ARCHIVES OF ACOUSTICS
17, 3, 343–356 (1992)

IN QUEST FOR CARDINAL VOWELS

W. JASSEM and M. KRZYŚKO

Department of Acoustic Phonetics
Institute of Fundamental Technological Research
Polish Academy of Sciences
Poznań, ul. Noskowskiego 10

Computing Centre, A. Mickiewicz University
Poznań

180 isolate voicings of (near) Cardinal Vowels – 10 of each: [ieɛayøœɶuoɔɒ ɯɤʌɒɨʉ] were described using four formant frequencies, which were measured from FFT spectra. In a 4-D space, the tokens could be correctly assigned 95% of the time using Bayes estimators of discriminant scores. The error increases with the reduction of the number of variables to three or two. The mean vectors characterizing the 18 vowel classes were, for the present speaker, somewhat different from those recorded and measured about 20 years ago, which may be due to aging.

1. The background

Until fairly recently, the Cardinal Vowels were used only for the purposes of Linguistic Phonetics. But for over 3 decades now specialists in Speech Acoustics, Speech Pathology and perhaps most of all, Speech Technology have been to a smaller or greater extent forced to use the transcription of the International Phonetic Association at some place or other in their routine work. These specialists often have decidedly insufficient knowledge of the system of IPA (International Phonetic Association) with the result that even if their primary problem is correctly solved, erronious transcription confuses phoneticians who wish to use their data.

In what follows we shall concentrate on those aspects of the Cardinal Vowels that are of interest to the *Speech Technologist*, the *Speech Pathologist* and the *Acoustician* who is dealing with the speaking or singing voice.

The Cardinal Vowels were devised by Daniel Jones some time between 1910 and 1920, and were based on the observation of four extreme vocalic articulations with respect to the natural, unmodified position of the tonge, its hump taking a maximally (1) high-front, (2) low-front, (3) low-back and (4) high-back position. Of these, (1), (3) and (4) can be indirectly controlled by the speaker's tactile sense, their articulations being minimally different from the palatal [ʝ], uvular [ʁ] and velar [ɣ]. The front-low position can only be controlled auditorily and, possibly, kinaesthetically. The four vowels were given the phonetic symbols [**i a ɑ u**]. The choice of the symbols was a mat-matter of convenience, and the three represented by letters of the roman alphabet are just as normal as the fourth. They were all intended as international reference values independent of any particular language. In 1917 Jones made X-rays of the four extreme vowels and these were published in the later editions of Jones's *The Pronunciation of English* (e.g. Jones 1956 [11] as a frontispiece. They formed the basis of the vowel quadrilateral, which, with minor changes, has been in use by those adhering to the principles of the International Phonetic Association.

Over a period of about 70 to 80 years, the set of Cardinal Vowels was made to include initially (beside the extreme four) an additional set of four, viz. [**e ɛ ɔ o**]. For some time it was maintained by many phoneticians, including D. Jones himself, that in the front series [**i e ɛ a**] the *articulatory* distances [i]–[e], [e]–[ɛ] and [ɛ]–[a] were equal, as were the distances in the back series: [u]–[o], [o]–[ɔ], [ɔ]–[ɑ]. The total system was, then:

i				**u**				close
	e				**o**			half-close
		ɛ				**ɔ**		half-open
			a				**ɑ**	open
			front				back	

The trapezoidal form of the arrangement, as seen above, reflects the assumed positions of the tongue hump and has here been slightly simplified (see, e.g. Jassem (1973), p. 124 [9]).

For several decades the trapezoid was used by phoneticians to describe the vowels of various languages e.g. for English in Jones (1956) [11], Roach (1983) [23], and Gimson (1988) [8], for French by Armstrong and Jones (1951) [2], for Russian by Jones and Ward (1969) [12], for Polish by Jassem (1973) [9] etc.

The articulatory basis for the 8-item set of Cardinal Vowels was never confirmed experimentally. What little X-ray work was done on their articulation actually falsified that basis (Butcher (1982) [3]). For some forty years now it has been maintained that the trapezoid represents *auditory* (i.e. psychoacoustic) relations. This, too, has been questioned by Butcher [3].

The 8 Cardinal Vowels (CVs) were not sufficient for the description of many languages because it was assumed that [**i ɛ a ɑ**] were unrounded (i.e. spoken with neutral position of the lips) whilst [**a ɔ o u**] have, in that order, increasing lip-roundig. Although this reflected a strong tendency, many languages have rounded front vowels and some have unrounded back ones. Such symbols as [y] for "close front rounded" or [ø] for "front half-close rounded" were used even towards the end of the 19th century within the International Phonetic Alphabet, but it was not before the end of the II World War that the complete set of 18 CVs was established. Ignoring the trapezoidal shape of the schematic, the full set of CVs is now arranged as follows:

	front		central close		back	
	unrounded	rounded	unrounded	rounded	unrounded	rounded
			ɨ	ʉ		
close	i	y			ɯ	u
half-close	e	ø			ɤ	o
half-open	ɛ	œ			ʌ	ɔ
open	a	ɶ			ɒ	ɑ

The front unrounded and back rounded vowels are often referred to as "primary" and the "newer" 8 as "secondary". In the above arrangement of the CVs, the secondaries lie *within*, the primaries if looked upon as placed in a plane.

For several decades the set of 8 primaries or the set of all the 18 CVs has been used in texts on the phonetics of many languages (see a comparative sample in JASSEM (1973) [9] 130 and 134). But X-ray studies of the CVs as well as those of real linguistically used vowels tended to show that the assumed articulatory basis was incorrect (BUTCHER (1985) [3]). It was also shown *loc.cit.* that the *perception* of the CVs was strongly affected by the first language of the hearer. It would seem that the only hopeful level at which the identity of the CVs could be sought was the *acoustic* level.

2. Earlier acoustical data

The earliest data on the acoustic properties of the Cardinal Vowels come from a paper by DELATTRE, LIBERMAN and COOPER (1951) [6] who produced them synthetically. Figures 1 and 2 in that paper show the synthesized 16 vowels, 15 of which were intended to represent the IPA (International Phonetic Association) CVs, in an (F_1F_2) *acoustical plane*. On a straight low–F_1 (first formant) line (250 Hz) lie [**i yɯ u**] – the close vowels. Along three other straight lines (1) at $F_1 = 440$ Hz (2) $F_1 = 550$ Hz and (3) $F_1 = 750$ Hz lie [**eøɤɔ**], [**ɛ œ ʌ ɔ**] and [**æ a ɒ ɑ**], in that order. These relations should be compared with the 8-vowel and 18-vowel arrangements above. Within each of the four subsets, F_2 decreases in the order indicated within brackets. Only the last subset of four requires comment. Cardinal [ɶ] had not yet been approved by the International Phonetic Association at the time of the Delattre–Liberman–Cooper experiment. In the two quadrangles – one for the unrounded and the other for the rounded, published in *Fundamentos* (1944) [7] and *Principles* (1949) [20], the lower-left corner in the "rounded" quadrilateral was left unmarked. Apart from [æ], the arrangement in the acoustical (F_1, F_2) plane in DELATTRE–LIBERMAN–COOPER (1951) [6] can be seen to correspond very well with the arrangement of the sixteen non-central vowels shown above, with one exception: the positions of [ɑ] and [ɒ] were reversed. A very important feature of the Delattre–Cooper–Liberman quadrangle was that the Primaries lie on the circumference while the

Secondaries lie within the quadrangle. The figure is a trapeze, with the bottom side considerably shorter that the top side, so that the [i]–[u] distance is over 3 times that between [æ]–[ɑ]. This should be confronted with the trapezoid on p. 2 above. In the Delattre–Liberman–Cooper trapeze, the distance D(**y**– **ɯ**) is only half that of D(**i**, **u**), D(**ø**– **ɤ**) is about half that of D(**e** – **o**) and D(**œ** – **ʌ**) is distinctly less than half of D(**ɛ** – **ɔ**). This reflects the fact usually tacitly admitted by phoneticians that the “secondary” CVs are *perceptually* less distinct – or less distant – than the Primaries of the same degree of openness (approx. the same value of F_1).

Lee (1968) [18] published an (F_1, F_2) chart of the Primaries pronounced by D. Jones. The geometry of the vowel “loop” turned out quite similar to the Delattre–Liberman–Cooper trapeze, though some of the absolute frequencies were different: [**e** – **o**] and [**a**-**ɑ**] do not lie on a const F_1 line, F_1[**a**] > F_1[**ɑ**] and F_1[**e**] > F_1[**o**]. Lee was trying to prove a point which is outside the scope of the present paper. But in order to proceed with his argumentation Lee took into account an important consideration, viz. that the frequencies of F_1 and F_2 vary not only with the phonetic quality of the vowel, but also with its *personal* quality (personal timbre).

It has been realized for some 30 years now that in speech perception a process of normalization takes place which permits phones that are acoustically different to be perceived as linguistically identical. Viewed differently, the hearer extracts from the acoustical speech signal simultaneously two kinds of information: linguistic and personal. Several attempts have been made to describe, sometimes in mathematical terms, the interaction between these two essential and intertwined sources of variation (see, e.g. Ladefoged and Broadbent (1957) [17], Ainsworth (1975) [1] and, especially, Nearey (1978) [19].

Catford (1981 [4], 1988 [5]) has devised two acoustic grids with F_1 and F_2 as non-orthogonal co-ordinates: one for the rounded vowels and the other for the unrounded, and used them to describe the vowels of several languages. But he does not say how he deals with inter-speaker differences. Further data have been published by Jassem (1973) [9], 1984 [10] including those on all the lower four formant frequencies. An acoustical map of all the 18 CVs, now including also [ɨ] and [ʉ], in an (F_1, F_2) plane was demonstrated. The topography is here very similar to that in the earlier studies, but the absolute values tend to be higher, possibly pointing to a smaller vocal tract. As the earlier investigations are limited to F_1 and F_2, Table 1 below includes only such data, for comparison. Values that had to be read off the (F_1, F_2) plots (the source giving no numbers) appear here in cursive. The values which appeared in Jassem (1973) [9] and (1984) [10] are here given under WJ1. For comparison, the mean values obtained in the present experiment for F_1 and F_2 also appear in Table 1 as WJ2.

Although in terms of absolute values the data differ between the individual sources, the following regularities may be observed.

(1) F_1 increases with the degree of openness:

[i] > [e] > [ɛ] > [a] (front unrounded)
[y] > [ø] > [œ] > [ɶ] (front rounded)
[ɯ] > [ɤ] > [ʌ] > [ɒ] (back unrounded)
[u] > [o] > [ɔ] > [ɑ] (back rounded)

(2) For each degree of openness, F_2 is higher in the unrounded than for the corresponding rounded vowel:

[i] > [y], [e] > [ø], [ɛ] > [œ], [a] > [ɶ] (front series)

[ɯ] > [u], [ɤ] > [o], [ʌ] > [ɔ], [ɑ] > [ɒ] (back series) (cf. above about the relations between the open vowels in DLC)

(3) For the front series,

[i] > [e] > [ɛ] > [a] (unrounded)

[y] > [ø] > [œ] > [ɶ] (rounded)

Table 1. F_1 and F_2 frequencies of Cardinal Vowels from five sources

vowel source		i	y	e	ø	ɛ	œ	a	ɶ
DLC	F_1	*240*	*240*	*360*	*360*	*520*	*520*	*730*	——
	F_2	*2900*	*1900*	*2500*	*1700*	*1650*	*1450*	*1320*	——
Lee	F_1	250	——	375	——	525	——	775	——
	F_2	2500	——	2250	——	1800	——	1100	——
Cat	F_1	240	235	390	370	610	585	850	820
	F_2	2400	2100	2300	1900	1900	1710	1610	1530
WJ_1	F_1	210	220	380	350	590	520	870	790
	F_2	2750	2550	2630	2320	2280	1950	1750	1650
WJ_2	F_1	217	249	417	422	559	511	921	511
	F_2	2775	2255	2538	1968	2151	1769	1560	1769
		ɯ	u	ɤ	o	ʌ	ɔ	ɑ	ɒ
DLC	F_1	*250*	*250*	*360*	*360*	*520*	*520*	*730*	*730*
	F_2	*1050*	*700*	*1100*	*800*	*1180*	*950*	*1050*	*1250*
Lee	F_1	——	250	——	350	——	525	775	——
	F_2	——	625	——	775	——	900	1450	——
Cat	F_1	300	250	460	360	600	500	750	700
	F_2	1390	595	1310	640	1170	700	940	760
WJ_1	F_1	280	270	450	400	570	550	800	710
	F_2	850	615	850	730	940	820	1050	900
WJ_2	F_1	369	308	475	427	591	535	740	680
	F_2	808	577	846	686	911	805	995	931

For comparison, we would cite some data contained in Papcun (1980) [2]. Figure 6.3 in this paper contains, in the form of a graph, measured F_1 and $F_{2'}$, of the English vowels as produced by two male speakers. Except, possibly, for [ɑ]/[ɒ] the vowels represent the same linguistic-phonetic entities. In keeping with the position taken by the Phonetics Laboratory at the Department of Linguistics, University of California Los Angeles, where the data was obtained, instead of the straight F_2 frequency, the other variable, beside F_1, is $F_{2'} = F_2 - F_1$.

The F_1 and $F_{2'}$ values, reads as closely as possible from the chart are presented below in Table 2.

Table 2. Frequencies of F_1 and $F_{2'}$ of English monophthongs

	vowel	[i]	[ɩ]	[E]	[æ]	[ɑ] [ɒ]	[ɔ]	[ω]	[u]
speaker 1	F_1	270	400	550	700	710	360	450	310
	F_2	1950	1550	1250	990	390	290	590	560
speaker 2	F_1	250	270	380	470	370	580	300	220
	F_2	1620	1500	1280	950	460	290	430	460

Both voices were male.

As can be seen from Table 2, the differences in the values for F_1 and $F_{2'}$ of linguistically equivalent vowels may be striking. They are due to the effect of *personal* timbre. If spoken by *the same* voice, the corresponding values in each of the columns would easily represent different, even very different linguistic-phonetic entities.

3. The present materials and their acoustic analysis

The purpose of the present experiment is fivefold:

(i) to establish reasonably narrow Cardinal Vowel subspaces in a four-dimensional vowel space.

(ii) to find for each CV a 4-element mean vector for each CV in the vowel space,

(iii) to perform a statistical discriminant analysis of the vectors representing the complete set of 18 Cardinal Vowels,

(iv) to see whether the CVs might be affected by the speaker's aging.

(v) to find, using statistical discriminant analysis, whether the frequencies of the higher formants, i.e. F_3 and F_4 contribute to the discriminability of CVs.

The materials for the present experiment consist of sets of two or three voicings intended to represent the Cardinal Vowels [**i e ɛ a**] (front unrounded) [**ɯ ɤ ʌ ɑ**] (back unrounded), [**y ø œ ɶ**] (front rounded), [**u o ɔ ɒ**] (back rounded) and the high central unrounded [ɨ] and rounded [ʉ], and vowels near enough to each Cardinal for the tokens to be transcribed by the respective Cardinal, possibly with such IPA-approved diacritics as [⊥, ⊤, +].

No set of two or three contained more than just one representation of any Cardinal. For convenience of analysis the experimenter (WJ) avoided combining into one set such vowels as differ strongly in intrinsic intensity level (e.g. close and open or even half-open). The total material included 10 tokens of each of the 18 CVs (or near-CVs). They were spoken in a silent (but not sound-treated) room and the S/N ratio was controlled to enable the extraction of even the weakest formants (F_3 and F_4 of [**u**]). The analysis was performed using the KAY Elemetric DSP 5500 Workstation. A 512-point FFT analysis was performed of all the 180 voicings (10 replications of 18 VCs). The measurements were made at a moment approx. 15% into the vowel from its beginning, in the middle of the vowel, and approx. 15% into the vowel from its end. This demonstrated that what is intended as a steady vowel cannot normally be produced as a perfectly stationary acoustic event (except by synthesis). At each point, the frequency of each of the four formants was estimated from the 512-point FFT spectrum with an accuracy of 10 Hz. The frequency of a given formant was the arithmetic mean of the three measurements, and this mean was assumed to represent the entire vowel-token. Averaged formant frequencies for each token could of course have been obtained more directly by taking an average cumulative spectrum of the entire voicing, but we were interested in how the formant frequencies are permitted actually to vary in what is intended (and perceived) as an isolate stationary vowel sound. This temporal variation is of little consequence for the issue at hand, but will be taken into account in further experiments which we propose to make with synthetic stationary vowels. We may, however, just mention in passing that these variations were mostly of the order of 2...7%, though occasional higher values of the calculated coefficient of variation were not uncommon in the case of F_1 of close vowels in which there is the well-known interaction between F_1 and F_0. Approx. 2% of the time the temporal variation in the course of one formant was zero within the measurement accuracy. The fundamental frequency, held steady for each voicing, varied among the individual voicings within the range 97...105 Hz (a somewhat low male voice). The duration of the individual vowel tokens varied between approx. 200...300 ms. Altogether, then, 18 (CVs) × 10 Replications × 3 moments in time × 4 Formants = 2160 measurements were made.

The results of the measurements are summed up in Table 3. The coefficient of variation in Table 3 pertains to within-class variability, not to the temporal variability within individual tokens.

Table 3. Mean formant frequencies and their dispersions

vowel category forman front unrounded		mean F f-quency	st.dev.	var.coeff.
[i]				
	F_1	217	28.8	0.1326
	F_2	2775	115.3	0.0416
	F_3	3645	108.7	0.0298
	F_4	4107	80.1	0.0195
[e]				
	F_1	417	33.8	0.0811
	F_2	2538	135.4	0.0534

[cont. Tabl. 3]

1	2	3	4	5
	F_3	2944	94.0	0.0319
	F_4	3805	45.5	0.0119
[ɛ]				
	F_1	559	31.8	0.0569
	F_2	2151	123.7	0.0575
	F_3	2743	83.9	0.0306
	F_4	3689	160.5	0.0435
[a]				
	F_1	921	40.3	0.0438
	F_2	1560	83.4	0.0534
	F_3	2741	45.2	0.0165
	F_4	3571	41.2	0.0364
front rounded				
[y]				
	F_1	249	38.3	0.1528
	F_2	2255	152.0	0.0674
	F_3	2663	200.5	0.0753
	F_4	3494	118.2	0.0338
[ø]				
	F_1	422	18.2	0.0431
	F_2	1968	66.1	0.0336
	F_3	2479	79.3	0.0320
	F_4	3647	53.5	0.0147
[œ]				
	F_1	511	32.2	0.0630
	F_2	1769	93.3	0.0527
	F_3	2466	75.8	0.0308
	F_4	3586	51.8	0.0457
[ɶ]				
	F_1	676	78.3	0.1164
	F_2	1375	105.1	0.0765
	F_3	2616	186.8	0.0714
	F_4	3501	179.8	0.0514
back unrounded				
[ɯ]				
	F_1	369	72.5	0.1965
	F_2	808	113.8	0.1408
	F_3	2525	131.7	0.0522
	F_4	3240	217.9	0.0673
[ɤ]				
	F_1	475	30.7	0.0647
	F_2	846	38.5	0.0455
	F_3	2497	68.3	0.0273
	F_4	3193	125.3	0.0393

[cont. Tabl. 3]

1	2	3	4	5
[ʌ]				
	F_1	591	48.7	0.0824
	F_2	911	74.5	0.0818
	F_3	2794	160.5	0.0575
	F_4	3313	79.3	0.0239
[ɑ]				
	F_1	740	23.0	0.0312
	F_2	995	58.8	0.0591
	F_3	2982	66.3	0.0222
	F_4	3538	214.2	0.0605
back rounded				
[u]				
	F_1	308	30.2	0.0982
	F_2	577	39.7	0.0688
	F_3	2467	121.0	0.0490
	F_4	3133	265.4	0.0847
[o]				
	F_1	427	19.0	0.0445
	F_2	686	78.0	0.1165
	F_3	2583	48.4	0.0593
	F_4	3166	126.9	0.0401
[ɔ]				
	F_1	535	27.2	0.0508
	F_2	805	28.8	0.0375
	F_3	2660	174.2	0.0655
	F_4	3253	54.3	0.0167
[ɒ]				
	F_1	680	46.9	0.0691
	F_2	931	50.6	0.0544
	F_3	3007	107.4	0.0357
	F_4	3516	111.5	0.0317
central unrounded				
[ɨ]				
	F_1	309	26.3	0.0851
	F_2	1936	198.5	0.1025
	F_3	2594	123.7	0.0477
	F_4	3603	104.7	0.0291
central rounded				
[ʉ]				
	F_1	306	43.2	0.1413
	F_2	1004	95.7	0.0953
	F_3	2399	173.8	0.0725
	F_4	3400	164.7	0.0484

Figure 1 represents the (F_1, F_2) means of the 18 CVs as pronounced by WJ for the purposes of the experiment, each point being a two-element vector representing the grand means of F_1 and F_2 of the 18 CVs.

4. Statistical discriminant analysis

The four discriminant linear combinations $w_1 \ldots w_4$ have been calculated. Their values expressed in terms of $F_1 \ldots F_4$ are as follows:

$$w_1 = -0.0068672 * F_1 + 0.0094832 * F_2 - 0.00081388 * F_3 + 0.00096338 * F_4 \quad (1)$$

$$w_2 = -0.023943 * F_1 - 0.27703 * F_2 + 0.0011695 * F_3 - 0.00084285 * F_4 \quad (2)$$

$$w_3 = -0.0015498 * F_1 - 0.0015240 * F_2 + 0.0082357 * F_3 - 0.0045712 * F_4 \quad (3)$$

$$w_4 = 0.040886 * F_1 + 0.016195 * F_2 + 0.064314 * F_3 + 0.87861 * F_4. \quad (4)$$

The above exact relations between the values of the discriminant variables and the formant frequencies should be compared with the coefficients of determination between the discriminant variables and the formant frequencies:

Table 4. Coefficients of determination between the formant frequencies and the calculated discriminant variables

	w_1	w_2	w_3	w_4
F_1	7.8	90.9	0.8	0.5
F_2	89.5	7.2	0.9	2.4
F_3	1.8	1.1	97.0	0.1
F_4	4.1	1.6	6.4	87.9

It transpires from equations (1)...(4) and the Table 4 that

The first discriminant variable w_1 *depends chiefly on* F_2, w_2 *on* F_1, w_3 *on* F_3 *and* w_4 *on* F_4.

The dependence of the discriminant variables on the formants may most simply be demonstrated by the following Table:

Table 5.

$w_1 \longrightarrow F_2 \rightarrow F_1 \rightarrow F_4 \rightarrow F_3$

$w_2 \longrightarrow F_1 \rightarrow F_2 \rightarrow F_4 \rightarrow F_3$

$w_3 \longrightarrow F_3 \rightarrow F_4 \rightarrow F_2 \rightarrow F_1$

$w_4 \longrightarrow F_4 \rightarrow F_2 \rightarrow F_1 \rightarrow F_3$

In Table 5, the F–frequencies are, in each row, arranged in order of decreasing effect.

We shall refrain, in this study, from presenting the mathematical foundations of Discriminant Analysis, as these may be found in various statistical texts such as LACHENBRUCH (1975) [16] KLECKA (1980) [13] or KRZYŚKO (1990) [14].

The discriminant variables are obtained, as can be seen above, as *linear combinations* of the original variables, so that each discriminant variable represents in varying degrees, each of the original variables. The discriminant variables are so calculated that they are uncorrelated and all their covariance matrices are unit matrices. As demonstrated above, it is usually he case that one discriminant variable (w_n) reflects one particular original variable more strongly than others.

Table 5 shows, in general terms, how the discriminant variables depend, in decreasing order, on the individual formant frequencies. The main advantage of the w_n's is that they make it possible to map the mean vectors of the classes under consideration *in a plane* though the original data vectors lie in a multidimensional space (four-dimensional in our case) since each object under observation has here been represented by measurements on four variables: F_1, F_2, F_3, and F_4.

The Mahalanobis distances were transformed into respective values of the T^2 statistic. The significance of the individual values of T^2 was verified using the method of simultaneous test procedure. The common critical value for T^2 in our case was$T^2_{(0.05)} = 112.14$.

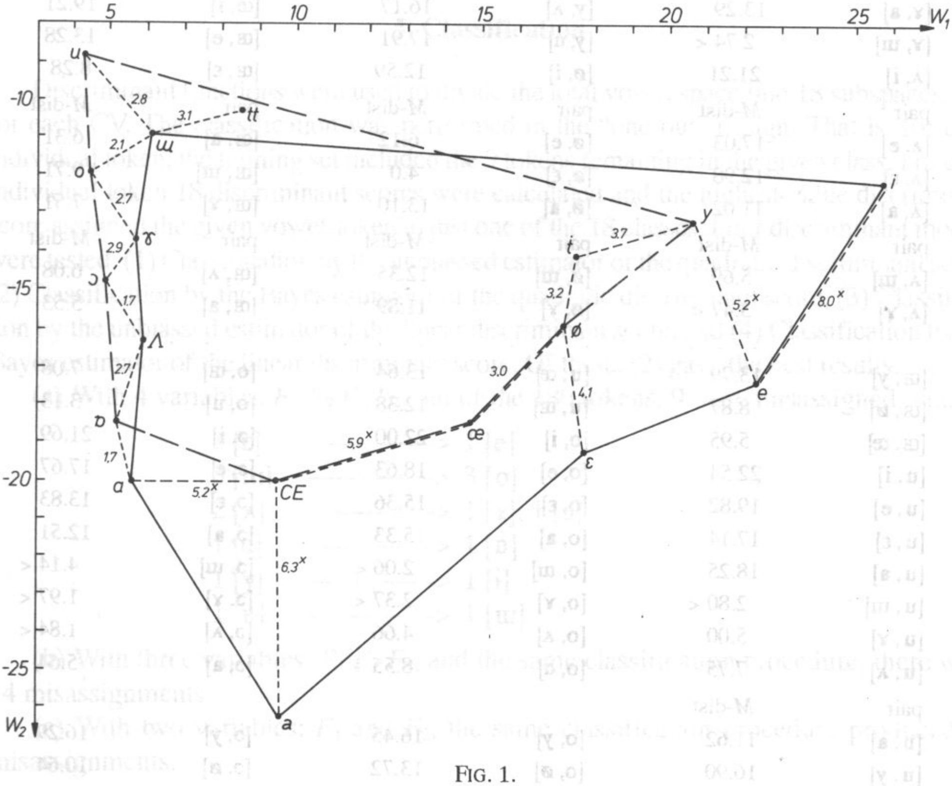

FIG. 1.

Out of the 153 distances between the mean vectors 14 were below the critical T^2 value, i.e., the 14 distances were not significantly different from 0. In Table 6 we give all the Mahalanobis distances ordered in columns rather than in a matrix arrangent in order to facilite the look-up. Those distances that at $\alpha = 0.05$ are not significantly different from 0 are marked <.

Table 6. Mahalanobis distances between the mean vectors

pair	*M*-dist	pair	*M*-dist	pair	*M*-dist
[e, i]	8.05	[ɑ, i]	22.18	[ø, ʌ]	12.52
[ɛ, i]	12.68	[ɑ, e]	17.59	[ø, ɑ]	13.67
[ɛ, e]	5.36	[ɑ, ɛ]	13.05	[ø, y]	5.65
[a, i]	22.38	[ɑ, a]	8.24	[œ, i]	14.86
[a, e]	15.91	[ɑ, ɯ]	9.58	[œ, e]	8.31
[a, ɛ]	10.75	[ɑ, ɤ]	7.30	[œ, ɛ]	4.10 <
[ɯ, i]	20.57	[ɑ, ʌ]	4.01 <	[œ, a]	10.44
[ɯ, e]	17.33	[y, i]	8.75	[œ, ɯ]	10.93
[ɯ, ɛ]	14.41	[y, e]	5.29	[œ, ɤ]	9.91
[ɯ, a]	15.73	[y, ɛ]	7.87	[œ, ʌ]	10.15
[ɤ, i]	21.10	[y, a]	18.07	[œ, ɑ]	10.10
[ɤ, e]	17.02	[y, ɯ]	14.81	[œ, y]	8.37
[ɤ, ɛ]	13.39	[y, ɤ]	15.16	[œ, ø]	2.99 <
[ɤ, a]	13.29	[y, ʌ]	16.17	[œ, i]	19.21
[ɤ, ɯ]	2.74 <	[y, ɑ]	17.91	[ɶ, e]	13.28
[ʌ, i]	21.21	[ø, i]	12.59	[ɶ, ɛ]	8.28
pair	*M*-dist	pair	*M*-dist	pair	*M*-dist
[ʌ, e]	17.03	[ø, e]	6.12	[ɶ, a]	6.31
[ʌ, ɛ]	12.98	[ø, ɛ]	4.0	[ɶ, ɯ]	9.71
[ʌ, a]	11.02	[ø, a]	13.10	[ɶ, ɤ]	7.51
pair	*M*-dist	pair	*M*-dist	pair	*M*-dist
[ʌ, ɯ]	5.69	[ø, ɯ]	12.35	[ɶ, ʌ]	6.08
[ʌ, ɤ]	3.47 <	[ø, ɤ]	11.89	[ɶ, a]	5.53
	. . .	. . .	. . .	. . .	
[ɶ, y]	13.79	[u, œ]	13.64	[o, ɶ]	7.08
[ɶ, ø]	8.87	[u, ɶ]	12.38	[o, u]	3.13
[ɶ, œ]	5.95	[o, i]	22.00	[ɔ, i]	21.69
[u, i]	22.54	[o, e]	18.63	[ɔ, e]	17.67
[u, e]	19.82	[o, ɛ]	15.36	[ɔ, ɛ]	13.83
[u, ɛ]	17.14	[o, a]	15.33	[ɔ, a]	12.51
[u, a]	18.25	[o, ɯ]	2.06 <	[ɔ, ɯ]	4.14 <
[u, ɯ]	2.80 <	[o, ɤ]	2.37 <	[ɔ, ɤ]	1.97 <
[u, ɤ]	5.00	[o, ʌ]	4.66	[ɔ, ʌ]	1.84 <
[u, ʌ]	7.75	[o, ɑ]	8.55	[ɔ, a]	5.64
pair	*M*-dist				
[u, a]	11.62	[o, y]	16.45	[ɔ, y]	16.29
[u, y]	16.90	[o, ø]	13.72	[ɔ, ø]	10.64

[cont. Tabl. 6]

[u, ø]	14.89	[o, œ]	11.98	[ɔ, œ]	7.23
[ɔ, ɶ]	7.23	[a, ɔ]	4.38 <	[ʉ, i]	18.71
[ɔ, u]	6.03	[ɨ, i]	11.01	[ʉ, e]	15.40
[ɔ, o]	2.94 <	[ɨ, e]	6.56	[ʉ, ɛ]	12.94
[ɒ, i]	21.82	[ɨ, ɛ]	6.51	[ʉ, a]	16.13
[ɒ, e]	17.67	[ɨ, a]	15.61	[ʉ, ɯ]	3.13 <
[ɒ, ɛ]	13.42	[ɨ, ɯ]	11.63	[ʉ, ɤ]	4.76
[ɒ, a]	9.81	[ɨ, ɤ]	11.86	[ʉ, ʌ]	7.64
[ɒ, ɯ]	8.19	[ɨ, ʌ]	12.93	[ʉ, a]	11.16
[ɒ, ɤ]	6.13	[ɨ, a]	14.76	[ʉ, y]	12.48
[ɒ, ʌ]	2.75 <	[ɨ, y]	3.67 <	[ʉ, ø]	10.24
[ɒ, ɑ]	1.73 <	[ɨ, ø]	3.22 <	[ʉ, œ]	9.36
[ɒ, y]	17.55	[ɨ, œ]	5.47	[ʉ, ɶ]	9.84
[ɒ, ø]	13.64	[ɨ, ɶ]	10.70	[ʉ, u]	4.94
[ɒ, œ]	11.12	[ɨ, u]	13.90	[ʉ, o]	5.14
[ɒ, ɶ]	6.19	[ɨ, o]	13.26	[ʉ, ɔ]	6.45
[ɒ, u]	10.12	[ɨ, ɔ]	12.99	[ʉ, a]	10.01
[ɒ, o]	7.09	[ɨ, ɒ]	14.36	[ʉ, ɨ]	9.26

5. Classification

Discriminant functions were used to divide the total vowel space into 18 subspaces, one for each CV. The classification was performed in the "one out" design. That is, for each individual token, the training set included the 9 tokens remaining in the given class. For each individual token 18 discriminant scores were calculated and the highest-value discriminant score assigned the given vowel-token to just one of the 18 classes. Four discriminant models were tested: (1) Classification by the unbiassed estimator of the quadratic discriminant score, (2) Classification by the Bayes estimator of the quadratic discriminant score, (3) Classification by the unbiassed estimator of the linear discriminant score, and (4) Classification by the Bayes estimator of the linear discriminant score. Of these, (2) gave the best results:

(a) With 4 variables, $F_1 F_2 F_3 F_4$, out of the 180 tokens, 9 were misassigned, viz.:

1 [ɛ] ————> 1 [e]
3 [ɯ] ————> 3 [o]
2 [ʌ] ————> 1 [ɤ], 1 [ɔ]
1 [ɑ] ————> 1 [ɒ]
1 [y] ————> 1 [ɨ]
1 [ʉ] ————> 1 [ɯ]

(b) With three variables: $F_1 F_2 F_3$, and the same classification procedure, there were 14 misassignments.

(c) With two variables: F_1 and F_2, the same classification procedure produced 19 misassignments.

6. Conclusions

The total 4–D vowel space can be divided into 18 subspaces representing the 18 Cardinal Vowels such that only about 5% of individual isolate voicings are erroneously assigned. In a 3–D space the error is about 8%. With only two variables, F_1 and F_2, the error is increased to almost 10%. So, in order to procure fewer mistakes, especially with fewer variables, such as F_1 and F_2, the statistical dispersions would have to be distinctly smaller than in the present materials.

References

[1] W.A. AINSWORTH, *Intrinsic and extrinsic factors in vowel judgement* in: Auditory analysis and perception of speech, [Eds.] G. Fant, M.A.A. Tatham, Academic Press, London 1975, 103–116.
[2] L. ARMSTRONG, D. JONES, *The phonetics of French*, G. Bell and Sons Ltd., London 1951.
[3] A. BUTCHER, *Cardinal Vowels and other problems*, in: Linguistic controversies, [Ed.] D. Crystal, E. Arnold, London 1982, 50–71.
[4] J.C. CATFORD, *Observations on the recent history on vowel classification*, in: Towards a history of phonetics. [Eds.] R.E. Asher and E.J. Henderson, Edinburgh University Press, Edinburgh 1981, 19–31.
[5] J.C. CATFORD, *A practical introduction to phonetics*, The Clarendon Press, Oxford 1988.
[6] P. DELATTRE, A.M. LIBERMAN, F.S.COOPER, *Voyelles systhétique a deux formantes et voyelles cardinales.* Le Maitre Phonetique, ser. 3, No. 96, 30–35 1951.
[7] *Fundamentos de escritura fonetica*, [Eds.] D. Jones and I. Dahl, Secretaria de la Asociacion Fonetics Internacional, London 1944.
[8] A.C. GIMSON, *The introduction to the pronounciation of English*, 4th ed. rev. S. Ramsaran, E. Arnold, London 1988.
[9] W. JASSEM, *Fundamentals of acoustic phonetics* (in Polish), PWN, Warszawa 1973.
[10] W. JASSEM, *The phonology of modern English*, PWN, Warszawa 1984.
[11] D. JONES, *An outline of English phonetics*, 8th ed. Heffner, Cambridge 1956.
[12] D. JONES, D. WARD, *The phonetics of Russian*, Cambridge University Press, Cambridge 1969.
[13] W. KLECKA, *Discriminant analysis*, Sage Publications, Beverly Hills 1980.
[14] M. KRZYSKO, *Discriminant analysis* (in Polish) 2nd ed. WNT, Warszawa 1990.
[15] W. LABOV, *Sources of inherent variation in the speech process*, in: Invariance and variability in speech processes, [Eds.] J. Perkell and D. Klatt, L. Erblaum, Hillside 1986, 402–423.
[16] P.A. LACHENBRUCH, *Discriminant analysis*, Hafner, New York 1975.
[17] P. LADEFOGED, D.E. BROADENT, *Information conveyed by vowels*, Journ. Acoust. Soc. Am., **9**, 98–104 1957.
[18] H.B. LEE, *A remark on the IPA Cardinal Vowels*. Le Maitre Phonetique III/130, 26–29 (1968).
[19] M. NEAREY, *Phonetic feature systems for vowels*, Indiana Univ. Linguistic Club, Bloomington, Indiana 1978.
[20] *The principles of the International Phonetic Association*, International Phonetic Association, London 1949.
[21] G. PAPCUN, *How do different speakers say the same vowels?* UCLA Working Papers in Phonetics No. 49, Los Angeles 1980.
[22] R. PLOMP, C.W. POLS, VAN DE GEER, *Dimensional analysis of vowel spectra*, Journ. Acoust. Soc. of America **41**, 707–712 (1967).
[23] P. ROACH, *English phonetics and phonology*, Cambridge University Press, Cambridge 1953.
[24] A.Q. SUMMERFIELD, M.P. HAGGARD, *Vocal tract normalization as demonstrated by reaction times*, in: Auditory analysis and perception of speech, [Eds.], G. Fant, M.A.A.A. Tatham, Academic Press, London 1975, 115–141.

Received December 31, 1991