a wide range, and the present material may not be representative of more than a fraction of

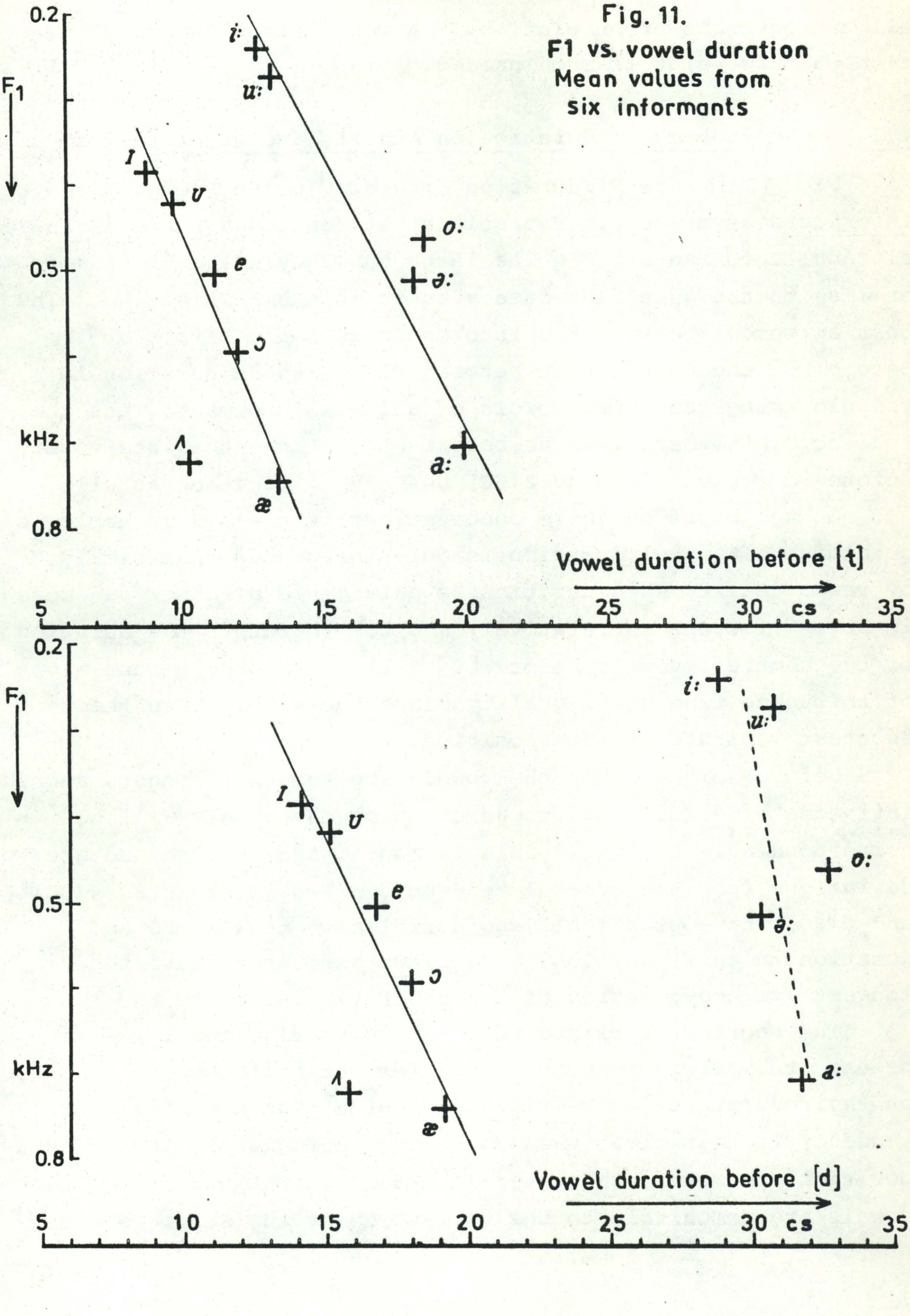



Fig. 11.
F1 vs. vowel duration
Mean values from
six informants

all the possible pronunciations. A more extensive investigation of this point is much needed.

4.2.3. Dependency of duration on quality and other factors

Figs. 12-15 show F1/duration diagrams for each of the six informants separately. Variations between the systems of individual informants are quite large but they all conform more or less to the general tendency as it is shown in Fig. 11. The best agreement between the informants is found before [t]. Before [d] the correlation between F1 and vowel duration is visible among the short vowels of all six informants, but the tendency is not as clear as before [t]. Among the long vowels before [d] only informant 2(GG) has any correlation at all.

On the basis of these observations it seems reasonable to conclude that vowel duration is only influenced appreciably by vowel quality when duration as determined by other and more important factors is relatively short. Therefore the duration of the shortest vowels before [t] will show the clearest signs of influence from vowel quality since the vowels have their shortest variants in this position.

Before voiced stops the vowels are somewhat longer, and the influence of quality on the duration of the vowels will be correspondingly smaller. This is manifested as more and greater deviations from the general tendency. (Two informants, 4(DC) and 6(RD), have practically no correlation between F1 and duration in this position. These two informants have the longest average duration of the short vowels before [d].)

The shortest variants of the long vowels, the variants before [t], are just short enough for the influence of quality on their durations to be visible. But at the same time the tendency is less clear than among the short vowels in the same position. However, the inter-informant variations of the long vowels are comparable to the variations of the short vowels before [d]: In this position the average duration of the short vowels is about equal to the average duration of the long vowels before [t].

Finally the long vowels before [d] are so long that any influence from quality on vowel duration has disappeared. Only the vowels of informant 2(GG), whose long vowels before [d] are very short indeed, show traces of the tendency..

## 4.2.4. Differences between long and short vowels

The mean duration of all the short vowels[27] amounts to 10.9 cs before [t] and 16.4 cs before [d]. The mean duration of the long vowels is 16.5 cs before [t] and 30.8 cs before [d].

These figures show that the two groups are not affected by the surrounding consonants in the same way. Thus in the present material the long vowels are 86.6 pct longer before [d] than before [t], while the short vowels are only 50.5 pct longer.

Another interesting point noticed by Wiik (1965) is that the lengthening effect of the voicing in the final consonant is of the same order of magnitude as the lengthening caused by the opposition between long and short vowels. This is confirmed in the present investigation. But it should be remembered that this rule applies only to the mean values of short and long vowels. Within these groups the durations of the individual vowels vary as a function of quality, and therefore no general percentage "lengthening" can be given which applies to all the individual vowels.

Thus it will be seen from Fig.11 that as long as the vowels are short enough for the influence of quality to be noticable the "lengthening effect" consists roughly of an added absolute value, the magnitude of which is independent of the "normal" duration of the vowel in question. For instance, the short vowels are all about 5 cs longer before [d] than before [t].

---

27) Including /æ/.

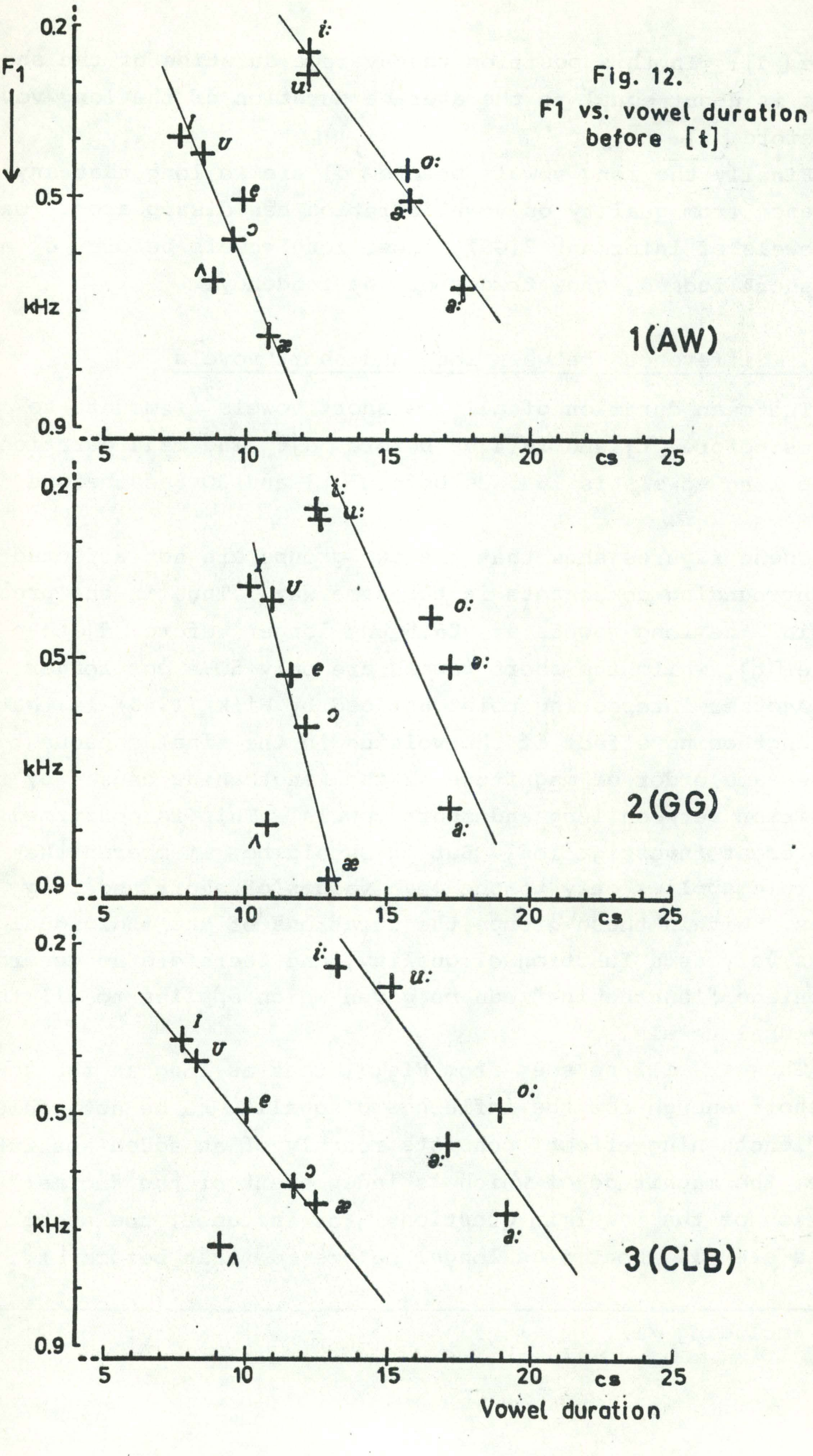


Fig. 12.
F1 vs. vowel duration
before [t]
F1
0.2
0.5
kHz
0.9
1(AW)
2(GG)
3(CLB)
5
10
15
20
cs
25
Vowel duration
i:
u:
I
U
e
ɔ
ʌ
æ
o:
ə:
a:

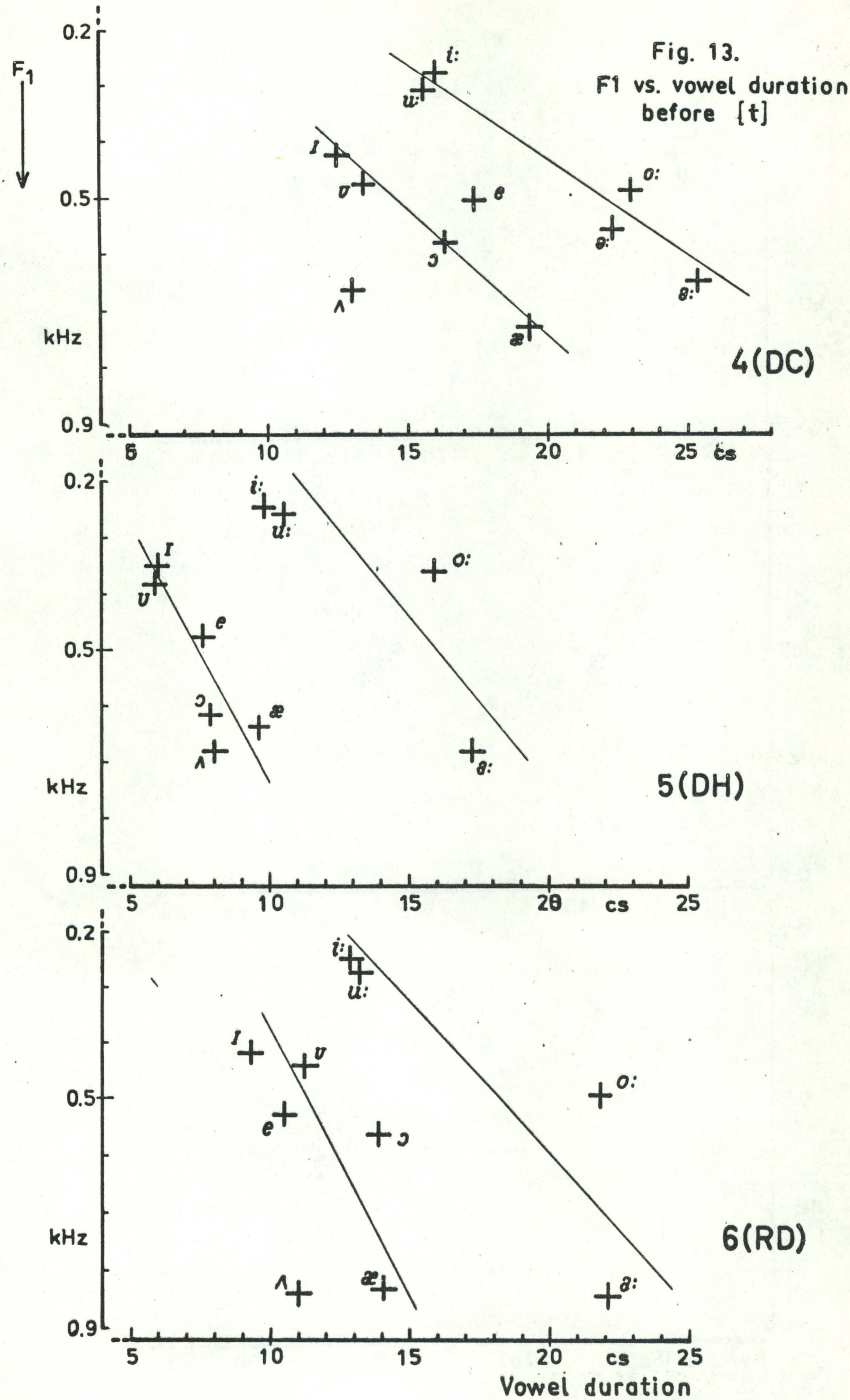



Fig. 13.
F1 vs. vowel duration
before [t]

Fig. 14. F1 vs. vowel duration before [d]

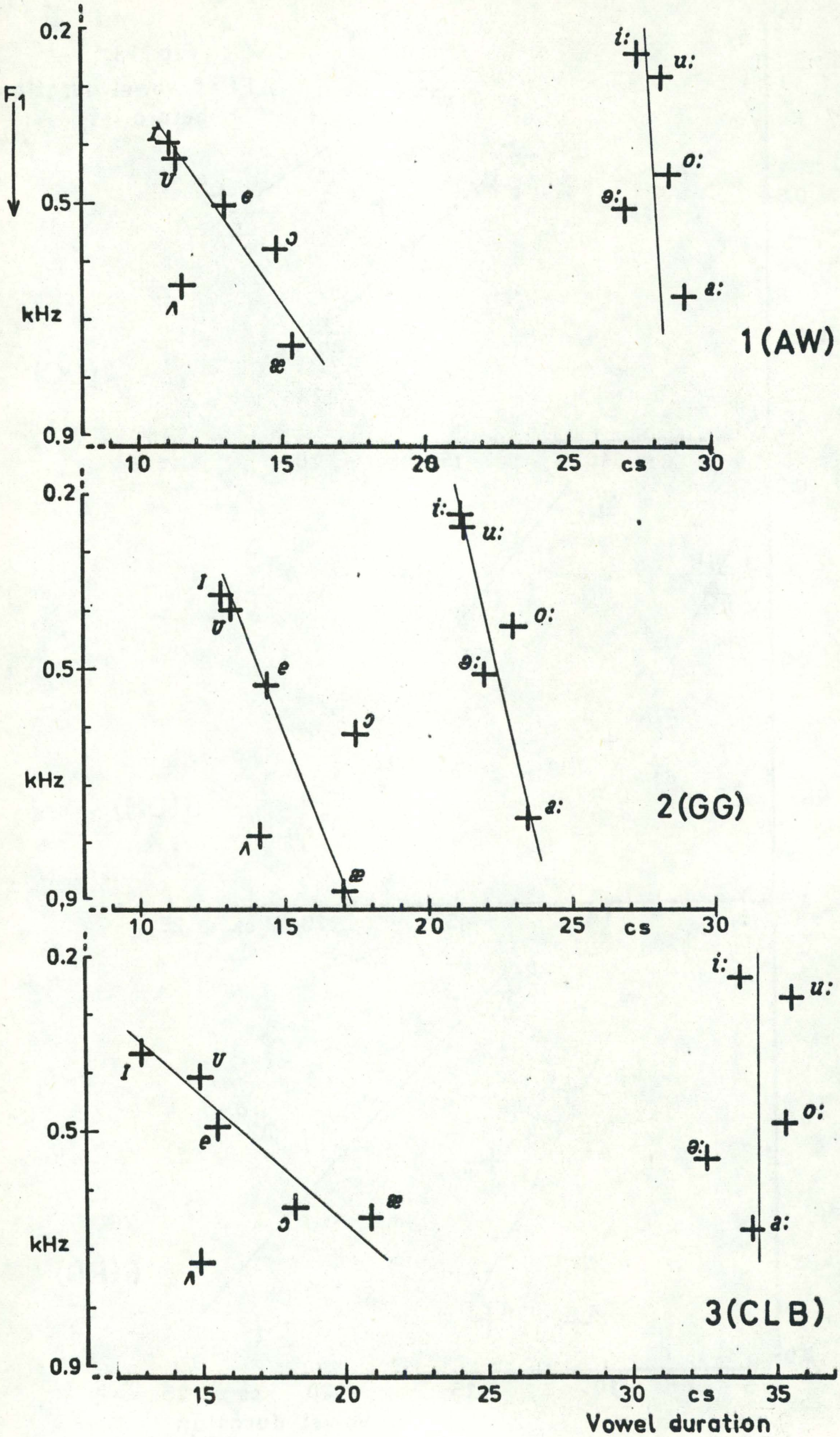
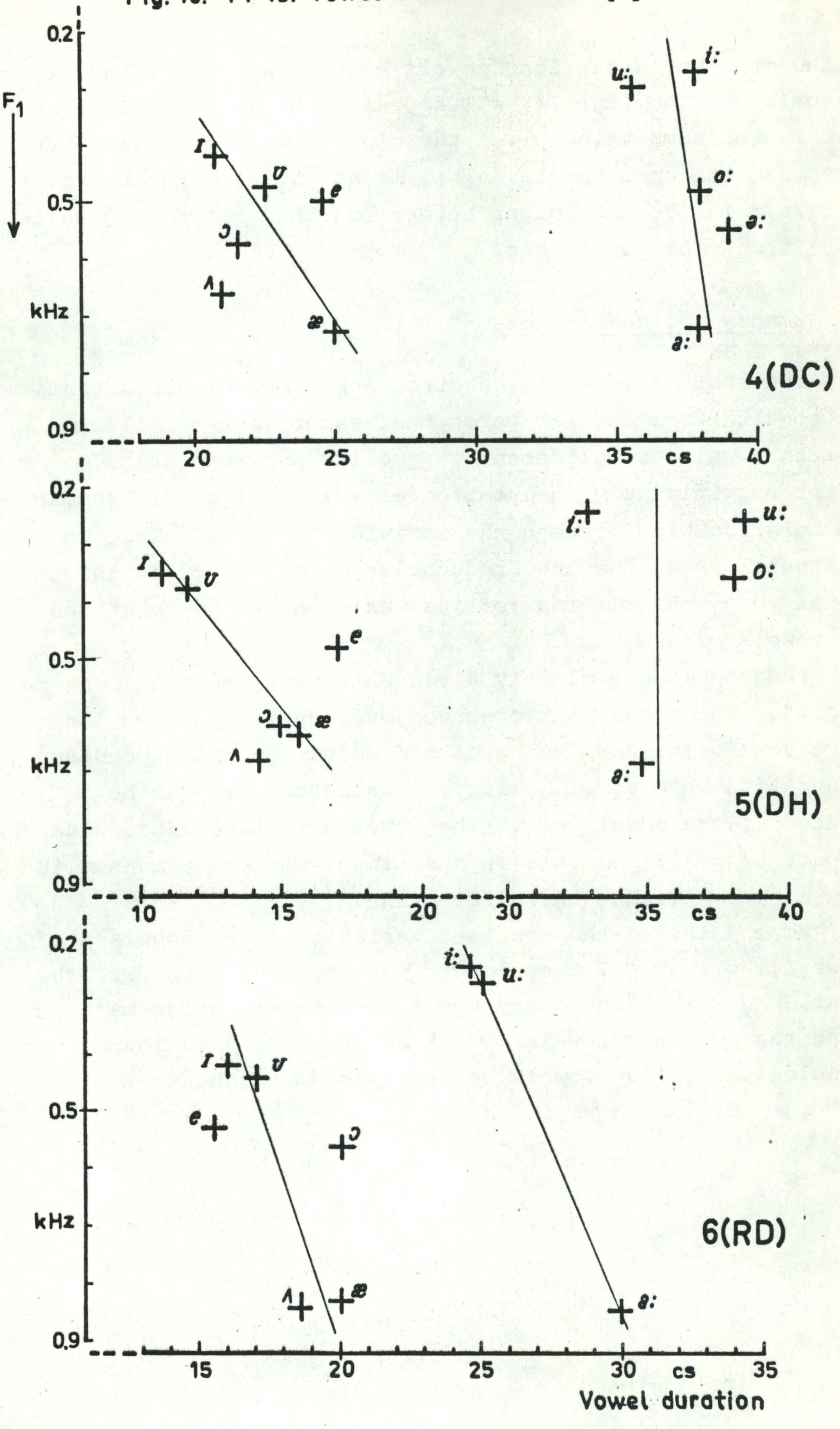


Fig. 15. F1 vs. vowel duration before [d]



On the other hand the long vowels before [d], where the effect of quality is negligible, all appear to be lengthened to approximately the same value, i.e. the close vowels are lengthened more than the open vowels. This means, to take two extremes, that /u:/ is 133 pct longer before [d] than before [t] while [a:] is only 59 pct longer!

## 5. Summary and conclusions

The results from the spectrographic measurements confirm the results from earlier perceptual experiments: All the English vowels are different in quality and some of them differ significantly in duration as well. Figure 16 summarises the relationships between the individual vowels. The two horizontal axes show the frequencies of F1 and F2 in the traditional vowel chart. The vertical axis shows the durations of the vowels before [t].

The vowels are clearly divided in two groups: "long" and "short". The frequently observed overlapping between long and short vowels is shown to be caused mainly by the dependency of duration upon vowel quality. Therefore it can be reduced if the durations of the vowels are normalised with respect to quality. This is done in a very crude manner in figure 16 by tilting the F1/F2-plane.

More detailed the shortest variants of the vowels are shown to be 1.25 - 1.5 cs longer per 100 Hz higher F1. The duration-quality dependency appears to be operating mainly among the shortest vowels. Thus no dependency is found among phonologically long vowels before voiced consonants.

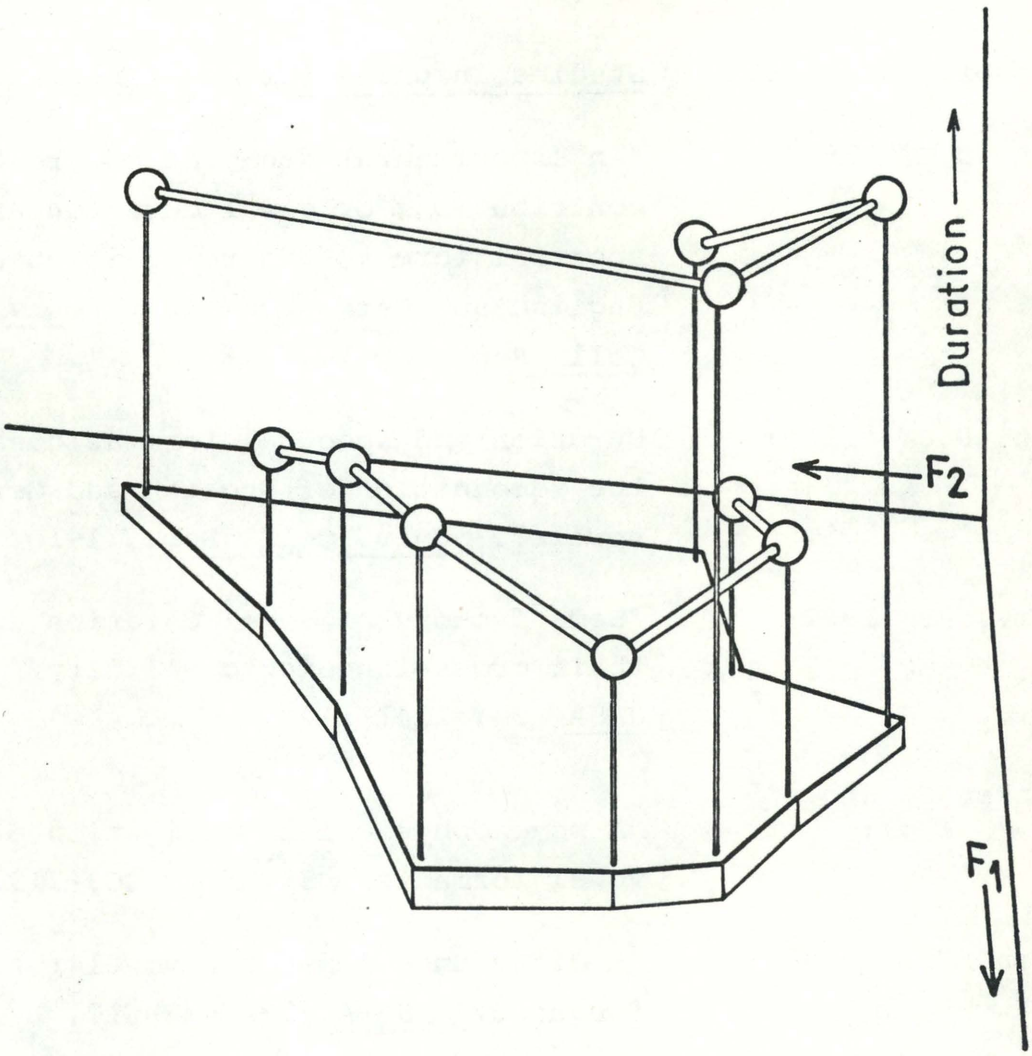



Fig. 16.

Three dimensional sketch of the distances between the individual vowels. The axes show the frequencies of F1 and F2 and the durations of the vowels before [t]. The scales used are about the same as those found in Figs. 12-15.

References


Abercrombie, D. 1965: Studies in phonetics and linguistics

Bennett, D.C. 1963: "An experimental study of the relative contributions of vowel duration and spectral form to the recognition of English and German vowels". Univ. Coll. Rep., 10-14.

Bennett, D.C. 1965: Duration and spectral form as cues in the recognition of English and German vowels". Univ. Coll. Rep., 1-10.

Delattre, P. 1962: "Some factors of vowel duration and their cross-linguistic validity" JASA 34, 1141-1143.

Fairbanks, G. and Grubb, P. 1961: "A psychophysical investigation of vowel formants" JSHR 4, 203-219.

Flanagan, J.L. 1955: "A difference limen for vowel formant frequency" JASA 27, 613-617.

Frøkjær-Jensen, B. 1963: "De danske langvokaler", Tale og Stemme, 2, 59-75.

Gimson, A.C. 1945-49: "Implications of the phonemic/chronemic grouping of English vowels" Acta Linguistica, 94-100.

Gimson, A.C. 1970: An introduction to the pronunciation of English (2nd ed.)

Heffner, R-M.S. 1937ff: "Notes on the length of vowels I". American Speech 128-135, II A.S. (1940 a) 74-80, III A.S. (1940 b) 377-381, IV A.S. (1941) 204-208, V A.S. (1942) 42-49, VI A.S. (1943) 208-216.

Heike, G. and Hall, R.D. 1969: "Vowel patterns of three English speakers: A comparative acoustic and auditive description" Linguistics, 54, 13-38.

House, A.S. 1961: "On vowel duration in English" JASA 33, 1174-1178.

House, A.S. and Fairbanks, G. 1953: "The influence of consonant environment upon the secondary acoustical characteristics of vowels" JASA 25, 105-113.

Jones, D. 1956: The pronunciation of English (4th ed.).

Jørgensen, H.P. 1969: "Die gespannten und ungespannten Vokale in der norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantfrequenzen", Phonetica, 19, 217-245.

Ladefoged, P. 1962 The nature of vowel quality (Coimbra).

Meyer, E.A. 1903: "Englische Lautdauer" Skrifter utg. af Kgl. Hum. Vetenskaps-Samf. i Uppsala, VIII, 3.

Peterson, G.E. and Barney, H.L. 1952: "Control methods used in a study of the vowels" JASA 24, 175-184.

Peterson, G.E. and Lehiste, I. 1960: "Duration of syllable nuclei in English". JASA 31, 428-435.

Stevens, K.N. 1959: "Effect of duration upon vowel identification" JASA 31, 109 (abstract)

Stevens, S.S. and Volkman, J. 1940: "The relation of pitch to frequency" Am. Journ. Psych. 53, 329-353.

Wells, J.C. 1963: "A study of the formants of the pure vowels of British English" Univ. Coll. Rep., 1-6.

Wiik, K. 1965: Finnish and English vowels.