W. Jassem, M. Krzyśko, P. Stolarski

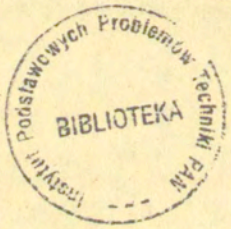
COMPUTER-AIDED CLASSIFICATION OF GENERAL BRITISH ENGLISH MONOPHTHONGS IN MONOSYLLABIC WORDS

WARSZAWA 1980

Praca wpłynęła do Redakcji dnia 19 sierpnia 1980 r.

Zarejestrowana pod nr 47/1980

Praca została wykonana w ramach problemu międzyresortowego I-24 podtemat 05.6



N a p r a w a c h r ę k o p i s u

Instytut Podstawowych Problemów Techniki PAN

Nakład 210 egz. Ark. wyd. 2,6 . Ark. druk. 4
Oddano do drukarni w listopadzie 1980 r.
Nr zamówienia 772/0/80

Warszawska Drukarnia Naukowa, Warszawa,
ul.Śniadeckich 8

Wiktor Jassem
Pracownia Fonetyki Akustycznej IPPT PAN
Mirosław Krzyśko
Przemysław Stolarski
Ośrodek Informatyki UAM w Poznaniu

COMPUTER-AIDED CLASSIFICATION OF GENERAL BRITISH ENGLISH MONOPHTHONGS IN MONOSYLLABIC WORDS

Abstract

The time-varying frequencies of the two lowest formants were measured at time intervals of 20 ms from spectrograms of 50 English monosyllabic words spoken by three native speakers of General British English. The data were processed in a general-purpose computer programmed to calculate the mean vectores and covariance matrices for each of the phonemes, separately for each speaker and also for data pooled for all three speakers. The two variables used were: (1) F_1 and F_2, (2) $\log F_1$ and $\log F_2$, (3) F_1 and $F_2 - F_1$. The mean vectors and covariance matrices were also calculated for conditions (2) and (3). The statistics were used to construct quadratic discriminant functions and vowel-classification charts. The charts, stored in the computer memory, were then used for the classification of each vocalic segment, represented as a trajectory in the vowel space by using a simple decision algorithm. If the charts were prepared separately for each of the three speakers, about 75% of the vocalic segments were correctly assigned to the appropriate (idio)phoneme. The three conditions gave almost identical overall results of the classification. The mean vectors, for all three conditions, were found to correspond very well to the description of the GBE monophthongs in terms of the IPA vowel quadrilateral.

1. The representation and the measurement of the signal.

As an acoustic event, the simplest vowel sound may be assumed to be

a periodic wave. The only deviation from strict periodicity, in the simplest case, is, or may be, the time limitation and gradual onset and decay. Such an 'ideal' vowel is almost entirely restricted to the laboratory experiment and is more likely to be the product of simplified synthesis in any of the numerous specialized electronic devices than the product of natural speech. Isolated vowels do occasionaly appear in natural speech, but they are then almost certain to deviate from strict periodicity also through the modulation of the fundamental frequency and the consequent variations in both the frequency and the amplitude of the higher components, even though the s p e c t r a l e n v e l o p e may remain the same. In continous speech, a vowel is practically always represented by an acoustical event which, while being still near-periodic, deviates through continuous modification of the spectral envelope and, mostly, through gradual temporal modulations of fundamental frequency. It is still not fully clear how fundamental frequency and spectral envelope interact with relation to phonetic quality, but it is still being fairly generally assumed that only spectral envelope features are relevant. Phase relations between the coponents of any near-periodic signal have long been known to be perceptually largely irrelevant, especially in phonetic signals. These circumstances have significant consequences for automatic speech recognition. If fundamental frequency and phase may be ignored, then the waveform is obviously not the most appropriate representation of the vowel in continuous speech. The spectral representation involves the trade-off resolution problem between frequency and time, which has found very effective solution in the application of broad-band filtering with considerable overlap, as implemented in the Sona-Graph. Although the heterodyning process is continuous in the Sona-Graph analysis, it may be compared to an analysis with approx. 400 passbands and a distance of 19 Hz between the mid-frequencies of successive bands. This kind of analysis has the advantage of sharply separating individual spectral-energy maxima that vary in time. If information were to be collected even with a restricted number of amplitude steps from each individual passband by digital methods, this would constitute a heavy computational load. The Sona-Graph representation of the speech signal in general, and of the time-varying vowels in particular, is however reputed for permitting easy visual tracking of the temporal

variations of each of the spectral peaks.

Energy peaks in the envelope of vocalic spectra have been related to the poles of the acoustical transfer function of the vocal tract (Fant, 1960; Stevens and House, 1961). It has also been shown (a) that there is a correlation between the frequency of the peaks and the spectral level of the peaks in a given vocalic spectrum (see, e.g., Fant, 1956) so that in a description of the spectrum in terms of its peaks, a specification of the spectral levels is largely redundant, (b) that the complete spectral envelope is predictable from the frequencies of the spectral peaks, and (c) that only a small number of the peaks need to be specified in order to relate the envelope to phonetic categories (Fant, 1960).

The spectral peaks, as related to the poles of the transfer function of the vocal tract, or to the normal modes of vibration of the oral (oro-nasal) cavities, are referred to as f o r m a n t s (see, especially, Hanne, 1965).

Several methods of tracking and measuring the formants in speech by automatic methods have been suggested (e.g., Flanagan, 1956; Markel and Gray, 1974) and they may replace visual measurements based on spectrograms. A simple method of extracting the time-varying formant frequencies, with analog and digital registration has been developed by Kubzdela (1973). Because these methods were still in the process of testing at the time when the measurements were made for the present study, the formant frequencies were here measured from conventional spectrograms with an accuracy of 50 Hz at intervals of Δt = 20 ms.

A different method of representing vowels in continuous speech has been used by several researchers in Holland (see, especially, Plomp, Pols and van de Geer, 1967; Pols, 1971). The output of 17 1/3-octave bandpass filters is continuously measured by digital methods and the vowels are thus originally represented as traces in a 17-dimensional feature space, which is then reduced to a plane, i.e., to two dimensions, by using principal component analysis. It was shown that the first two components carried sufficient information and were correlated with the first two formant frequencies (Pols, 1977).

In a recent publication (Papçun, 1980) vowels are also analyzed by means of 1/3-octave filters and specified in terms of coefficients of discriminant functions. The first two of such functions were found

sufficient to specify various steady-state vowels and they, too, are reasonably well correlated with the F_1 and F_2 frequencies - the frequencies of the first two formants.

The methods of describing vowels by using the outputs of $^1/_3$-octave filters have the important advantage of not relying on any particular shape of the spectrum. Consequently, they are available for a continuous description and representation of the speech signal quite independently of the phonetic category of the underlying segments. Indeed, an early segmentation into phonetic elements is basically not necessary for such a method of analysis. But so far no description of extended stretches of speech using this method is available. When - and if - it is, and if segmentation is introduced, there is no way of knowing, and indeed it is doubtful that consonants can satisfactorily be classified on the basis of two statistically relevant features. However that may be, an analysis of the speech signal by means of small-overlap continuous $^1/_3$-octave bandpass filters requires very considerable computing power, just as do the linear prediction methods.

One of the main premises of the work on Automatic Speech Recognition carried on in the Acoustic Phonetics Research Unit is that the method of indentification of any speech elements (whether segments or words or something else) should be highly e c o n o m i c a l , the main consideration being the requirement of low p r o c e s s i n g c o s t . Since this cost tends to increase exponentially with the number of features used for the classification (recognition) of any objects when probabilistic identification methods are used, we decided, before proceeding to a fully automatic recognition of spoken language, to examine in detail the possibilities of feature reduction, using semi-automatic methods in which the measurements are made by man while all the remaining procedures involved in processing the numerical data which lead to classificatory decisions are left to the computer. In our case a small general-purpose machine, the ODRA 1204 was used.

2. Previous work.

The methods used in the present project were first tested on Polish material.

To begin with, near-stationary, isolated vowels were analyzed, described in terms of the four lower formants, and clasified both in design and test sets using various statistical procedures (Jassem,

Krzyśko, Dyczkowski, 1972; Jassem, Krzyśko, Dyczkowski, 1976). It was found that entirely satisfactory results could be obtained in classifying and identifying these vowels on the basis of F_1 and F_2 only, the higher formants carrying barely significant additional information. In subsequent work, only F_1 and F_2 were therefore used.

The six Polish vowel phonemes were next classified and identified in running speech represented by specially constructed sentences with equal frequency of occurrence of each of the phonemes spoken by three selected male voices at three rates of delivery (Łobacz, 1976). Although, as could be expected, the recognition scores decreased with increasing speech rate, the recognition of the vowels was quite satisfactory even in very fast speech (for the three selected voicess the results were 90% correct at fast, though only one of the speakers was a trained phonetician).

Next, on the basis of a large-scale statistical analysis of spoken texts leading to estimates of the occurrence frequency of all phonemes and their sequences up to 4-th order aggregates (Jassem and Łobacz, 1976), 6 typical sentences were constructed which were statistically representative of the Polish language. The vowels in these sentences were measured, classified and identified for a larger number of casual male voices (Jassem, Gembiak, Dyczkowski, 1979) and the results indicated that for normal rate of utterance, using again F_1 and F_2 only, 90% of the vowels could be identified correctly.

Polish has 6 vowels phonemes: /i ɨ e a o u/, which are all monophthongal. The monophthongs and, to a lesser degree, also the diphthongs of American English have been extensively described and classified in terms of the formant frequencies F_1 and F_2 (e.g., Potter and Steiberg, 1950; Peterson and Barney, 1952; Broad and Fertig, 1970, Gerstman, 1968; etc.). Few similar data are available on the monophthongs of General British English (Jones' 'RP', see Jones 1956; Gimson 1962). Arnold et al. (1958) made experiments with synthetic speech trying to find optimum F_1, F_2 and F_3 frequencies for isolated 'RP' monophthongs. Gimson (1962) gives typical formant values based on data collected by J. Wells. No attempt at automatic or semi-automatic classification (recognition) of General British - GBE - vowels has appeared to date.

According to all recent phonological analyses, the monophthongs of GBE can be classified into 12 phonemes (cf., e.g., Jones, 1956; Gimson, 1962; Wells and Colson, 1971; Jassem, 1979), which we will transcribe

/i ɪ e æ ʌ ɑ ɒ ɔ ʊ u ɜ ə/. Of these, the last one is only used in unaccented syllables and will not be considered here. With 11 GBE as against 6 Polish phonemes, assuming similar overall occupation of the feature space, distinctly poorer results might be anticipated for GBE when the same identification methods are used. It was one of the main aims of the present study to see whether in fact the possible deterioration of the results would be so strong as to make the method inapplicable for the GBE vowels.

If both the GBE and Polish syllabic vocoids are described in terms of the 'Cardinal Vowels', then the latter are found to be optimally spread in the articulatory-perceptual vowel space (see Linblom and Sundberg, 1969; Jassem, 1973).

The GBE systems might perhaps be regarded as slightly less economical because the 'close' articulation area, especially the 'back-close' part of it is not used there (the back-close area corresponds to low F_1 with low F_2). On the other hand, a 'half-open-front' [æ]-like vocoid is only used in Polish as a relatively rare allophone of /a/, and no Polish vocoid is as open-back as GBE /ɒ/.

3. Materials.

Three native speakers of GBE were used for the present experiment. They will be denoted by the letters M, C and B. Each of them read, at a fast but natural rate, 50 words in a previously prepared list. The words were as follows: 1. bib /bɪb/, 2. court /kɔt/, 3. zed /zed/, 4. shoot /ʃut/, 5. good /gʊd/, 6. zip /zɪp/, 7. vase /vaz/, 8. set /set/, 9. seethe /sið/, 10. push /pʊʃ/, 11. dove /dʌv/, 12. soot /sʊt/, 13. sieve /sɪv/, 14. farce /fɑs/, 15. curve /kɜv/, 16. teeth /tiθ/, 17. cause /kɔz/, 18. thick /θɪk/, 19. bog /bɒg/, 20. that /ðæt/, 21. daub /dɔb/, 22. shirk /ʃɜk/, 23. gag /gæg/, 24. cough / kɒf/, 25. teethe /tið/, 26. death /deθ/, 27. zag /zæg/, 28. purse /pɜs/, 29. food /fud/, 30. shock /ʃɒk/, 31. tease /tiz/, 32. this /ðɪs/, 33. thud /θʌd/, 34. burp /bɜp/, 35. soothe /suð/, 36. goose /gus/, 37. serve /sɜv/, 38. thorp /θɔp/, 39. dog /dɒg/, 40. pack /pæk/, 41 fez /fez/, 42. path /pɑθ/, 43. tag /tæg/, 44. gush /gʌʃ/, 45. should /ʃʊd/, 46. cash /kæʃ/, 47. carve /kɑv/, 48. tough /tʌf/, 49. buzz /bʌz/, 50. thief /θif/.

The list was so constructed as to fulfil , approximately, the following requirements:

(1) All words are monosyllabic.

(2) All words have the structure CVC.

(3) Each of the 11 vowels /i ɪ e æ ʌ ɑ ɒ ɔ ɷ u ə/ is represented an equal number of times.

(4) The consonants are stops and fricatives, and both types are resented equally.

(5) The consonants are lenis and fortis, and both types are represented equally.

(6) The four places of articulation are: (I) (bi)labial, (II) dental or alveoral, (III) alveolo-palatal, (IV) velar. These are represented equally.

It having been decided that the words should come from the real GBE lexicon though with no consideration of text frequency , a perfect fulfilment of the above conditions was an obvious impossibility.

The number of times that each phoneme was represented was as follows:

Vowels		Consonants	
/i/	5	/p/	7
/ɪ/	5	/t/	11
/e/	4	/k/	10
/æ/	6	/b/	6
/ʌ/	5	/d/	9
/ɑ/	4	/g/	9
/ɒ/	4	/f/	6
/ɔ/	4	/θ/	7
/ɷ/	4	/s/	10
/u/	4	/ʃ/	7
/ɜ/	5	/v/	6
	50	/ð/	4
		/z/	8
		/ʒ/	-
			100

Under condition (4) above, the distribution was: stops 52, fricatives 48. Under condition (5), the distribution was: fortis 58, lenis 42. Under condition (6), the distribution was: (bi)labial 25, dental/alveolar 49, alveolopalatal 7, velar 19. Condition (6) was fulfilled very poorly for reasons of (a)'asymetry' of the GBE consonant system (b) very different lexical frequencies of the consonants, (c) differences depen-

dent on lexical frequencies, (d) the non-occurrence of /ʒ/ in real CVC words with stops and fricatives.

The word list therefore represents a compromise between two general conditions: (1) equal distribution of the types of phonemes and (2) lexicon (and text) dependent frequencies of occurrence. For the purposes of the present study such a compromise, necessary as it was, was not without a certain advantage considering its size: Ideally, for general purposes of automatic speech recognition, the distribution should reflect the text-dependent frequencies of the phonemes and their aggregates whilst making the number of different phonemes so large that the size of the individual classes phonemes was not statistically relevant.

The list of words, spoken with a very brief pause between them, was read in an anechoic chamber and recorded using a high-quality condenser microphone and hi-fi recorder.

4. Analysis and measurements.

The analysis was performed with a Kay Electric Co. Sona-Graph using the wide filter. No scale enlargement was considered necessary. The F_1 and F_2 values were both measured at the same points along the time axis for each vowel beginning from the CV boundary, at steps of 20 ms with an accuracy of 50 Hz (see above, p.5). A bivariate reading at each given point along the time scale can thus be represented by a point in an (F_1, F_2) plane. For each vowel there were between 3 and 18 readings according to the duration of the vocalic segment.

5. The statistical model.

It is possible to divide a feature space such as an (F_1, F_2) plane into classification regions (each region corresponding to one object such as one phoneme) categorically. In this case, each point in the plane is assigned to one of the given number of classes (objects - in our case, phonemes) on the basis of the data obtained in the analysis using the criterion of majority. If, for example, 100 readings of (F_1=450 Hz, F_2=1550 Hz) are taken from the measurements of /ɜ/, 25 from the measurements of /æ/ and 12 from the measurements of /ʌ/, the point (F_1=450 Hz, F_2=1550 Hz) is assigned to /ɜ/, etc. One obvious disadvantage of such a procedure is that the assigment of some of the points depends too much on the particular sample. Categorical division of the (F_1, F_2) plane was used by Forgie and Forgie (1959). The borders between the areas delimited categorically are, in such a case, somewhat

irregular and may have to be smoothed by eye (cf. Plomp, 1978). But the main disadvantage of such a procedure is that the p r e d i c t i v e p o w e r of the results is i n d e t e r m i n a t e. A probabilistic approach to the problem of classification (identification, recognition) leads to results that are p r e d i c t i v e, to the extent that (a) the sample is representative of the population and (b) the statistical model is appropriate to the situation. In any statistical treatment of data, it is necessary to make certain assumptions whose rationale often has a purely deductive or a heuristic basis. One such assumption is the type of distribution of the random variable. In many cases there is good reason to assume that the distribution is normal (Gaussian), i.e., that, for the univariate case, the density function is

$$f_i(x) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left[-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right] \qquad (1)$$

where x is the value of the variable X, μ_i is the mean of variable X in the population π_i and σ_i^2 is the variance of that variable, whilst for a multivariate case, the density function is

$$f_i(\underline{x}) = (2\pi)^{-\frac{p}{2}} \left|\underline{\Sigma}_i\right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu}_i)'\underline{\Sigma}_i^{-1}(\underline{x}-\underline{\mu}_i)\right] \qquad (2)$$

where $\underline{x} = [x_1, x_2, \ldots x_p]'$ is the value of the p-dimensional variable $\underline{X}$, which, in the population π_i has a spread defined by the covariance matrix $\underline{\Sigma}_i$, and p is the number of variables.

In the simplest case of two univariate populations π_i and π_j, with equal variances $\sigma_i^2 = \sigma_j^2 = \sigma^2$, the probability of misassignment is a function of the difference $|\mu_i - \mu_j|$, i.e., the difference between the means, and σ^2. This probability increases with the variance σ^2 and with the inverse of the difference $|\mu_i - \mu_j|$ as shown in Fig.1. The projection on the x axis of the intersection between the density functions $f_i(x)$ and $f_j(x)$ is the d i s c r i m i n a n t p o i n t, which lies at $a = |\mu_i - \mu_j|/2$. If the variances are $\sigma_i^2 \neq \sigma_j^2$, then the position of the point a is shifted towards the mean of the population with the smaller variance, as shown in Fig. 2. A given value x_0 of the variable X is assigned, in the process of identification (classification, recognition), to the population π_i or to the population π_2 according

to its relation to point a on the x axis, i.e.,

$$\begin{cases} x_o < a \Rightarrow x \in \pi_i \\ x_o > a \Rightarrow x \in \pi_j \end{cases} \qquad (3)$$

We can calculate the probabilities

$$P(i|i) = \int_{-\infty}^{a} f_i(x)dx \qquad (a)$$

$$P(j|j) = \int_{a}^{+\infty} f_j(x)dx \qquad (b)$$

$$P(j|i) = \int_{a}^{+\infty} f_i(x)dx \qquad (c) \qquad (4)$$

$$P(i|j) = \int_{-\infty}^{a} f_j(x)dx \,. \qquad (d)$$

$P(i|i)$ and $P(j|j)$ are the respective probabilities of correctly assigning a value x_o to the appropriate class whilst $P(i|j)$ is the probability of assigning to population π_i a value x_o assumed or known to belong to population π_j and $P(j|i)$ is the probability of assigning to population π_j a value x_o assumed or known to belong to population π_i. These probabilities are predictive, i.e., they make it possible to estimate an error of misassignment in a similar future experiment.

The probability of correct assignment is related to the s t a t i s t i c a l d i s t a n c e between the two populations π_i and π_j to which a given value of the variable X may belong. In the simplest case of univariate populations with equal variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$, this distance is

$$\Delta = \frac{|\mu_i - \mu_j|}{\sigma} \qquad (5)$$

and if $\sigma_1^2 \neq \sigma_2^2$, then

$$\Delta = \frac{|\mu_i - \mu_j|}{\sqrt{\sigma_1 \sigma_2}} \,. \qquad (6)$$

The square of (5) is

$$\Delta^2 = \frac{(\mu_i - \mu_j)^2}{\sigma^2} \qquad (7)$$

which may be written

$$\Delta^2 = (\mu_i - \mu_j)\frac{1}{\sigma^2}(\mu_i - \mu_j) \cdot \qquad (8)$$

The statistical distance (or discriminance) between two multivariate populations is expressed, in the case of equal covariance matrices $\underline{\Sigma}_i = \underline{\Sigma}_j = \underline{\Sigma}$ by

$$\Delta^2 = (\underline{\mu}_i - \underline{\mu}_j)' \underline{\Sigma}^{-1} (\underline{\mu}_i - \underline{\mu}_j) \cdot \qquad (9)$$

Note the similarity between (9) and (8).

If $\underline{\Sigma}_i \neq \underline{\Sigma}_j$, the squared distance between the populations π_i and π_j is given by

$$\Delta^2 = (\underline{\mu}_i - \underline{\mu}_j)' \left[y\underline{\Sigma}_i + (1-y)\underline{\Sigma}_j \right]^{-1} (\underline{\mu}_i - \underline{\mu}_j) \qquad (10)$$

where y is the unique solution of

$$\underline{b}' \left[y^2 \underline{\Sigma}_i - (1-y)^2 \underline{\Sigma}_j \right] \underline{b} = 0 \qquad (11)$$

As in the univariate case, probabilities of correct and incorrect assignment (classification) may be calculated for the multivariate case by using the appropriate integrals, but the computations then become more laborious.

If the variable $\underline{X}$ is two-dimensional, its feature space is a plane, and the point a dividing two populations π_i and π_j is replaced by a line whose shape depends on the type of discriminant function. In the present instance, quadratic discriminant functions are used so that the dividing curve is a segment of an ellipse circle , parabola or hyperbola . For 3 variables, the 3-dimensional space is divided by a surface, and for 4 or more variables the regions of the feature space are separated by a hypersurface. The hypersurface is described by

$$\underline{x}' \left(\underline{\Sigma}_j^{-1} - \underline{\Sigma}_i^{-1} \right) \underline{x} + 2\left(\underline{\mu}_i' \underline{\Sigma}_i^{-1} - \underline{\mu}_j' \underline{\Sigma}_j^{-1} \right) \underline{x} +$$
$$+ \underline{\mu}_j' \underline{\Sigma}_j^{-1} \underline{\mu}_j - \underline{\mu}_i' \underline{\Sigma}_i^{-1} \underline{\mu}_i + \ln |\underline{\Sigma}_j| / |\underline{\Sigma}_i| + 2 \ln q_i / q_j = 0 \qquad (12)$$

where $\underline{\mu}_i$, $\underline{\mu}_j$ are the mean vectors, $\underline{\Sigma}_i$ and $\underline{\Sigma}_j$ are the covariance matrices of the populations π_i and π_j, and q_i and q_j are the respective apriori probabilities of occurrence of the objects representing the respective populations.

Expression (12) relates to the more general case of $\underline{\Sigma}_i \neq \underline{\Sigma}_j$.

If the number of different populations (classes) is n, then it is nessessary to calculate $n(n-1)/2$ discriminant functions of the type (12). The procedure may be simplified, as described in Jassem, Krzyśko, Dyczkowski (1976).

6. The variables and their transformations.

The original data, as stated above (p.10), consist of direct measurement values of F_1 and F_2, but since perception tends to discriminate along a logarithmic rather than a linear scale, there is justification for transforming the formant frequencies accordingly. Various authors have used a log frequency scale for the description of vowels in terms of F_1 and F_2, and some use transformed data (see, recently, especially Broad and Wakita, 1978). Ladefoged (1975) has suggested that the traditional decription of vowels in terms of the IPA quadrilateral is really based on perception rather than articulation, that the positioning of the vowels in the plane of the quadrilateral is more strongly correlated with the variables F_1 and $(F_2 - F_1)$ than with F_1 and F_2, and furthermore, that formant data should be transformed according to the mel scale. Ladefoged's description is heuristic, but the results are very encouraging. For the purposes of the present study, straightforward frequences F_1 and F_2-F_1 (untransformed to the mel scale) were used, but some of the results have been brought onto Ladefoged's vowel chart.

The classification of our vowels has been performed three times, separately for each of the three sets of variables: F_1 and F_2, $\log F_1$ and $\log F_2$ as well as F_1 and F_2-F_1.

7. The mean formant frequencies.

Table 1 includes, for each of the 11 accentable GBE monophthongs, the formant frequencies F_1 and F_2 according to Arnold et al. (1958), according to Wells (in Gimson, 1964) and the corresponding values for each of our 3 speakers. Each formant frequency for M, C and B is the mean of all the measurement data for the given vowel and the given speaker in our materials. Additionally, mean values are given for the

measurement data averaged over the three speakers.

There is broad overall agreement in that the F_1 values increase with the degree of opening whilst the F_2 values increase with the degree of fronting, in all sources. Some of the most essential divergencies of the Arnold et al. data from those of Wells as well as from ours are as follows:

1 /ʌ/ has the highest F_1 value.
2 /u/ has very low F_1.
3 /i/ and /ɪ/ have the same F_2 values.
4 /æ/ and /ʌ/ have the same F_2 values.
5 F_2 of /ɜ/ is high enough to reflect an almost fully fronted vowel of the [ɛ] type.

In terms of the absolute values of F_1 and F_2, there are otherwise no drastic differences between the Arnold at al. data and the other figures in the Table.

The Wells figures differ from ours in a stronger differentiation of the front vowels /i ɪ e æ / from each other and from the rest of the vowels.

Table 2 contains the mean logF frequencies for our 3 speakers. The means are calculated after a linear-to-log transformation of the original data, which is equivalent to calculating geometric means of the original data. The Table also contains these means expressed in Herz i.e., the geometric means.

Figs 3, 4 and 5 show our GBE vowels in a ($\log F_1$, $\log F_2$) plane according to the mean values included in Table 2, and Fig. 6 is an adaptation of the data for 'RP' as shown, typically, e.g., in Gimson (1964), Wells and Colson 1971 and Jassem (1979), in terms of the IPA quadrilateral which, in order to compare directly with a ($\log F_1$, $\log F_2$) chart, has to be reversed to a mirror image and turned anticlockwise by 90°.

Comparing Figs. 3,4,5 and 6, a very good agreement may be found, but assuming the usual simple F_1/F_2 : high-low/front-back relation, the following divergencies of our M vowels from the textbook IPA-quadrilateral description should be noted:

(1) /i ɪ e/ lie on a straight line indicating that either all three are front (i.e. that there is no centralization of /ɪ/) or that /e/ is also somewhat centralized.

(2) The distance between /e/ and /æ/ is quite small and comparable only to the distance between /ɑ/ and /ɒ/.

(3) /ʌ/ and /ɑ/ have the same degree of opening, so that the relevant difference between them is only that of front-retracted vs. back-advanced.

(4) /ɑ/ is just as back as /ɒ/, but /ɔ/ is considerably more retracted than both.

(5) /u/ is rather more fronted.

Our speaker C's vowels differ from those indicated in Fig. 6 essentially in the following respects:

(1) /i ɪ e æ/ lie very nearly on a straight line (cf. the comment above).

(2) The distance between /ʌ/ and /ɑ/ is the shortest of all the distances between any two vowels but the /ʌ/ differs from /ɑ/ in hight as well as in the front-back feature .

(3) /æ/ has the same degree of opening as /ɑ/.

(4) /ɑ/, /ɒ/ and /ɔ/ lie on a straight line from low-back-advanced to mid-fully-back.

The main differences between our speaker B's vowels and the 'textbook' ones may be summarized as follows:

(1) Whilst /i ɪ e/ very nearly lie on a straight line, /æ/ is more fronted, thus indicating that both /ɪ/ and /e/ are slightly centralized.

(2) /ɑ ɒ ɔ/ are on a straight line from open-back-fronted to mid-back.

(3) /u/ is much more fronted than /ɑ/, with the /u/-/ɪ/ distance strinkingly small.

Note, also, that for all speakers, the distance between /e/ and /ɜ/ is very much smaller than between /ɜ/ and /ɔ/. This is probably a reflectation of the fact that the IPA has two quadrilaterals - one for 'rounded' and one for 'unrounded' vowels. It is obvious that if some kind of /ɤ/ were substituted for the regular GBE /ɔ/ = [ɔ̟], then the distances would be more equal (cf. Jassem, 1973). The vowels are here brought into a two-dimensional space, whilst the pseudo-articulatory space is three-dimensional (and so is probably the perceptual vowel space).

Table 3 contains our means of F_1 and F_2-F_1.

Ladefoged suggests a reversal of the scales for both F_1 (which he makes descrease upwards) and (F_2-F_1) (which he makes descrease to the right). For the variables F_1 and F_2 this trick was used by Joos (1948)

and has been followed by many specialists since (e.g., Peterson and Barney, 1952). The result is an identity of orientation with the traditional IPA quadrilateral according to which close vowels are placed above open vowels and front vowels to the left of back vowels.

A surprising effect of Figs. 7, 8, 9 and 10 is that the replacement of F_2 by F_2-F_1 has brought about very little topological difference: The relations between and among the vowels are very much the same as they are in Figs. 4, 5 and 6. Also the deviations described above on p.15-16 remain essentially the same, though now /ɔ/ is more in line with the position on the IPA quadrilateral. We have decided to present a 'combined' vowel chart with the collapsed means because this may be regarded as a more 'neutral' representation, with interspeaker differences removed. But it is not certain that the procedure is necessarily justified.

8. The dispersion of formant frequency values.

The variance σ^2 of the univariate statistical analysis is replaced by the covariance matrix Σ if objects are described by two or more features. This matrix shows, in addition to the variances of the individual variables, the r e l a t i o n s b e t w e e n the variables, in the form of covariances. The covariance is simply related to the variances between the two, or any two of a larger number of variables. If the objects under consideration are described with p quantitative features, then there are $p(p-1)$ covariances of the type σ_{ij} and σ_{ji} with $i \neq j$. Since, however, $\sigma_{ij} = \sigma_{ji}$, there are only $\frac{p}{2}(p-1)$ (possibly) different values of covariance. As in our experiment, there are two variables (two features): F_1 and F_2 or $\log F_1$ and $\log F_2$ or F_1 and F_2-F_1, we obtain $\frac{2}{2}(2-1) = 1$ covariance for each object (each phoneme) under each condition. We symbolize this covariance by σ_{12}, which is the population covariance. The actual value of the population covariance is not known and we estimate it from our sample. The estimated covariance is symbolized s_{12}, whilst the estimates of the variances $\sigma_1^2 = \sigma_{11}$ and $\sigma_2^2 = \sigma_{22}$ are s_1^2 and s_2^2, also denoted as s_{11} and s_{22}.

Thus, the dispersion of the formant frequencies (or the variables derived from them) for each phoneme is described by the covariance matrix

$$\underline{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}, \qquad (13)$$

estimated by

$$\underline{S} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}. \qquad (14)$$

Since $s_{12} = s_{21}$, the sample dispersion of the variables is fully described, in our case, by three values, viz. $s_{11} = s_1^2$, $s_{12} = s_{21}$, and $s_{22} = s_2^2$.

The covariance is related to the variances and the c o r r e l a t i o n between the two variables by

$$s_{12} = r_{12} s_1 s_2 , \qquad (15)$$

where r_{12} is the estimate of the classical Pearson correlation coefficient between the variables, and s_1 and s_2 are the square roots of the respective variances, i.e., the standard deviations. The Pearson ϱ, whose estimate is r, is a standarized measure varying between +1.0 and -1.0, with $\varrho = 0$ indicating complete indenpendence of the two variables. The values +1.0 and -1.0 indicate complete dependence, i.e. a functional relationship between the variables. With a positive correlation coefficient an inccease in one variable is accompanied by an increase in the other variable. A negative correlation coefficient indicates that one variable increases with the decrease of the other.

Since the discriminant functions (cf. above, p.15) are based on the covariance matrices, they depend on all the elements of the matrices, the variances as well as the covariances. Since, further, the classificatory decisions are based on the discriminant functions, it follows that a classificatory decision depends on the values of all the variances and covariances which are the elements of the actual covariance matrices.

Tables 4-6 contain the values of the elements of the covariance matrices in all the three systems of variables used in the present study. The a b s o l u t e values of the variances and covariances are not easily interpretable, but some tendences in their r e l a t i v e values may be of interest. When the $(F_1, F_2 - F_1)$ design is

compared with the (F_1,F_2) design, two tendences may be noted: (a) a generally greater variance of $(F_2 - F_1)$ than that of F_2 and (b) almost invariably negative correlation between $(F_2 - F_1)$ and F_1 as against often positive correlation between F_2 and F_1.

Note that a relatively low value of covariance need not necessarily indicate relative independence of the two variables since it depends on the values of the variances. Thus, for instance, s_{12} of C's /a/ in the (F_1,F_2) design is relatively low, viz. -358. As $s_{12} = r_{12}s_1s_2$, r_{12} can be calculated as $r_{12} = \frac{s_{12}}{s_1s_2}$, which, in this case, is

$r_{12} = \frac{-358}{\sqrt{3168 \cdot 2155}} = -0.137$. For /ɔ/ in the same design $s_{12} = -951$

and $r_{12} = \frac{-951}{\sqrt{3818 \cdot 19540}} = -0.110$.

A detailed analysis of the correlations has not been carried out, but the several values that have been calculated seem to indicate that the correlations between the variables are weak or very weak.

It can be seen from Tables 4-6 that the variance values for F_2 in the (F_1F_2) design and of (F_2-F_1) in the (F_1,F_2-F_1) design are mostly greater (often by an order of magnitude) than those of the corresponding F_1 in the respective vowels, whilst in the $(\log F_1, \log F_2)$ design, though mostly smaller, they tend to be less drastically different. A logF model of vowel description may be therefore be preferable. It would also be interesting to compare the variances of F_1 and F_2-F_1 after the raw data have been transformed to the mel scale, as suggested by Ladefoged (1975).

An important application of the variance in univariate analysis and the covariance matrix in multivariate analysis is the e s t i m a t i o n of the p o p u l a t i o n m e a n (mean vector). In most situations in which statistical inference is employed, we are interested the true mean (mean vector) of a population though we only actually know the mean (mean vector) of the s a m p l e. If we can be reasonably sure that the sample is really representative of the population, we can estimate, at a given level of confidence, the interval, of region, which includes the true value the population

value (of the mean vecor) . In the univariate case, the confidence interval is a distance along the $\bar{x}$ axis between two values on either side of the sample mean. In the multivariate case, the confidence region is an ellipse (for two variables) or an ellipsoid (for more than two variables) with the centroid at the sample mean vector. If the variables are completely uncorrelated, the ellipse goes to a circle and the ellipsoid to a (hyper) sphere. If they are completely correlated, the confidence regions become reduced in dimensionality (e.g., an ellipse goes into a line) . Whilst the shape and the orientation of the confidence ellipses (ellipsoids) depend on the variances and covariances, their a r e a s also depend on the s a m p l e s i z e. The greater the sample size, the smaller the area of the ellipse (ellipsoid), i.e., the more confident we may be that the sample mean vector is a good representation of the population mean vector.

If the distribution of the variable(s) is normal, then, according to the sampling theorem, the means, or mean vectors, of independent samples of a given population are also normally distributed. In the univariate case, the sample means of a given population are distributed according to

$$\phi\left(\bar{x}\right) = \frac{\sqrt{N}}{2\pi\sigma} \exp\left[-\frac{N\left(\bar{x}-\mu\right)^2}{2\sigma^2}\right]. \qquad 16$$

In the multivariate case there is an analogous distribution of mean vectors:

$$\phi\left(\underline{\bar{x}}\right) = 2\pi^{-\frac{p}{2}} \cdot N^{\frac{p}{2}} \cdot \left|\underline{\Sigma}\right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} N \left(\underline{\bar{x}}-\underline{\mu}\right)' \underline{\Sigma}^{-1} \left(\underline{\bar{x}}-\underline{\mu}\right)\right]. \qquad (17)$$

In the expressions 16 and 17 , $\bar{x}$ and $\underline{\bar{x}}$ are the sample mean and the sample mean vector, and μ and $\underline{\mu}$ are the population mean and the population mean vector.

The distributions $\phi\left(\bar{x}\right)$ and $\phi\left(\underline{\bar{x}}\right)$ make it possible to obtain confidence intervals and confidence ellipses (ellipsoids). In the univariate case, we can find the interval on the $\bar{x}$ axis which, with a probability of $(1-\alpha)$, includes the population mean by taking

$$CI = \bar{x} + \left(1-\alpha\right)\frac{s}{\sqrt{N}}, \qquad (18)$$

where $\bar{x}$ is the sample mean, α is the significance level , s is the sample standard deviation and N is a (sufficiently large) sample size.

The confidence region in the bivariate case, in the form of an ellipse is given by

$$ax_1^2 + 2bx_1x_2 + cx_2^2 + 2dx_1 + 2ex_2 + f = 0 \qquad (19)$$

where

$$a = s_{11}\,, \qquad b = s_{12}\,, \qquad c = s_{22}\,,$$

$$d = -\left(s_{11}\bar{x} + s_{12}\bar{x}_2\right),$$

$$e = -\left(s_{12}\bar{x}_1 + s_{22}\bar{x}_2\right),$$

$$f = s_{11}\left(\bar{x}_1\right)^2 + 2s_{12}\bar{x}_1\bar{x}_2 + s_{22}\left(\bar{x}_2\right)^2 - \frac{T_o^2(\alpha)}{N}$$

In the above expression for f, $T_o^2(\alpha)$ is the value of the Hotelling T^2 criterion, the multivariate analogue of the univariate criterion t:

$$T_o^2(\alpha) = \frac{(N-1)\,p}{N-p}\, F_{\alpha\, p,N-p}(\alpha) \qquad (20)$$

where p is the number of variables. In our case, p = 2. α is the significance level, and F is a criterion (with values tabulated for various significance levels) used in various statistical tests (such as the test of equality of variances).

In (19) the notations s_{11}, s_{12}, s_{21} and s_{22} have special meanings. Rather than the usual elements of the covariance matrix, they are here the elements of the i n v e r s e covariance matrix $\underline{S}^{-1}$, i.e.,

$$\underline{S}^{-1} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}. \qquad (21)$$

Fig. 11 shows confidence ellipses for the mean formant frequency vectors for speaker M in the design (F_1,F_2), i.e., for the vectors $(\bar{F}_1,\bar{F}_2)$ of each of the 11 vowel phonemes.

It is also possible to draw ellipses of equal probability density of the vectors (F_1,F_2) for each of the phenemes. They have exactly the same shape, orientation and centroids as the confidence ellipses.

By drawing families of such equal probability density ellipses it is possible to see at what stage the probability density figures corresponding to those in Fig.1 cross. Just as, in the univariate case, a projection of the crossover of two probability density curves represents a discriminant point (see p.11), so, in the multivariate case, the projections of lines, or surfaces, along which the probability density regions cross, indicate the divisions between the identification regions described by the discriminant functions. The confidence ellipses are, consequently, a good indication of the probabilities of missassignment that are analogous to the probabilities (c) and (d) in (4).

9. Classification of the vowel trajectories in the feature spaces.

The discriminant functions were used to divide the feature spaces into identification regions, one for each of the 11 phonemes under investigation. Since, in each of our three designs, there were always just two variables, our spaces were reduced to planes and these were divided into identification regions by quadratic curves (see above, p.13). There were three different feature spaces: (F_1, F_2), $(\log F_1, \log F_2)$, and $(F_1, F_2 - F_1)$. Each of these three spaces was divided into identification regions separately for each speaker and, additionally, for the three speakers combined. In each of the different identification planes, points were selected which were spaced 50 Hz apart along each of the two orthogonal co-ordinates. Each such point corresponded to a possible bivariate measurement and was identified as representing exactly one phoneme according to the identification region in which it was situated. In this way, 12 vowel charts were composed. Computer printouts of some of these charts appear in Figs.12 to 17. For purposes of computer processing the phonemes here investigated are coded as follows: /i/ 1, /ɪ/ 2, /e/ 3, /æ/ 4, /ʌ/ 5, /a/ 6, /ɒ/ 7, /ɔ/ 8, /ɑ/ 9, /u/ 10, /ə/ 11. The individual phoneme areas are separated by stepwise lines corresponding to the dividing curves as determined by the discriminant functions. The classification of the vowel trajectories in the identification planes will be illustrated by an example. The formant frequencies F_1 and F_2, as measured every 20 ms in the word 'dove' spoken by M are as follows: (500, 1750), (600, 1600), (700, 1450), (650, 1350), (650, 1350), (600, 1350); (600, 1350), (550, 1300). By referring to the chart in Fig. 12 it may be found that these points are identified as representing successively the following phonemes: /e ʌ ʌ

ʌʌɒɒɒ /(3 5 5 5 5 7 7 7). Each of the C measurements is thus mapped onto the chart and represents a phonemic 'microsegment'. In this case, the longest succession of phonemically identical microsegments is /ʌʌʌʌʌ/, consequently the entire succession of microsegments is identified as representing the phoneme /ʌ/. Occasionally (especially in tokens of /i/ and /ɔ/), all microsegments represented the same phoneme: For instance, M's vowel in the word 'teethe' was microsegmenally /i i i i i i i i i i i/ and in a case like this the final decision is obvious. But there are also cases of subsequences of equal length. For instance, the vowel in M's 'bog' was represented by /ɜ ɔ ɒ ɑ ɑ ɑ ɒ ɒ ɒ /. The final decision here was /ɒ/ not because there was altogether one /ɒ/ more in the entire sequence, but because the subsequence /ɒ ɒ ɒ/ came a f t e r /ɑ ɑ ɑ/. The algorithm included the assumption that of two (or more) subsequences of equal length it was the l a s t one that lead to the final decision. This reflects the assumption that initial consonants in CVC sequences affect the formant 'movements' more strongly than do final consonants, and this assumption was borne out by the results of the identification. In this way, each of the 3(speakers) x 50(words) = 150 tokens was identified in each of the three designs: (F_1, F_2), $(\log F_1, \log F_2)$ and (F_1, F_2-F_1). The 12 vowel charts were stored in the computer's memory and the procedure of trajectory identification was performed by a separate program.

The three charts for the three designs and a given speaker do not differ substantially, as can be seen from Figs.12, 13 and 17 which pertain to speaker M.

As the range of F_2 normally used in speech is almost three times the range of F_1, the overall shape of the vowel space as delimited by the potential (F_1, F_2) values is disproportionately extended along the F_2 axis. This cannot be helped even after the F-to-logF transformation because the original measurements were made along a linear scale of formant frequency. Since, however, the positions of the vectors of the individual phonemes in a $(\log F_1, \log F_2)$ plane reflect the positions of the vowels on the IPA quadrilateral more directly than do those in the (F_1, F_2) plane, vowel charts were computer-composed with $\log F_1$ aud $\log F_2$ as the linear co-ordinates and equal-logF distances between the points in the $(\log F_1, \log F_2)$ plane. These appear in Figs 13-16. They were not used for the purposes of vowel classification in the present experiment as the points in this plane do not correspond to the measurement

data, but they are presented here because they are apparently a good representation of the areas taken by the individual phonemes in a reduced bidimensional perceptual and articulatory space (cf. Fig. 18).

10. Results of classification.

When the procedure described in the preceding section was applied to all the vowel tokens in our materials, in all three designs, identification results were obtained which are presented in Tables 7, 8 and 9. These results may be summarized as follows:

Percent overall scores

design → / ↓ speaker	F_1F_2	$\log F_1 \log F_2$	F_1 F_2-F_1
M	74	76	74
C	76	74	76
B	74	70	74
comb.	64	65	64

It may be seen from the above figures that there is very little, if any, overall effect of the transformations of the original formant data. The somewhat lower overall scores for B in the $\log F_1$, $\log F_2$ design cannot be taken as indicative of a distinct difference. Also, the differences in the o v e r a l l scores between the speakers are very small and may be largely due to chance. For the i n d i v i d u a l vowels, the percent scores based on data from the individual speakers are:

vowel → / ↓ design	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	o	u	ɜ
F_1, F_2	100	80	75	83	14	100	42	100	92	67	60
$\log F_1$, $\log F_2$	100	80	67	89	27	100	42	100	75	67	60
F_1, F_2-F_1	100	80	75	83	27	100	42	100	92	67	60

There is a good deal of consistency in the above figures. There is again no, or very little, effect of the design. Vowels /i/, /ɑ/ and /ɔ/ are always identified correctly, /ɪ/, /æ/ and /o/ are also quite well identified and /ʌ/, irrespective of the design, fares worst.

The confusion matrices in Tables 7,8 and 9 reflect the similarities

between the vowels as evidenced in Figs. 3,4,5,7,8 and 9. Note that there is no confusion between the pair /i/ - /ɪ/. On the other hand, /ʌ/ is mostly identified as /ɑ/ showing the acoustical closeness between these sounds.

It is not possible to give a simple answer to the question whether an overall score of about 75% correct identification is acceptable or not. The answer depends on the context within which the question is being asked. Consider, first, that neither a naive native speaker, nor the specialist, always identifies the vowel correctly using his biological perceptual mechanism, although he receives considerable more information at the acoustical level, viz. the entire time varying spectrum (cf., e.g., Fairbanks and Gruubs, 1961). Second, a semiautomatic identification of vowels as presented here, is made in order to prepare the methodology of fully automatic recognition, which would have to be related to a specific project or at least to a specific aim. If, for instance, the purpose of a specific system of ASR is to recognize the meanings of complete utterances from a limited set of sentences with certain semantic, lexical and syntactic constraints, then the redundancy of the information at the phonological level is so high that even with modest introduction of 'top-down' procedures, complete success may be possible if only a small proportion of the vowels are correctly identified. If a given system is expected to recognize isolated words from a sizable vocabulary containing a number of minimally contrastive pairs, such as 'hut' - heart', 'cod' - 'cord', then the scores would have to be better than those here obtained. But it is difficult to envisage any practical applications of such a system. Minimal contrastive pairs differing by just the vowel, even if numerous in English, are almost invariably related to semantic distinctions which would make it rather implausible for such pairs to belong to a somewhat limited vocabulary selected for some practical purpose. One also has to consider the correct-identification scores of the consonants, if the ultimate aim of the system is the recognition of complete lexical entities. High rates of phoneme- by-phoneme recognition of the speech signal would be needed for purposes of a phonetic typewriter, but now, after 30 years of work in the area of ASR, it becomes clear that such devices, depending almost entirely on 'bottom-up' analysis, are - at least for the near future - not practical propositions.

Our method, being based on information on the time-varying frequencies of F_1 and F_2, entirely ignores d u r a t i o n. In the traditional description, the GBE monophthongs are referred to as being either 'long' or 'short' at the phonemic level (Jones, 1956; Gimson, 1964; O'Connor, 1967) and the 'long-short' relation was particularly emphasized by Jones (1950, esp. Chap. XXIII). Although it has always been admitted that no two stressable GBE vowel phonemes are the same in quality, many authors have considered the length differences to be both theoretically relevant and important in language teaching. Indeed, for some 50 years an IPA transcription of the GBE monophthongs prevailed in which the distinction between bead' and bid' was shown as /i:/ vs. /i/, between 'cord' and 'cod' as /kɔ:d/ vs. /kɔd/, and between 'shoed' and 'should' as /ʃu:d/ vs. /ʃud/. At one time, also the difference between 'card' and 'cad' was merely indicated by the length mark or an equivalent doubling of the phonetic letter (e.g., MacCarthy, 1947). The vowels of General American English have been almost unanimously transcribed by different letters, often without a length mark, in IPA. It has never been shown experimentally that there is a significant difference between the two variaties of English with respect to the quality-quantity relation. Thirty years ago one of the present authors (Jassem 1950) suggested a transcription of GBE (which he then called 'Educated Southern British') which obviated the colon. Since then, at least two influencial authors have abandoned it: Abercrombie (1964) and Wells (Wells and Colson, 1971). It can be seen from the confusion matrices (Tables 7, 8 and 9) that /ɪ/ is never identified as /i/ and that /ɒ/, if misassigned, is much more frequently identified as /ɑ/ than as /ɔ/ ; /æ/ and /ɑ/ are never confused.

One way of looking at 'quasi-phonemic' length in GBE is to consider the following system (Jassem, in print):

1. /ə/
2. /ɪ/ /ʊ/
3. /e/ /ʌ/ /ɒ/ /æ/
4. /i/ /ɑ/ /ɔ/ /ɜ/ /u/.

(1) is unstressable and may occur finally.

(2) are stressable and may occur in unstressed final position.

(3) never occur in final position.

(4) may occur in final position, stressed or unstressed.

/ɒ/ is at present, with some GBE speakers, in transit from group (3) to group (4) in connection with the monopthongization of /ɛə/.

The duration of a vowel is the result of several interworking factors of which the following are probably the most significant: (A) membership in one of the groups (1) to (4) above increasing in steps from (1) to (4), (B) position (final vs. non-final) in rhythm unit ('phonological word') and (C) the following context (Ø, neutral consonant, lenis co., fortis co.). Although some studies of the effect of these factors on GBE vowels do exist, sufficiently reliable statistical data are not yet available, and interactions among these factors are not sufficiently well known (cf. Imiołczyk, forthcoming). Some of the confusions in our materials are such that a vowel belonging to group (3) (or (2) in the case of /ɑ/) is identified as the nearest vowel belonging to group (4). It is possible (though by no means certain) that when duration is treated as an additional feature with proper weight, the results of (semi)automatic identification may be improved.

The figures in Table 10 indicate the number of times that the vowel in the particular word was assigned to the correct phoneme, i.e., the absolute 'recognition scores' for the words, regardless of the speaker, but split up according to the 3 different designs. For convenience, the words are here arranged according to the vowel. In 42 out of the 50 words, the vowels were correctly recognized the same number of times in all three designs. In the remaining 8 words, one or the other of the three designs had a different score. 4 of these 8 words were those containing /ɜ/, and 2 containing /u/. There may therefore be some very slight interaction between design and vowel, with (F_1, F_2-F_1) doing better work for /ɜ/. A similar arrangement of the scores, according to speaker (rather than the design) shows that there is very little, if any, interaction between word and speaker. A statistical treatment of such interaction would require larger materials.

Of the three phonemes which got 100% scores, /i/ mostly occurs in the environment of a dental or alveolar consonant which might make the results look too optimistic, but in the /ɑ/ and /ɔ/ words the consonants are variously (labio)dental, alveolar, or velar, so that probably the contextual homogeneity did not have an anhancing effect on the /i/ scores. Although /ʌ/ is a decided loser, the results are quite consis-

tent, as may be seen from Table 10. Whilst the present results do not seem to point to any contextual effects in the sense of place of articulation, there is probably an influence of the final consonant in the way it is related to the duration of the vowel. The shorter allophones tend to be misassigned more frequently than the longer allophones (compare 'bib' and 'sieve' with 'zip', 'thick' and 'this', 'bog' and 'dog' with 'cough' and 'shock'.

The interplay of linguistic and personal acoustical features of speech sounds, especially vowels, has recently attracted a good deal of attention. Several suggestions have been made to extract the linguistic invariants by normalizing for the personal characteristics. The problems of the interaction between linguistic and personal effects and the normalization procedures have recently been discussed (beside Broad and Wakita, 1978, mentioned earlier, p.14), by Disner (1978), Nearey (1978) and Papçun (1980). Again, our materials are not sufficient to enable us to state with any statistically defensible confidence whether our three voices were distinct enough with respect to their formant frequencies to exclude the possibility of a combined treatment. A few general observations may, however, be in order.

If the mean formant frequencies and their dispersions were not significantly different among our 3 speakers, then the statistics obtained for the individual vowels by collapsing the data for the three speakers, when used in the discriminant functions, would lead to recognition scores which were about the same as those for the individual speakers. The scores would be similar for the individual speakers and the individual vowels, and what little differences did persist would be attributable to chance. It is symptomatic that although there was hardly any difference among the speakers in the overall scores, these dropped by about 10 % for the 'combined' case (see p. 24). Note also that the variances for the 'combined' case are larger than for the individual cases (see Tables 4, 5 and 6). It is also very doubtful that the realizations of the individual phonemes by the three speakers are p h o n e t i c a l l y e q u i v a l e n t. If they were, the mutual relations on the vowel charts would be similar and either some reasonably simple transformation of the data or the statistical parameters, or of the scales of the co-ordinates would place the three mean vectors for each vowel very

near each other. A casual study of the topology of the relations among the mean vectors as shown in Figs. 3, 4, 5, 7, 8 and 9 will lead to a strong suspicion that the realizations of at least some of the phonemes by each of our three speakers are not phonetically equivalent.

11. Conclusions.

Using data from spectrographic measurements of the time-varying formant frequencies and a statistical model of classification based on discriminant functions, it was possible to identify semi-automatically the vowels in 50 monosyllabic words uttered at a fast rate by 3 native speakers of General British English with an overall score of about 75 % correct when each voice was treated separately. This score, which may be sufficient for some purposes of automatic speech recognition because of the natural redundancy at the phonemic level, goes down by about 10 % when data are collapsed for the three speakers indicating that the phonetic realizations of at least some of the phonemes by our speakers were not phonetically equivalent. All the same, when the mean vectors representing the individual realizations the 'idiophonemes' are brought onto charts with $\log F_1$ and $\log F_2$ or F_1 and F_2-F_1 as co-ordinates, a large measure of agreement with the textbook description of typical GBE representations of the monophthongs in the IPA quadrilateral can be observed.

Table 1. Formant frequencies F_1 and F_2 of the General British English monophtongs.

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
Arnold et al	315 2200	380 2200	570 2100	940 1750	960 1750	720 1150	720 980	620 920	500 1150	250 800	550 1900
Wells in Gimson	280 2620	360 2220	600 2060	800 1760	760 1320	740 1180	560 920	480 760	380 940	320 920	560 1480
M	351 2286	411 2007	522 1837	600 1778	651 1397	655 1195	600 1242	481 967	348 1258	317 989	533 1518
C	300 2156	470 1714	576 1657	664 1530	643 1321	667 1204	610 1110	493 902	441 1281	328 1234	573 1359
B	325 2105	393 1712	476 1586	620 1628	553 1292	641 1189	569 1065	480 901	393 1232	365 1477	504 1417
average	324 2178	426 1791	523 1683	630 1638	613 1334	654 1196	593 1135	484 921	396 1257	336 1246	537 1425

Table 2. Mean log F_1 and F_2 frequencies and the geometric means of the original F_1 and F_2 data.

M

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
log F_1	2.538	2.610	2.714	2.772	2.811	2.815	2.774	2.679	2.535	2.496	2.722
log F_2	3.358	3.302	3.263	3.249	3.141	3.076	3.092	2.975	3.094	2.981	3.187
F_1	345	407	517	592	647	653	594	478	343	314	528
F_2	2282	2003	1874	1774	1385	1191	1235	944	1241	957	1509

C

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
log F_1	2.472	2.664	2.756	2.816	2.803	2.823	2.781	2.690	2.640	2.504	2.756
log F_2	3.333	3.233	3.218	2.182	3.117	3.081	3.043	2.951	3.107	3.090	3.133
F_1	296	461	572	654	635	665	603	490	437	319	570
F_2	2152	1710	1653	1521	1310	1204	1105	892	1279	1229	1357

B

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
log F_1	2.501	2.587	2.670	2.787	2.735	2.806	2.749	2.678	2.588	2.551	2.697
log F_2	3.319	3.233	3.199	3.210	3.109	3.075	3.024	2.949	3.087	3.167	3.150
F_1	318	386	468	612	544	639	561	476	388	356	498
F_2	2085	1710	1580	1623	1285	1187	1056	889	1221	1468	1414

comb

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
log F_1	2.503	2.622	2.713	2.793	2.781	2.814	2.768	2.682	2.540	2.518	2.725
log F_2	3.336	3.251	3.224	3.212	3.122	3.077	3.051	2.957	3.097	3.084	3.152
F_1	318	418	516	621	604	652	586	417	389	330	531
F_2	2166	1783	1676	1628	1324	1194	1126	906	1246	1213	1419

Table 3. Mean frequencies F_1 and (F_2-F_1)

M

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
F_1	351	411	522	600	651	655	600	481	348	317	533
F_2-F_1	1935	1596	1315	1178	746	540	642	486	910	671	984

C

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
F_1	300	470	576	664	643	667	610	493	441	328	573
F_2-F_1	1856	1245	1081	866	679	537	500	408	841	906	785

B

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
F_1	325	393	476	620	553	642	569	479	393	365	504
F_2-F_1	1780	1319	1109	1008	739	547	496	421	839	1112	913

comb

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
F_1	324	426	523	630	613	654	593	484	396	338	537
F_2-F_1	1853	1365	1160	1007	721	541	543	437	861	908	889

Table 4. Elements of the covariance matrices for the (F_1, F_2) design.

M

	i	ɪ	e	æ	ʌ	a	ɒ	ɔ	ʊ	u	ɜ
s_1^2	3928	3333	4083	9327	4931	3324	6613	2921	3582	2345	5674
s_{12}	-2180	1453	-736	-8125	-8105	985	-8548	3801	-1395	6305	-1507
s_2^2	16024	16866	9989	15145	39182	9281	21147	50203	44493	68689	29627

C

	i	ɪ	e	æ	ʌ	a	ɒ	ɔ	ʊ	u	ɜ
s_1^2	2054	7505	4172	13609	8404	3168	7765	3818	3661	5506	4107
s_{12}	1429	3761	1870	-5583	676	-358	-1500	-951	-659	4240	799
s_2^2	18488	17149	15315	29660	33447	2155	12471	19540	7528	12357	5126

B

	i	ɪ	e	æ	ʌ	a	ɒ	ɔ	ʊ	u	ɜ
s_1^2	3706	5300	7397	9541	10395	3348	8897	3728	4577	7030	5670
s_{12}	2719	2862	7216	227	-2065	-220	7337	963	146	5410	-3982
s_2^2	87166	7140	1715	16519	17900	3033	19974	24386	31892	29262	12029

comb

	i	ɪ	e	æ	ʌ	a	ɒ	ɔ	ʊ	u	ɜ
s_1^2	3586	6669	7001	11584	9951	3356	7961	3505	5269	5429	5936
s_{12}	1827	1597	4199	-7024	-1447	156	89	1114	-193	954	-2439
s_2^2	46994	30169	24831	31104	31078	4552	22972	30970	27098	73894	18335

Table 5. Elements of the covariance matrices for the (log F_1, log F_2) design. Multiplied by 10^2.

M

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
s_1^2	0.6387	0.4007	0.3673	0.5631	0.246	0.157	0.379	0.2425	0.5436	0.463	0.3917
s_{12}	-0.0552	0.0306	-0.0195	-0.1481	-0.166	0.270	-0.214	0.1678	-0.0667	0.3077	-0.2321
s_2^2	0.0627	0.0815	0.0566	0.0869	0.344	0.118	0.218	0.852	0.588	1.2185	0.2297

C

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
s_1^2	0.4665	0.7522	0.246	0.6066	0.4613	0.1441	0.4235	0.2873	0.3802	0.9755	0.240
s_{12}	0.0487	0.0986	0.0388	0.0949	0.0524	0.0097	-0.0442	-0.0319	-0.0178	0.1880	0.0219
s_2^2	0.0825	0.1139	0.2757	0.2268	0.3273	0.0288	0.1957	0.4177	0.0864	0.1510	0.0530

B

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
s_1^2	0.9372	0.763	0.696	0.5062	0.7308	0.1636	0.5644	0.2927	0.5265	0.9576	0.4872
s_{12}	0.0667	0.0902	0.2113	0.0033	0.5572	0.0064	0.2752	0.0330	0.0056	0.2072	-0.1175
s_2^2	0.3646	0.0473	0.1356	0.1140	0.1841	0.0402	0.3631	0.4864	0.3487	0.2361	0.1083

comb

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
s_1^2	0.7472	0.7665	0.5743	0.5873	0.6046	0.1585	0.4690	0.2751	0.6477	0.8605	0.4327
s_{12}	0.0511	0.0541	0.1161	-0.1263	-0.1755	0.0040	0.0372	0.4994	-0.0004	0.3898	-0.0626
s_2^2	0.1993	0.1726	0.1710	0.2210	0.2946	0.0590	0.3365	0.5787	0.3302	1.0605	0.1546

Table 6. Elements of the covariance matrices in the F_1, F_2-F_1 design.

M

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
s_1^2	3928	3333	4083	9326	4931	3323	6613	2021	3582	2345	5674
s_{12}	-6107	-1880	-4818	-17452	-1304	-2338	-15161	880	-4977	3960	-7181
s_2^2	24311	17293	15543	40722	60322	10634	44856	45522	50865	58424	38314

C

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
s_1^2	2054	7505	4172	13609	8404	3168	7765	3818	3661	5506	4107
s_{12}	-625	-3745	-2302	-19101	-7728	-3526	-9265	-4769	-4320	-1266	-3308
s_2^2	17684	17134	15748	54435	40498	6039	23235	25260	12507	9383	7635

B

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
s_1^2	3706	5300	7397	9541	10395	3348	8897	3728	4577	7030	5670
s_{12}	-987	-2438	-181	-9314	-12460	-3569	-1560	-2766	-4431	-1620	-9651
s_2^2	85433	6715	9980	25606	32426	6822	14196	26188	36177	25473	25662

comb

	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
s_1^2	3586	6669	2001	11584	9951	3355	7961	3505	5269	5429	5936
s_{12}	-1758	-5071	-2802	-18608	-11398	-3200	-2872	-2390	-5462	3625	-8375
s_2^2	46925	33643	23434	56736	43922	7595	30755	32246	32753	61215	29149

Table 7. Confusion matrices in the F_1, F_2 design.

M

classified \ spoken	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
i	5										
ɪ		5									
e			3								
æ			1	6							1
ʌ					2						2
ɑ					3	4					
ɒ							1				1
ɔ							3	4			
ʊ									3	1	
u									1	3	
ɜ											1

K

classified \ spoken	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
i	5										
ɪ		2									
e		3	4	2							
æ				4							
ʌ					1						1
ɑ					3	6	1				
ɒ							3				
ɔ								4			
ʊ									4	1	
u										3	
ɜ					1						4

spoken

classified	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
i	5										
ɪ		5	1							1	
e			2								
æ				5							
ʌ					1						1
ɑ					3	4	2				
ɒ							1				
ɔ							1	4			
ʊ					1				4	1	
u										2	
ɜ			1	1							4

F

spoken

classified	i	ɪ	e	ɐ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
i	5	3									
ɪ		8	1							1	
e		3	5	1							1
æ			3	13	1						
ʌ				2	2						2
ɑ					10	11	8				1
ɒ						1	3		2		
ɔ							1	12			
ʊ		1			1				9	4	
u			1						1	7	
ɜ			2	2	1						11

comb

Table 8. Confusion matrices in the $\log F_1$, $\log F_2$ design.

M

spoken

classified	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
i	5										
ɪ		5									
e			4								1
æ			1	6							
ʌ					2						2
ɑ					3	4	3				
ɒ							1				
ɔ								4			
ʊ									3	1	
u									1	3	
ɜ											2

C

spoken

classified	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
i	5										
ɪ		2									
e		3	4	2							
æ				4							
ʌ					1						1
ɑ					3	4					
ɒ							1				
ɔ							3	4			
ʊ									3	1	
u										3	
ɜ					1				1		4

spoken

classified	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
i	5										
ɪ		5	1							1	
e			1							1	
æ			2	6							1
ʌ					1				1		1
ɑ					2	4	2				
ɒ					1		1				
ɔ							1	4			
ʊ					1				3		
u										2	
ɜ											3

B

spoken

classified	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
i	15	3									
ɪ		8	1							1	
e		3	6	1							1
æ			3	13	1						
ʌ				2	1						2
ɑ					11	11	7				1
ɒ						1	4				
ɔ							1	12			
ʊ		1			1				9	3	
u									1	7	
ɜ			2	2	1				2	1	11

comb

Table 9. Confusion matrices in the F_1, F_2-F_1 design.

M

spoken

classified	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
i	5										
ɪ		5									
e			3								1
æ			1	6							
ʌ					2						2
ɑ					3	4	4				
ɒ							1				1
ɔ								4			
ʊ									3	1	
u									1	3	
ɜ											1

C

spoken

classified	i		e							u	
i	5										
ɪ		2									
e		3	4	2							
æ				4							
ʌ					1						1
ɑ					3	4	1				
ɒ							3				
ɔ								4			
ʊ									4	1	
u										3	
ɜ					1						5

spoken

classified	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
i	5										
ɪ		5	1							1	
e			2								
æ				5							
ʌ					1						1
ɑ					3	4	4				
ɒ							1				
ɔ								4			
ʊ									4	1	
u					1					2	
ɜ			1	11							4

B

spoken

classified	i	ɪ	e	æ	ʌ	ɑ	ɒ	ɔ	ʊ	u	ɜ
i	15	3									
ɪ		8	1						1	1	
e		3	5	1							
æ			3	13	1						2
ʌ				2	2						1
ɑ					10	11	8				
ɒ						1	3		2		
ɔ							1	12			
ʊ			1		1				9	4	
u		1							1	7	
ɜ			2	2	1						11

comb

Table 10. Correct assignment of the vowel in individual words.

variables / word	F_1 F_2	$\log F_1$ $\log F_2$	F_1 F_2-F_1	variables / word	F_1 F_2	$\log F_1$ $\log F_2$	F_1 F_2-F_1
1. teethe	3	3	3	26. vase	3	3	3
2. seethe	3	3	3	27. carve	3	3	3
3. tease	3	3	3	28. farce	3	3	3
4. teeth	3	3	3	29. path	3	3	3
5. thief	3	3	3				
6. bib	3	3	3	30. bog	2	2	2
7. sieve	3	3	3	31. dog	2	2	2
8. zip	2	2	2	32. cough	ø	ø	ø
9. thick	2	2	2	33. shock	1	1	1
10. this	2	2	2	34. daub	3	3	3
11. zed	1	1	1	35. cause	3	3	3
12. fez	2	2	2	36. court	3	3	3
13. set	3	3	3	37. thorp	3	3	3
14. death	3	3	2	38. good	3	3	3
15. gag	2	2	2	39. should	3	3	2
16. zag	3	3	3	40. push	2	2	2
17. tag	3	3	3	41. soot	3	3	2
18. that	2	2	3	42. food	2	2	2
19. pack	3	3	3	43. soothe	3	3	3
20. cash	2	2	2	44. shoot	ø	ø	ø
21. dove	1	1	1	45. goose	3	3	3
22. thud	ø	ø	ø	46. curve	2	2	2
23. buzz	ø	ø	ø	47. serve	2	2	3
24. gush	3	3	3	48. shirk	2	1	1
25. tough	ø	ø	ø	49. purse	3	2	1
				50. burp	3	2	2

Fig 1

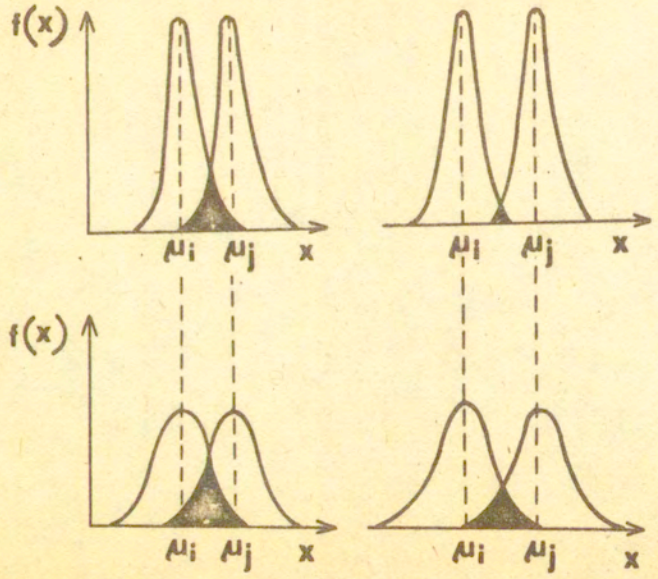


Fig. 1. The effect of the differences between means and differences between the variances on the probability of misassignment.

Fig 2

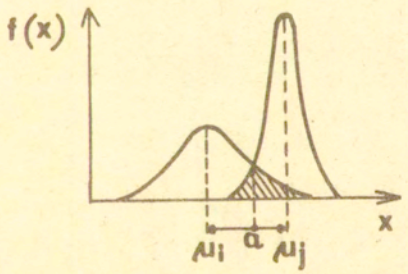


Fig. 2. The effect of a difference in variances on the position of the 'discriminant point'.

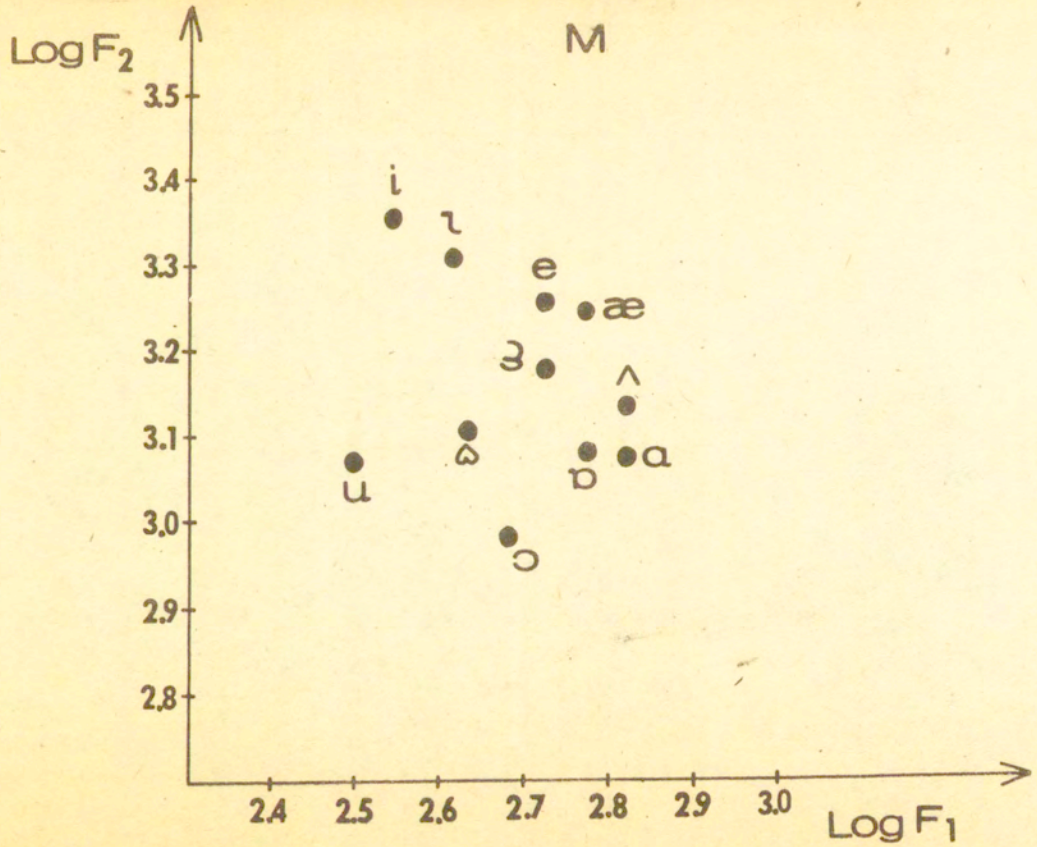


Fig. 3. Mean log frequencies of the formants of the 11 GBE monophtongs. Speaker M.

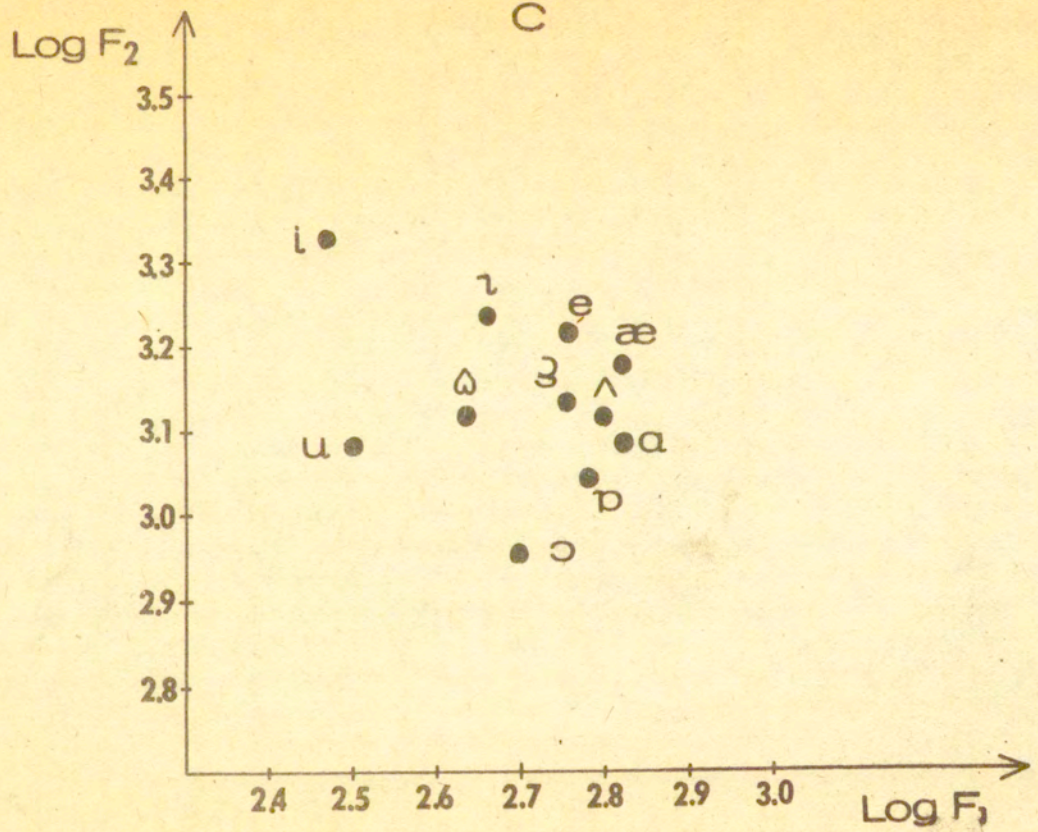


Fig. 4. Mean log frequencies of the formants of the 11 GBE monophthongs. Speaker C.

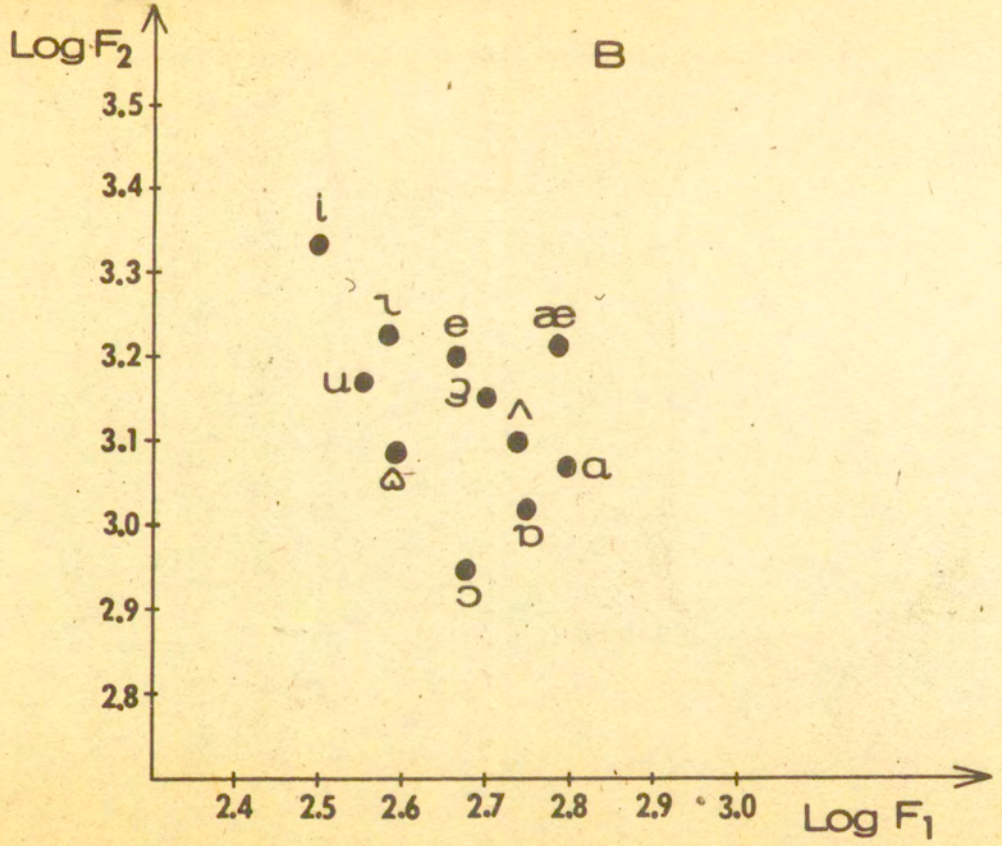


Fig. 5. Mean log frequencies of the formants of the 11 GBE monophthongs. Speaker B.

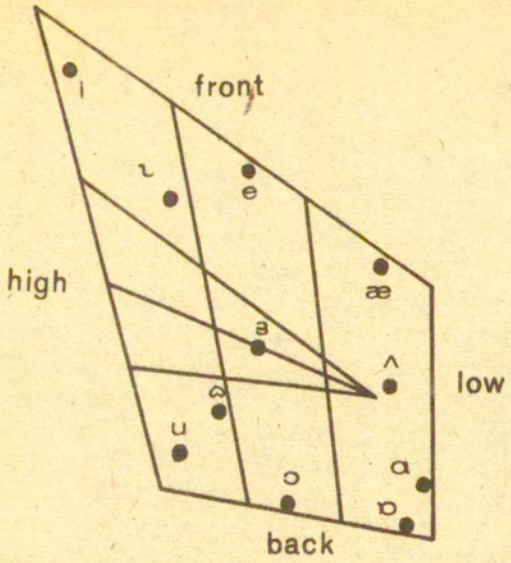


Fig. 6. The GBE monophthongs on the IPA quadrilateral.

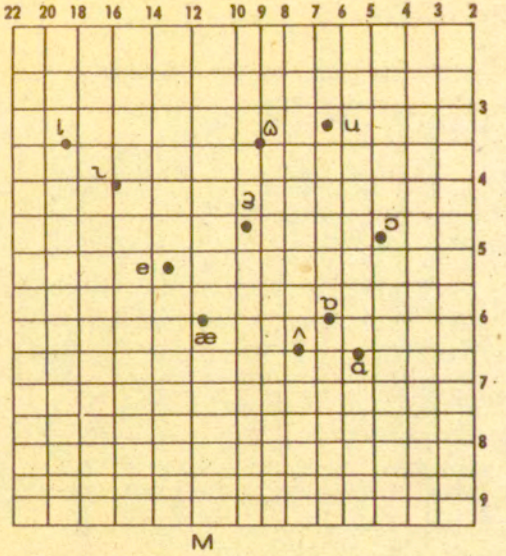


Fig. 7. Mean values of M's F_1 and (F_2-F_1) on Ladefoged's (1975) chart.

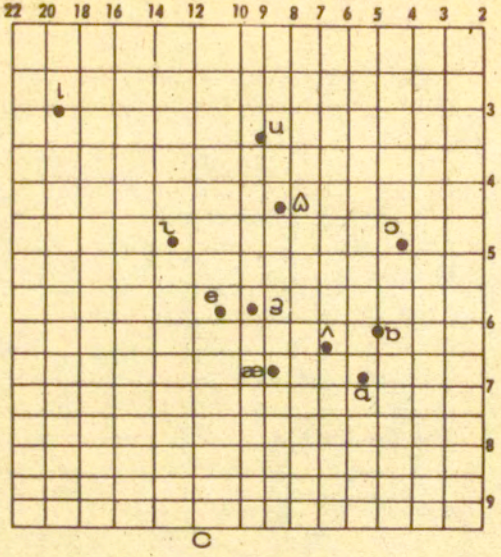


Fig. 8. Mean values of C's F_1 and (F_2-F_1) on Ladefoged's (1975) chart.

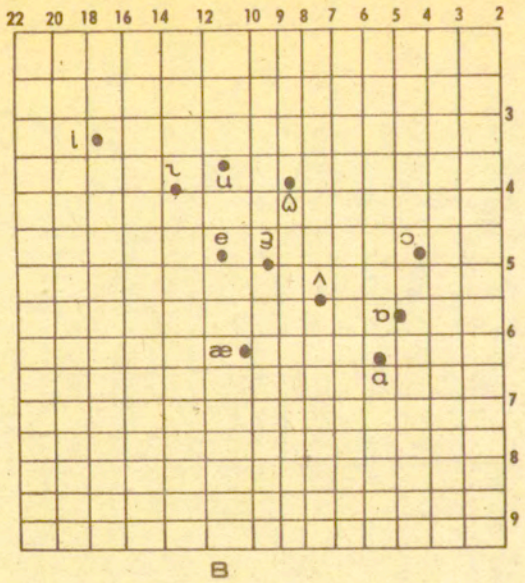


Fig. 9. Mean values of B's F_1 and (F_2-F_1) on Ladefoged's (1975) chart.

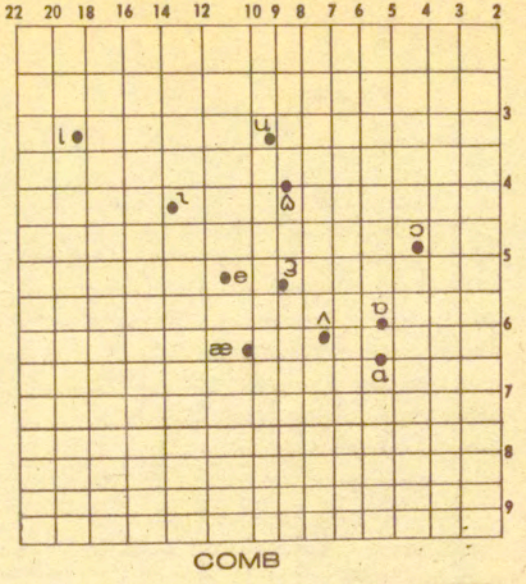


Fig. 10. Mean values of F_1 and (F_2-F_1) on Ladefoged's (1975) chart. Pooled data.

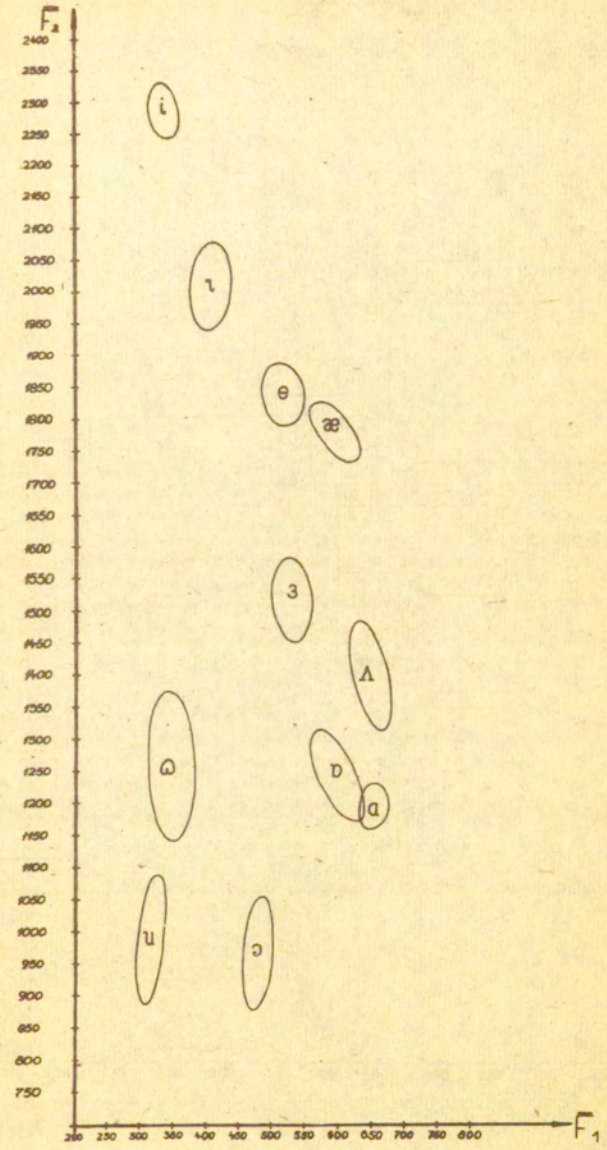


Fig. 11. 95 % confidence ellipses for the mean formant frequency vectors in the (F_1,F_2) design. Speaker M.

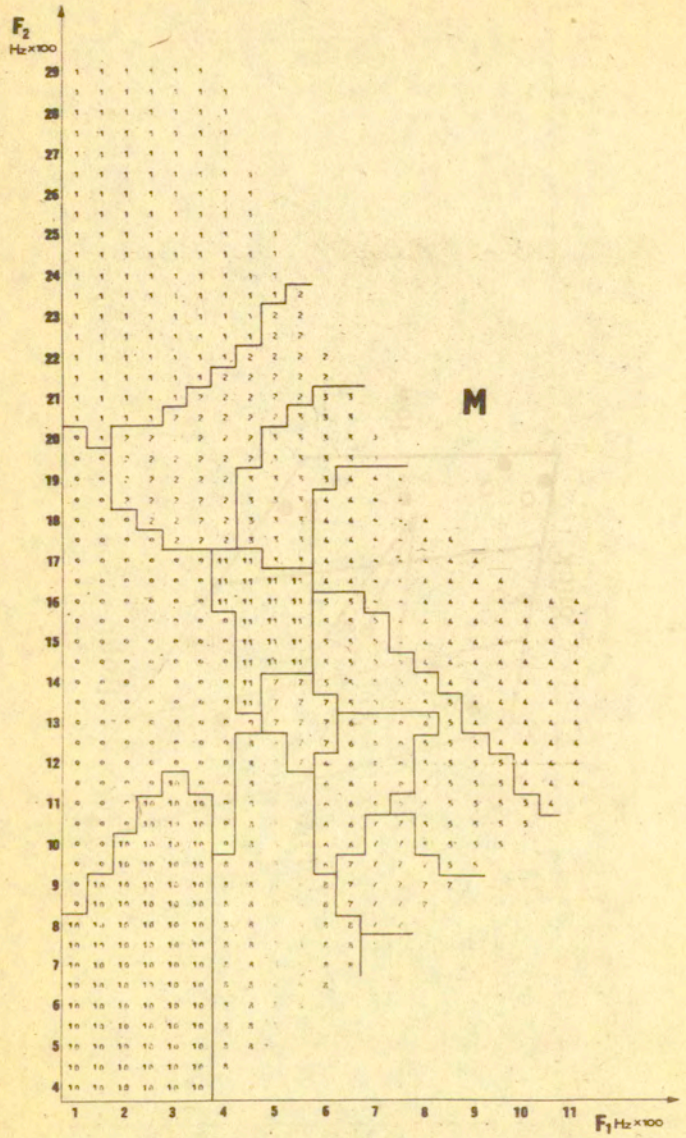


Fig. 12. Vowel chart for the (F_1, F_2) design. Speaker M.

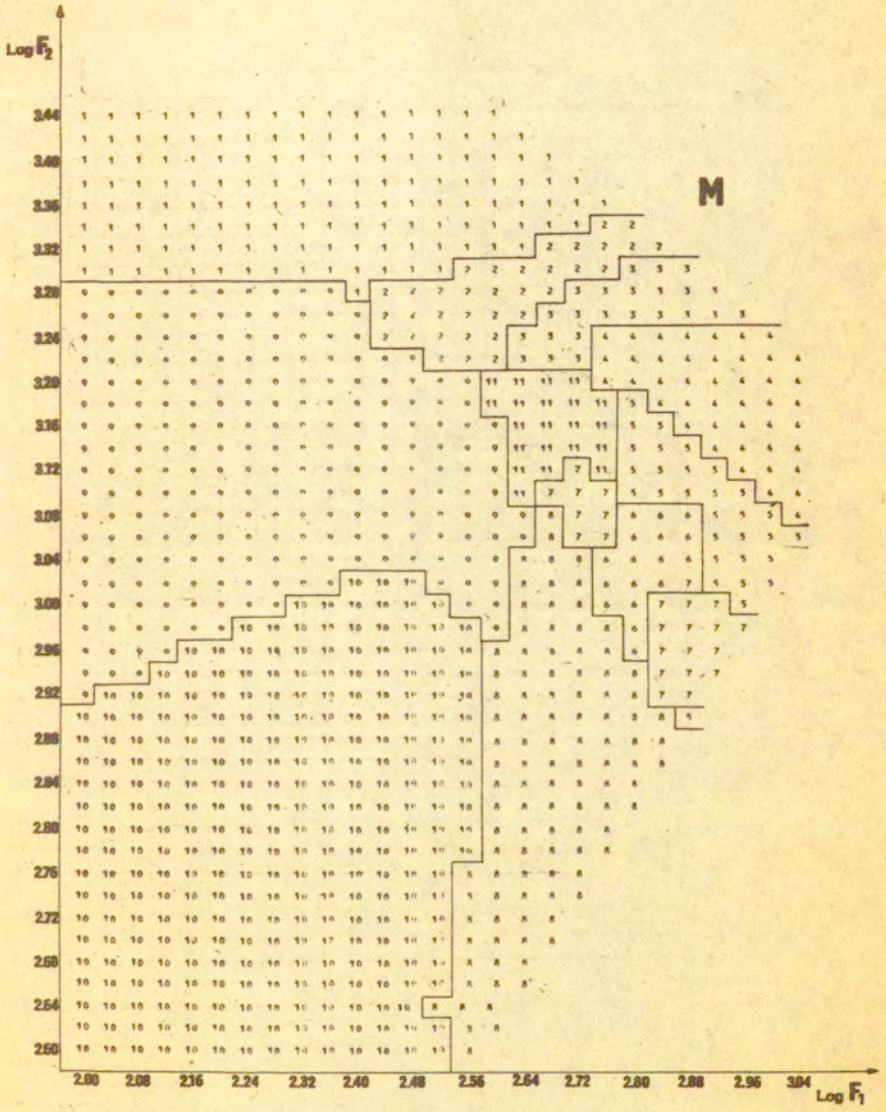


Fig. 13. Vowel chart for the ($\log F_1$, $\log F_2$) design. Speaker M. Equal logF distances.

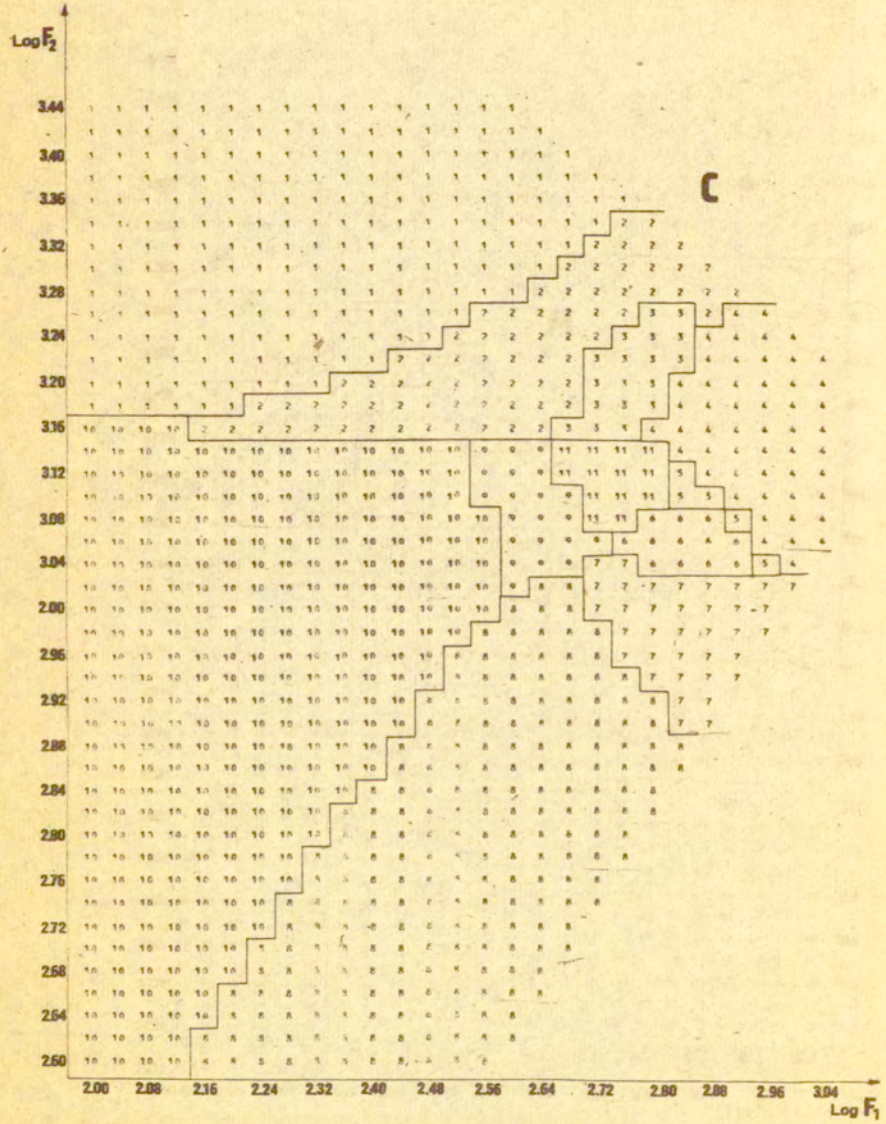


Fig. 14. Vowel chart for the ($\log F_1$ $\log F_2$) design. Speaker C. Equal logF distances.

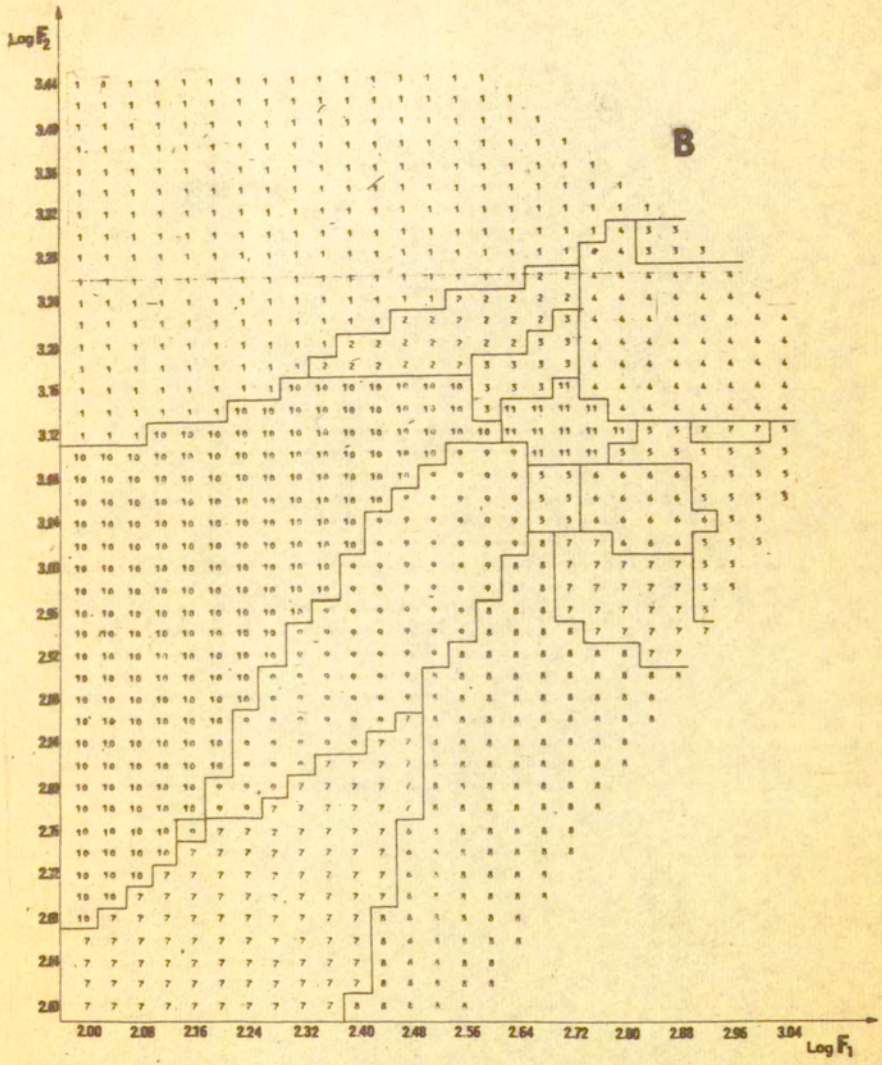


Fig. 15. Vowel chart for the ($\log F_1$, $\log F_2$) design.
Speaker B. Equal logF distances.

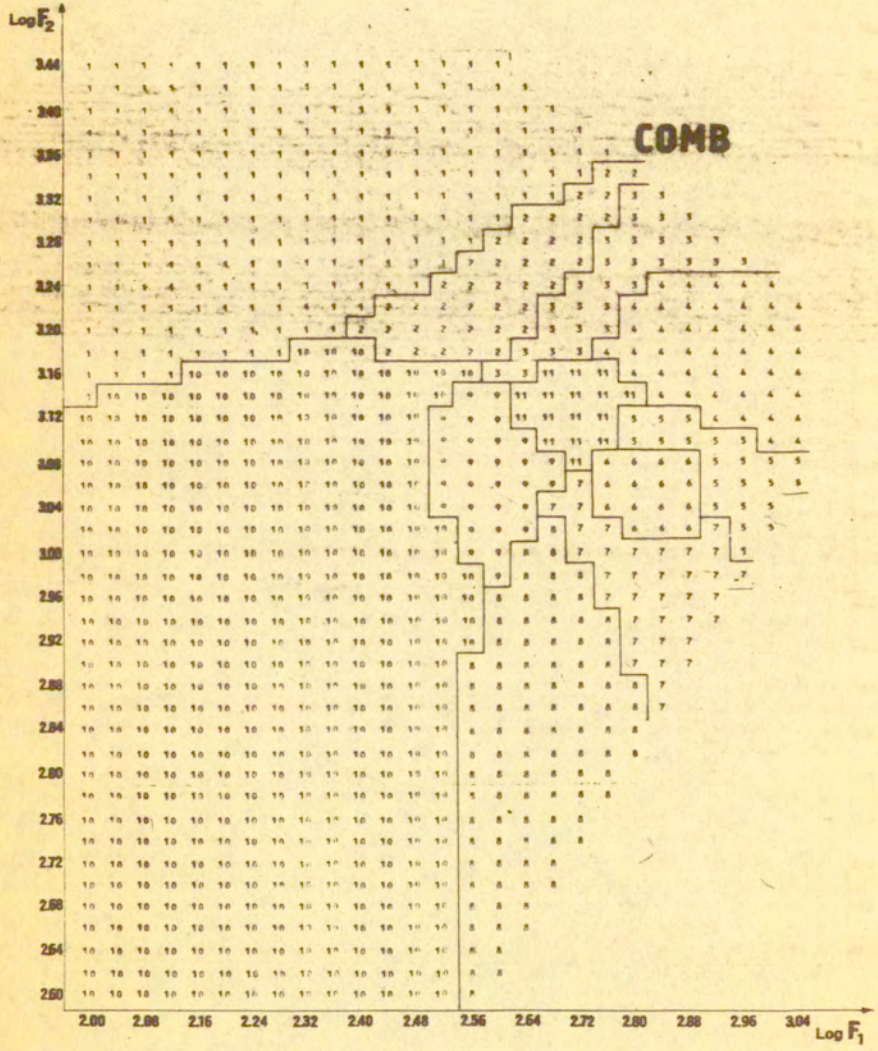


Fig. 16. Vowel chart for the ($\log F_1$, $\log F_2$) design. Combined data. Equal logF distances.

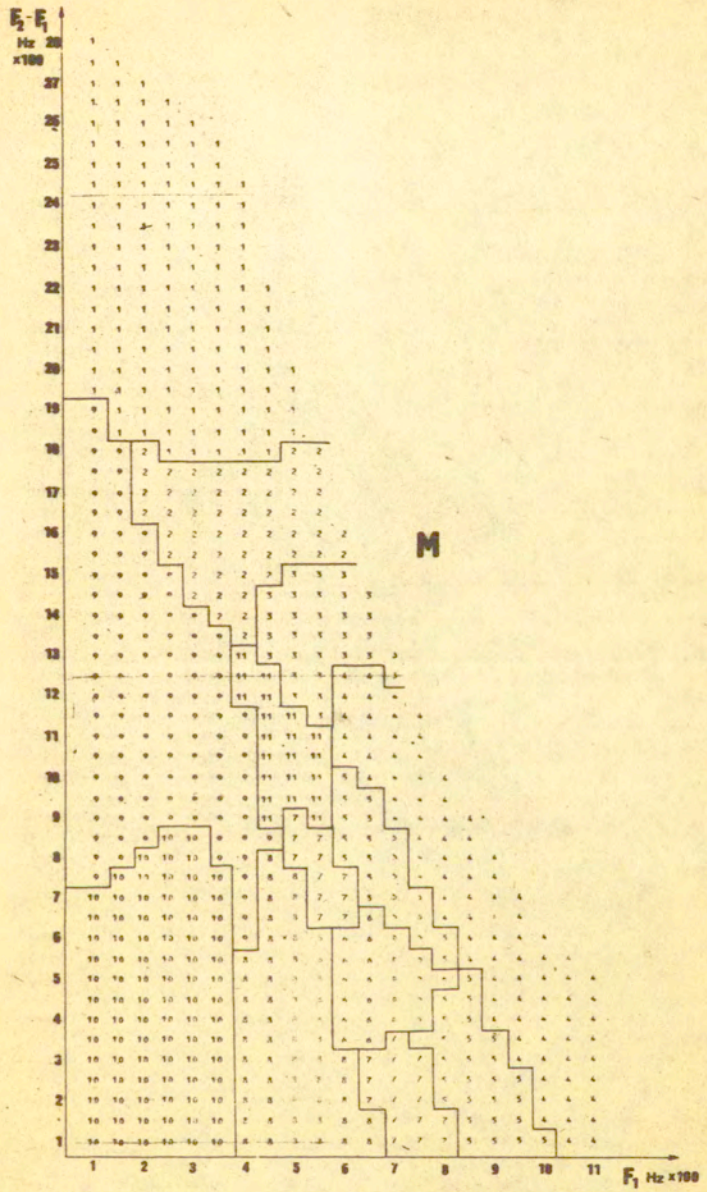


Fig. 17. Vowel chart for the (F_1, F_2-F_1) design. Speaker M.

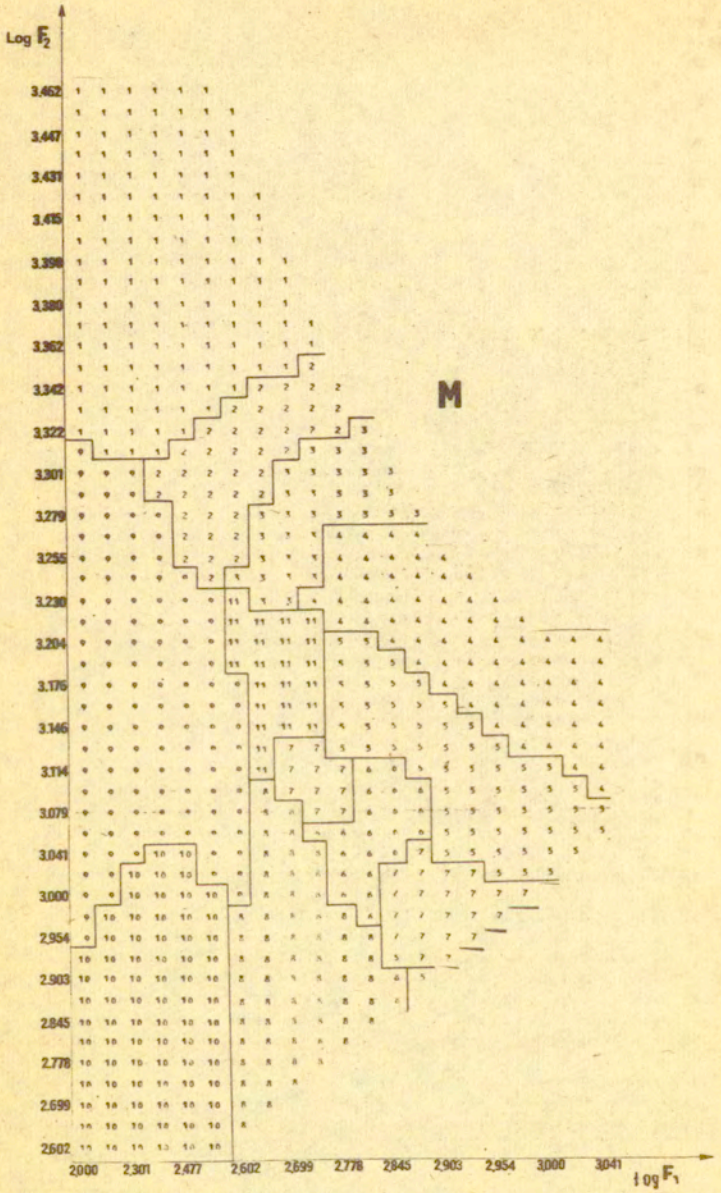


Fig. 18. Vowel chart for the ($\log F_1$, $\log F_2$) design. Equal F = 50 Hz distances.

REFERENCES

ARNOLD, G.F., DENES, P., GIMSON, A.C., O CONNOR, J.D., TRIM, J.C.M. (1958). The synthesis of English vowels, Language and Speech, 1, 114-125.

ABERCROMBIE, D. (1964). English Phonetic Texts, Faber and Faber, London.

BROAD, D. and FERTIG, R.M. (1970). Formant frequency trajectories, Journ. Acoust. Soc. Am., 47, 1572-1582.

BROAD, D. and WAKITA, H. (1978). A phonetic approach to vowel recognition, in: Speech Communication with Computers L.Bolc, ed. , Hauser Verl. München/Wien, 54-92.

DISNER, S.F. (1978). Vowels in Germanic Languages, Univ. of California Los Angeles Working Papers in Phonetics No.40.

FAIRBANKS, G. and GRUBBS, P. (1961). A psychological investigation of vowel formants, Journal of Speech and Hearing Research, 4, 203-219.

FANT, G. (1956). On the predictability of formant levels and spectrum envelope from formant frequencies, For Roman Jakobson, Mouton, s'Gravenhage, 109-120.

FANT, G. (1960). Acoustic Theory of Speech Production, Mouton, s'Gravenhage.

FLANAGAN, J.L. (1956). Automatic extraction of formant frequencies from continuous speech, Journ. Acoust. Soc. Am. 28, 110-118.

FORGIE, J.W. and FORGIE, C.D. (1959). Results obtained from a vowel recognition computer program, Journ. Acoust. Soc. Am., 31, 1480-1489.

GERSTMAN, L.S. (1968). Classification of self-normalized vowels, IEEE Trans. AU-16, 76-80.

GIMSON, A.C., (1962). An Introduction to the Pronunciation of English, Arnold, London.

Hanne, J.R. (1965). Formant Analysis, Univ. of Michigan Communication Sciences Laboratory, Report No.12.

IMIOŁCZYK, J. forthcoming . The duration of GBE monophthongs in monosyllabic words.

JASSEM, W. (1950). Phonemic transcription of the vowel sounds

of Educated Southern English, Le maître phonétique, III/93, 10-12.

JASSEM, W. (1973). Podstawy fonetyki akustycznej Bases of acoustic phonetics , PWN, Warszawa.

JASSEM, W. (1979). Podręcznik wymowy angielskiej A Handbook of English Pronunciation , 3rd ed., PWN, Warszawa.

JASSEM, W. (in print). The Phonology of Modern English, PWN, Warszawa.

JASSEM, W., GEMBIAK, D., DYCZKOWSKI, A. (1979). Computer-aided recognition of Polish vowels in continuous speech, Archives of Acoustics 4, 39-54.

JASSEM, W., KRZYŚKO, M., DYCZKOWSKI, A. (1972). Klasyfikacja i identyfikacja samogłosek polskich na podstawie częstotliwości formantów Classification and identification of Polish vowels on the basis of their formant frequencies , Prace IPPT No.64/1972, Warszawa.

JASSEM, W., KRZYŚKO, M., DYCZKOWSKI, A. (1976). Identification of isolated Polish vowels, Speech Analysis and Synthesis W. Jassem, ed. IV, PWN, Warszawa, 107-134.

JASSEM, W. and ŁOBACZ, P. (1976). Frequency of phonemes and their sequences in Polish texts, Speech Analysis and Synthesis (W. Jassem, ed.) vol. 4, 241-249.

JOOS, M. 1948 . Acoustic Phonetics, Suppl, to Language 24, Baltimore.

JONES, D. (1950). The Phoneme, its Nature and Use, Heffner Cambridge.

JONES, D. (1956). Outline of English Phonetics, 8th ed., Heffer, Cambridge.

KUBZDELA, H. (1973). Automatyczna ekstrakcja tonu podstawowego oraz pierwszych trzech formantów sygnału mowy (Automatic extraction of fundamental frequency and the first three formants from the speech signal), Prace IPPT No. 51/1973, Warszawa.

LADEFOGED, P. (1975). A Course of Phonetics, Harcourt, Brace, Jovanovich, New York.

LINDBLOM, B., SUNDBERG, J. (1968). A qualitative model of vowel production and the distinctive features of Swedish vowels, Speech Transmission Lab., Quarterly Progress and Status Report KTH No. 1, 14-30.

ŁOBACZ, P. (1976). Speech rate and vowel formants, Speech

Analysis and Synthesis (W.Jassem, ed.) IV, PWN, Warszawa, 187--218.

MacCARTHY. (1947). English Pronunciation, 3rd ed., Heffner Cambridge.

MARKEL, J.D., GRAY, A.H., (1974). Linear Prediction of Speech. Springer Verlag Berlin.

NEARLY, T.M. (1978). Phonetic feature systems for vowels, Indiana Univ. Linguistics Club, Bloomington, Indiana.

O'CONNOR, J.D. (1967). Better English Pronunciation, Cambridge Univ. Press.

PAPÇUN, G. (1980). How Do Different Speakers Say the Same Vowels? Univ. of California Los Angeles Working Papers on Phonetics No. 48.

PETERSON, G.B., BARNEY, H.L. (1952). Control methods used in a study of vowels, Journ. Acoust. Soc. Am. 24, 175-184.

PLOMP, R., POLS, L.C., van de GEER, J.P. (1967). Dimensional analysis of vowel spectra, Journ. Acoust. Soc. Am. 41, 707-712.

POLS, L.C.W. (1971). Dimensional representation of speech spectra, Proc. 7th Intern. Congress on Acoustics, Budapest, 25C7.

POLS L.C.W. (1977). Spectral Analysis and Identification of Vowels in Monosyllabic Words, Institute for Perception TNO Soesterberg.

POTTER, R.K., STEINBERG, J.C. (1950). Towards the specification of speech, Journ. Acoust. Soc. Am. 22, 807-820.

STEVENS, K.N., HOUSE, A.S. (1961). An acoustical theory of vowel production and some of its implications, Journ. of Speech and Hearing Res., 4, 303-320.

WELLS, J.C., COLSON, G. (1971). Practical Phonetics repr. Pitman, London.