CONSERVATION OF VOWEL CONTRAST IN VARIOUS SPEECH CONDITIONS*

J.S.C. van Dijk

INTRODUCTION

In her thesis Koopmans-van Beinum (1980) gives plots in the F1 - F2 plane of the vowel systems of four speakers in eight different speech conditions. At first glance these figures lead to the global impression that internal relations between vowels in the vowel system of a specific speaker are independent of the speech conditions. In other words, it appears that vowel systems of the same speaker but in different speech conditions are similar to each other to a certain degree. If this is true it must be possible to make vowel systems of different speech conditions comparable with each other by application of an appropriate scaling technique. In this article a simple scaling technique is presented which follows from main properties of vocal tract model studies. Using the afore-mentioned formant frequency data (Koopmans-van Beinum, 1980) it will be shown that a first glance tends to love at the first sight.

Acoustical properties of vowels have been studied thoroughly from different points of view. More or less sophisticated vocal tract model studies, for instance Dunn (1950), Fant (1960), Ungeheuer (1962), Flanagan (1972), Bonder (1983), show that it is sensible to identify a resonance frequency of the vocal tract with a formant frequency of the speech signal. Moreover, theory shows that these resonances always follow from equations of the kind

$$(1) \quad f\left(\frac{Fl}{c}\right) = 0,$$

in which the function f depends on the geometry of the vocal tract and the boundary conditions at both the glottis and the opening of the mouth. l is the length of the vocal tract and c the velocity of sound in air. F stands for the frequency.

---

* This is an extended version of a paper read at the 10th International Congress of Phonetic Sciences, Utrecht, August 1983.

Resonances are found by solving eq. (1). Let us denote subsequent solutions of this equation with A_j, where j = 1,2, ... Then it is obvious that a resonance frequency takes the shape

$$(2) \quad F_j = A_j \frac{c}{l} \quad ; \quad j = 1, 2, ...$$

Hereafter we shall assume that during the production of vowellike utterances the boundary conditions do not change dramatically. That means, the symbols A_j only depend on the (ordinal) number j and the geometry of the vocal tract. Ungeheuer (1962), Vieregge (1970) and Bonder (1983) for instance, showed that a change of the vocal tract shape (in general) leads to a shift of the resonance frequencies (2). So, if c and l are fixed quantities articulatory differences between vowels are reflected in different values of A_j.
Ungeheuer models the actual vocal tract as a perturbation of the straight tube. In his frame of mind resonance frequencies of a vocal tract are disturbances of corresponding 'straight tube' resonances. This type of model shows that greater weighting of a disturbance can be accorded to lower resonance frequencies. Similar conclusion have been drawn by Paige and Zue (1970). As a consequence it stands to reason to characterize a vowel with the ordered n-tuple (3) consisting of the first n resonance frequencies of the vocal tract.

$$(3) \quad \text{Vowel} = \{ F_1, F_2, ... , F_n \}.$$

By doing this we meet remarks made by Lord Rayleigh at the end of the past century. For, Rayleigh (1894) stated in his 'Theory of Sound': "it is not unlikely that the complete characterization of a vowel is of multiple nature". It was the same Lord who pointed out ideas of the phonetician Lloyd (1890) about the identity of a vowel. In Lloyd's opinion "the identity of a given vowel depends not upon the absolute frequency* of one or more resonances, but upon the relative frequency of two or more". In this manner Lloyd

* In Lloyd's original work pitch was used instead of frequency.

solves the striking problem that the articulation of a given vowel appears to be the same for an infant and for a grown man, (Rayleigh, 1894).

VOWEL INDICES

In my opinion it is worth-while to follow Lloyd's original proposal. Therefore we introduce the dimensionless quantities

$$(4) \quad I_j = \frac{F_j - F_j^r}{F_j^r} \quad ; \quad j = 1, 2, \ldots\ldots ,$$

which are relative deviations of formant frequencies with respect to a given reference vowel V^r. This point of reference is defined by the n-tuple

$$(5) \quad V^r = \{ F_1^r , F_2^r , \ldots\ldots\ldots F_n^r \} ,$$

where the symbols F_j^r ; $j = 1, 2, \ldots.. , n$ stand for formant frequencies which are unknown at this stage. For the sake of convenience I shall call the numbers (4) the vowel indices. If the point of reference belongs to the vowel system of a speaker, the indices (4) are roughly independent of the factor c/l. For, substitution of (2) in (4) yields

$$I_j = \frac{A_j - A_j^r}{A_j^r} \quad ; \quad j = 1, 2, \ldots.. .$$

Here the symbols A_j^r ; $j = 1, 2, \ldots.$ are solutions of eq. (1) which belongs to the vocal tract of the reference vowel. In our applications it appears that the shape of this vocal tract is not far from the straight tube. So, the indices approximately measure the relative perturbation of the vocal tract shape with respect to the shape of the straight tube. In this manner Lloyd's proposal has been combined with the vocal tract model studies of Ungeheuer.

COMPARISON OF DATA

In the next section we shall transform formant frequency data to vowel indices. The data are formant frequencies of the twelve Dutch vowels in eight different speech conditions (appendix A, table 1). The number of speakers is four. The conditions are:

1. vowels spoken in isolation.
2. vowels spoken in isolated monosyllabic words.
3. vowels in stressed position in a text read aloud.
4. vowels in unstressed position in a text read aloud.
5. vowels in stressed position in a retold story.
6. vowels in unstressed position in a retold story.
7. vowels in stressed position in normal conversation.
8. vowels in unstressed position in normal conversation.

Plots of these data in the F_1 - F_2 plane are found in Koopmans-van Beinum (1980). These plots show that in the (plane) formant frequency space the vowel system of a speaker performs a motion as a function of the speech condition. While travelling in that space the system conveys its own centre at the same time contracting or dilating its volume. Before comparing vowel systems of different speech conditions with each other it seems to be sensible to include these properties in the data. I shall do this in two steps. The first step consists in defining a coordinate system which moves- parallel to the formant frequency axes - with the centre of the vowel system in the formant frequency space. This property of convection is most easily included if we identify in the indices (4) the vowel of reference (5) with the centre of the vowel system. Because this step is only a translation (in the mathematical sense), the volume of the vowel system with respect to the new coordinates - the indices (4) - has been conserved. Scaling of this volume per speech condition to a unit value is the second step.

OUTLINE OF THE CALCULATIONS

When defining the reference vowel (5) several possibilities occur. One of these is to consider V^r as an appropriately chosen working point, which has characteristics of a centre point of a vowel system. This can be accomplished by application of the general formula for the mean (Abramowitz, 1964) to the formant frequencies with the same (ordinal) number over the whole vowel system of a speaker per speech condition. The formula reads

$$(6) \quad F_j^r = \{ \frac{1}{N} \sum_{i=1}^{N} (F_j^i)^p \}^{\frac{1}{p}} \quad , \quad -\infty < p < +\infty \quad ,$$

where F_j^i denotes the j-th formant frequency of the i-th vowel of a system consisting of N vowels. In the applications I selected in (6) the value $p = 0$. It is well-known that in this limiting case formula (6) defines the geometric mean (Abramowitz, 1964). At this stage the point of reference was calculated for the vowel system of a speaker per speech condition (appendix A, table 2). Vowel indices were calculated according to (4). From the sets of indices with the same (ordinal) number I determined the standard deviations (appendix A, table 2) and normalized the indices to this 'intrinsic' measure.

RESULTS

The next figures show per speaker the vowel systems of eight different speech conditions in one plot. The figures have been arranged in four sets of two. The first figure of each pair gives the formant frequency data in the F1 - F2 plane for one speaker. The second one shows the same data, but this time as normalized vowel indices in the I1 - I2 plane.

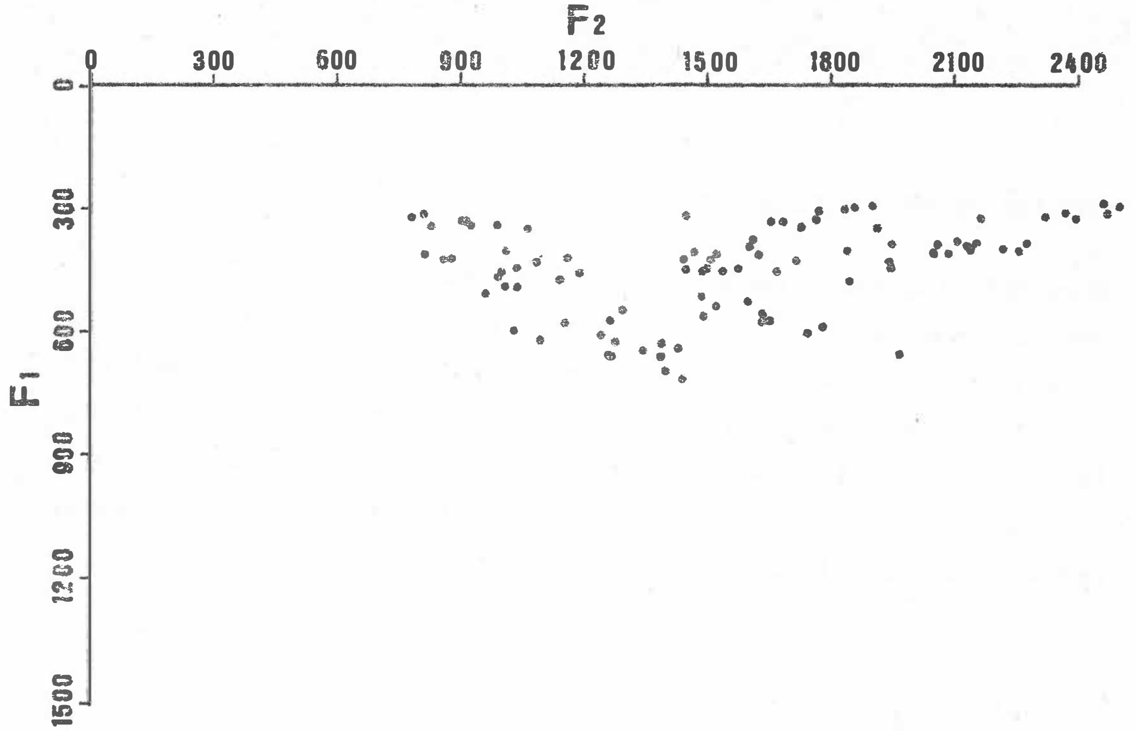

Fig. 1. Plot of the vowel frequencies of twelve Dutch vowels in eight different speech conditions. Speaker 1 (male, trained).

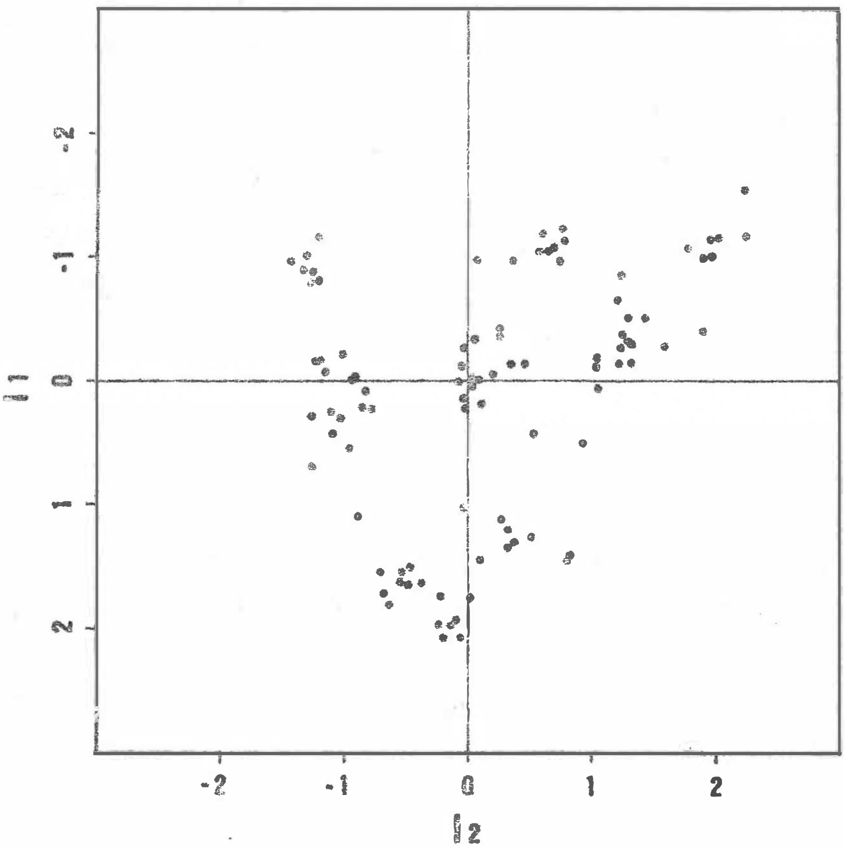

Fig. 2. Plot of the same data as in Fig. 1 in the plane of the scaled vowel indices. Speaker 1.

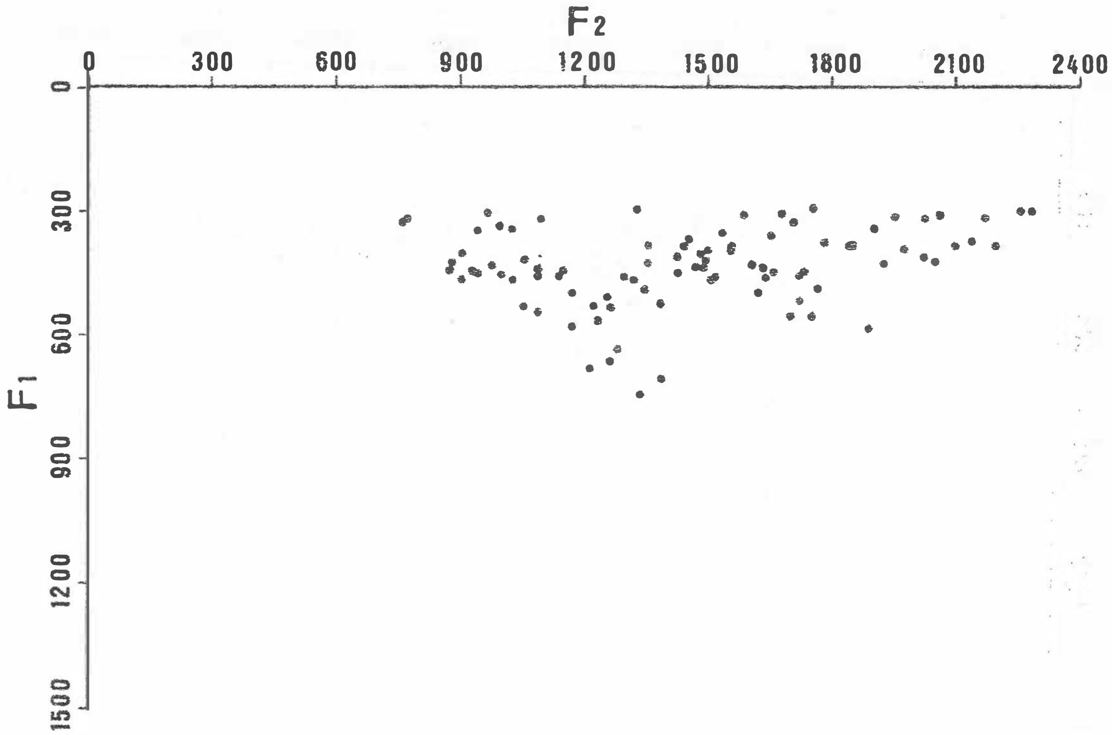

Fig. 3. Plot of the vowel frequencies of twelve Dutch vowels in eight different speech conditions. Speaker 2 (male, untrained).

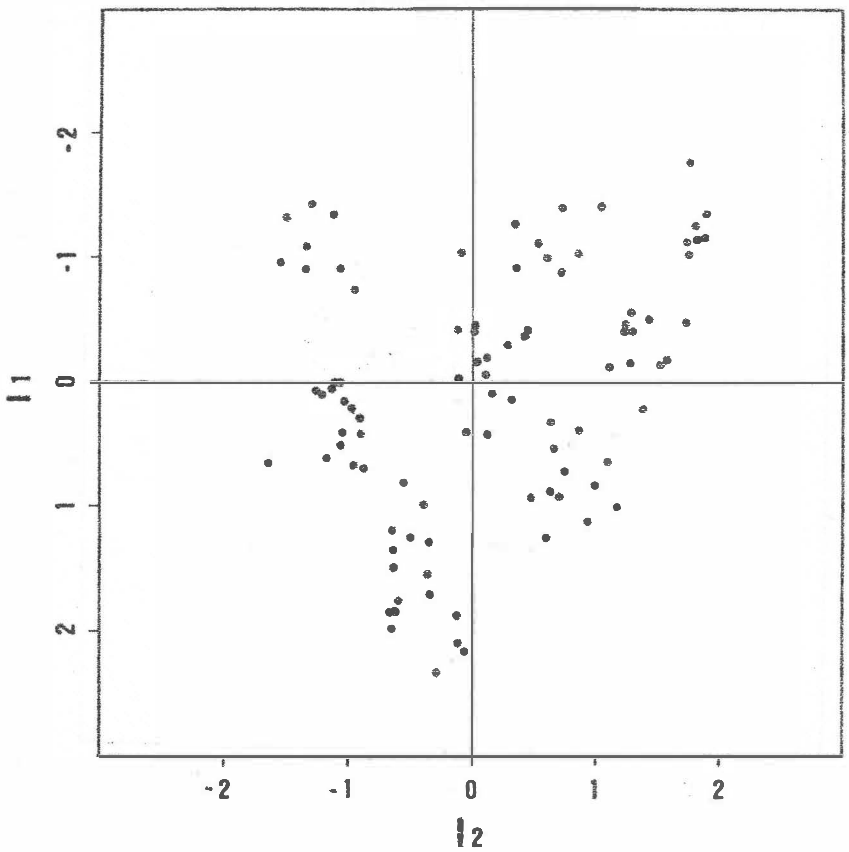

Fig. 4. The same data as in Fig. 3 in the scaled vowel indices plane. Speaker 2.

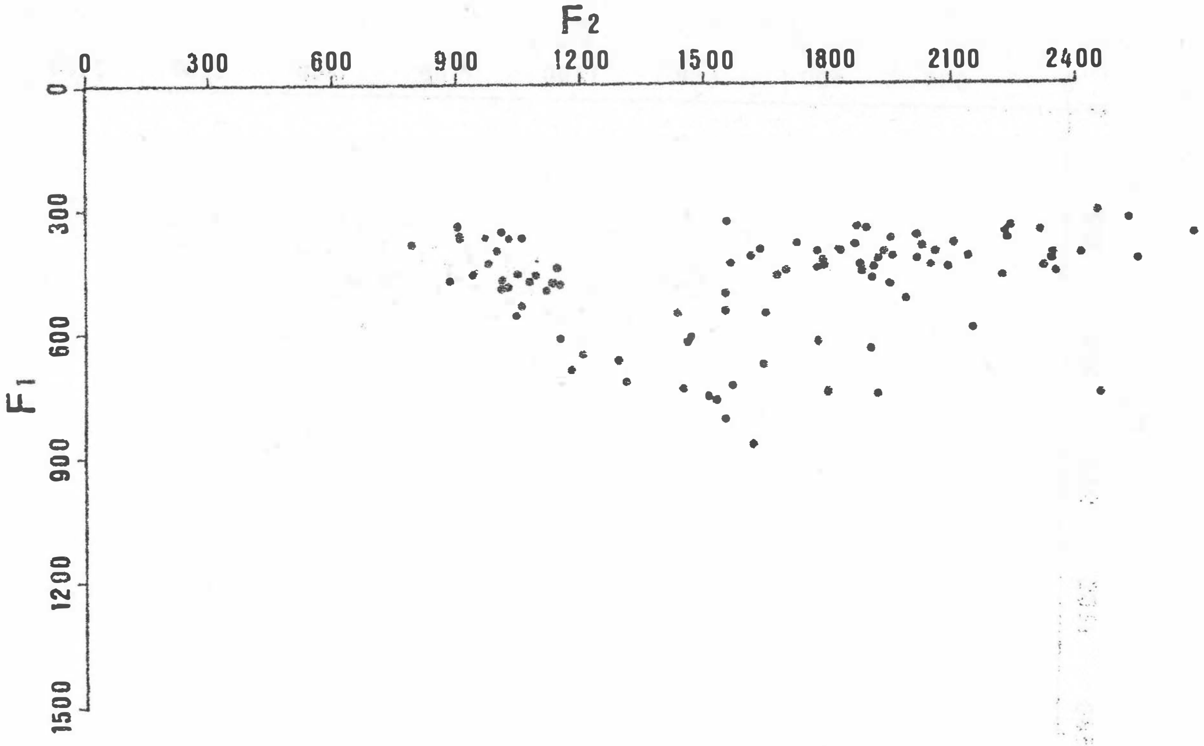

Fig. 5. Plot of the vowel frequencies of twelve Dutch vowels pronounced by speaker 6 (female, trained) in eight different speech conditions.

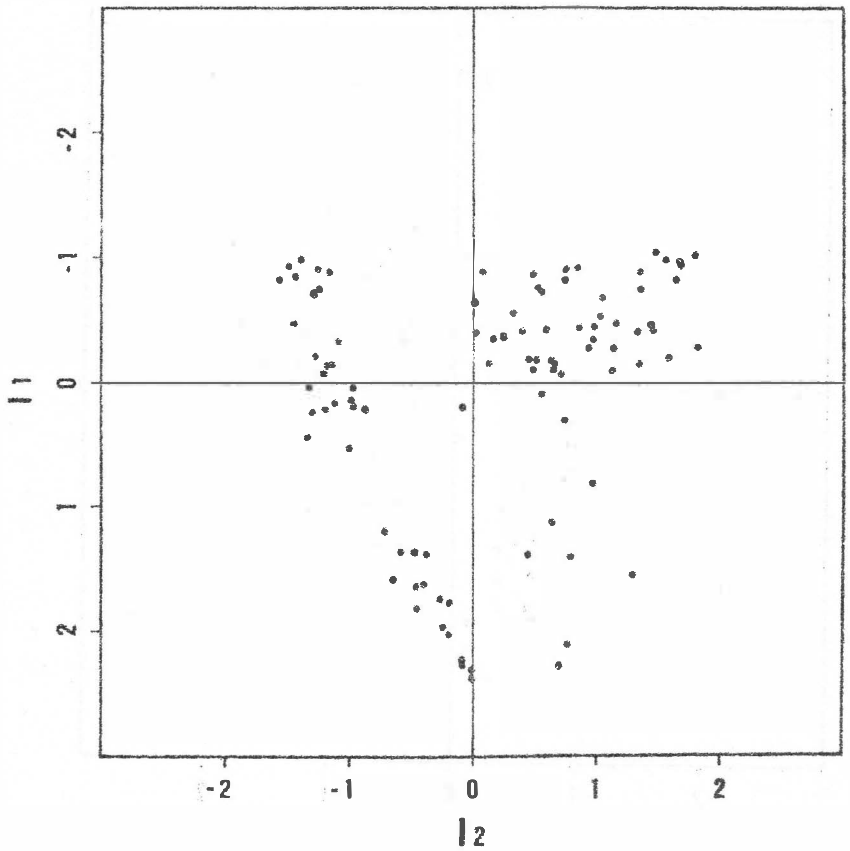

Fig. 6. The data from Fig. 5 expressed in terms of scaled vowel indices. Speaker 6.

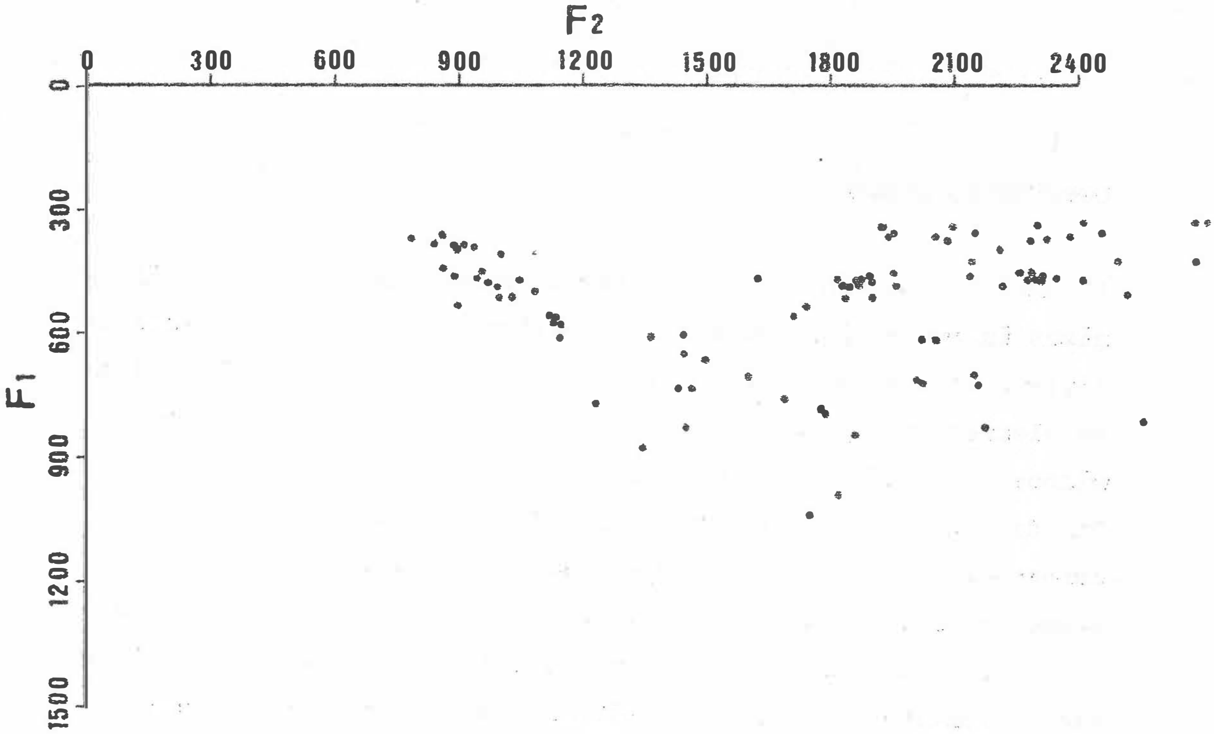

Fig. 7. Plot of the vowel frequencies of twelve Dutch vowels pronounced by speaker 9 (female, untrained) in eight different speech conditions.

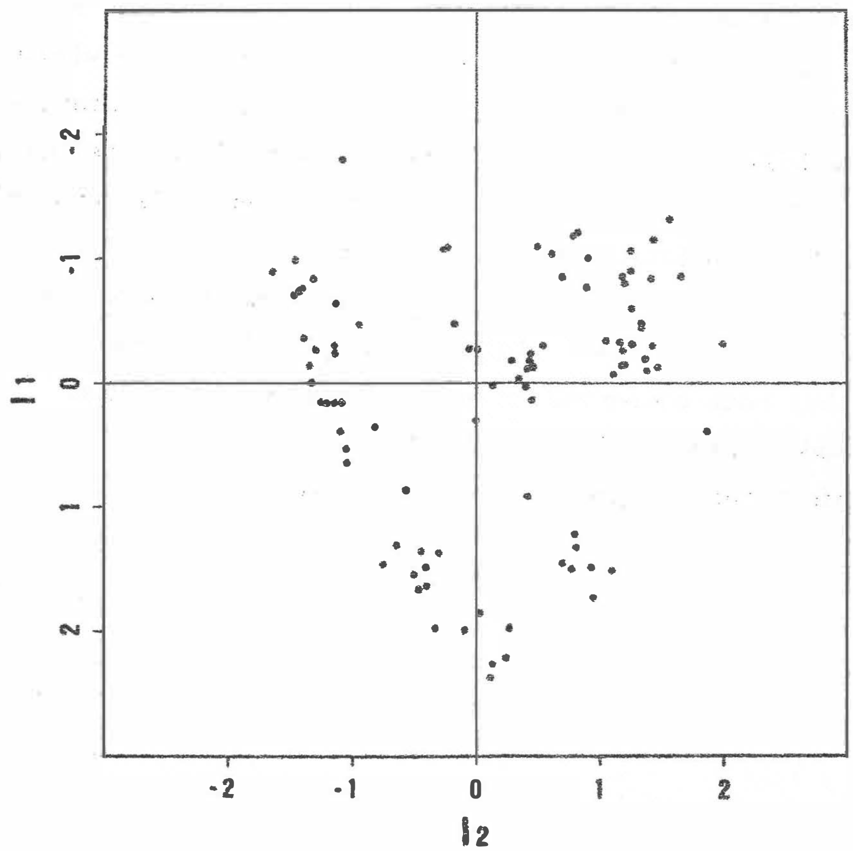

Fig. 8. The data from Fig. 7 in terms of scaled vowel indices. Speaker 9.

CONCLUDING REMARKS

A vowel defined in terms of formant frequencies or vowel indices gives in an implicit manner information about the act of articulation. Because we hunt for regularity in articulatory behaviour we plotted only points in both the F1 - F2 and I1 - I2 plane without assigning a (vowel)name to any point.
The distributions of points in the F1 - F2 planes seem to be rather diffuse. However, a distribution of points in an I1 - I2 plane tends to be a collection of clusters. From a statistical point of view a distribution in the index plane is a sample drawn from a continuous population with a density function which looks like a landscape with well defined hills. Every hill represents the preference of a speaker to a rather specific act of articulation. We shall call this region of preference a vowel. When scaling back the landscape of one speaker per speech condition to the original values of the vowel indices a set of nearly similar landscapes results. As a common property we find that - within a certain statistical margin - the mutual relations of distances between the hills have been conserved. So, I tend to consider the notion of 'vowel contrast' of a speaker as an ensemble consisting of preferent articulation movements which - after a period of development - are independent of growth and speech condition. In this sense 'vowel contrast' is not reduced (Koopmans-van Beinum, 1980) but has been conserved in various speech conditions.
In an extended version of this article these results will be discussed against the background of results from literature.

APPENDIX A

Speech condition		1		2		3		4		5		6		7		8	
		F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
sp. 1	u	317	783	308	809	342	827	342	924	329	914	347	1062	326	903	338	989
	o	421	861	407	817	421	877	401	1012	469	1140	452	1189	429	1079	417	1162
	o	600	1030	506	959	440	1035	454	1002	490	1007	489	1039	461	994	443	1045
	α	658	1260	657	1266	612	1093	597	1153	627	1274	572	1262	609	1238	549	1296
	a	692	1398	717	1437	649	1344	631	1388	661	1383	514	1488	642	1429	566	1490
	y	292	1902	296	1861	306	1774	301	1836	332	1655	346	1731	321	1764	332	1687
	ø	425	1715	372	1614	415	1625	388	1604	404	1520	453	1491	454	1535	456	1674
	œ	443	1573	451	1446	424	1509	426	1442	447	1500	434	1515	405	1464	408	1467
	i	293	2503	292	2470	322	2450	323	2399	311	2475	320	2329	309	2370	323	2170
	I	403	2258	385	2156	387	2280	409	2087	386	2060	387	1954	379	2110	384	1918
	e	393	2220	392	2129	399	2143	409	2053	429	1945	475	1847	441	1948	403	1846
	ε	653	1965	605	1741	590	1780	559	1634	577	1633	522	1599	575	1649	540	1525
sp. 2	u	317	773	325	761	349	948	307	967	344	1026	319	1098	336	990	339	993
	o	443	875	454	942	427	886	420	1057	471	1028	459	1299	459	1135	448	1084
	o	468	907	533	1052	446	937	406	903	435	977	449	1003	458	1084	451	1147
	α	685	1217	638	1281	549	1096	504	1170	537	1225	489	1351	533	1225	469	1316
	a	750	1338	707	1385	665	1268	531	1384	589	1170	532	1270	572	1226	512	1255
	y	298	1760	307	1679	323	1715	312	1588	300	1759	350	1541	311	1680	364	1649
	ø	465	1522	459	1513	423	1361	430	1492	388	1355	380	1451	457	1421	437	1474
	œ	467	1515	438	1490	407	1434	372	1452	399	1501	382	1565	408	1478	393	1557
	i	297	2335	302	2264	296	2298	322	2174	323	2028	343	1915	315	2063	315	1952
	I	378	2143	394	1977	382	2207	388	2103	389	1846	372	1792	387	1849	386	1855
	e	425	2057	430	1928	409	2034	416	2022	526	1721	448	1742	491	1768	444	1633
	ε	588	1897	560	1754	554	1710	465	1639	503	1622	428	1618	456	1726	454	1659
sp. 6	u	383	792	370	910	353	903	366	899	371	963	357	1006	369	1056	371	1015
	o	473	885	459	939	486	1009	436	973	476	1068	486	1147	481	1130	404	990
	o	537	1060	562	1044	507	1006	495	1017	458	1042	464	1092	499	1113	448	1135
	α	723	1315	743	1447	673	1204	622	1149	758	1505	630	1455	612	1462	554	1542
	a	877	1620	817	1549	704	1177	672	1288	766	1523	736	1564	682	1640	557	1425
	y	398	1693	349	1876	350	1555	384	1721	373	1953	349	1893	388	1864	375	2014
	ø	432	1883	420	1617	453	1787	454	1694	472	1909	451	1878	445	1909	425	1914
	œ	437	1793	511	1550	463	1777	434	1560	465	1674	408	1829	405	1775	443	1876
	i	363	2697	334	2534	330	2460	385	2109	351	2319	344	2246	358	2234	378	2235
	I	423	2560	414	2420	414	2034	419	2017	439	2053	406	1936	407	2061	391	2100
	e	453	2360	417	2352	448	2349	442	2325	465	2226	424	1961	447	2095	419	2139
	ε	715	2407	601	2154	776	1919	750	1798	643	1903	629	1772	486	1947	530	1982
sp. 9	u	362	860	365	787	386	837	381	881	394	930	385	915	393	889	409	998
	o	443	860	465	944	462	889	468	1043	450	948	476	971	514	1023	486	988
	o	533	900	613	1149	585	1145	559	1135	579	1125	560	1125	514	997	500	1085
	α	773	1233	880	1346	830	1450	729	1463	737	1424	651	1452	609	1359	667	1479
	a	1000	1825	1040	1758	850	1862	776	1777	764	1686	705	1609	793	1784	610	1442
	y	333	2417	340	2101	357	1955	362	1940	344	1923	362	2068	354	2153	374	2087
	ø	433	2150	469	1429	486	1959	471	1857	458	1947	470	1884	474	1902	488	1872
	œ	562	1720	537	1746	518	1840	489	1849	475	1815	486	1841	512	1903	471	1896
	i	318	2810	339	2700	334	2723	394	2213	349	2305	366	2397	376	2288	364	2465
	I	432	2700	510	2533	472	2305	458	2142	479	2280	482	2233	369	2324	472	2283
	e	395	2757	432	2508	477	2315	464	2316	482	2417	469	2364	456	2288	455	2264
	ε	818	2567	830	2178	727	2027	710	2016	711	2148	621	2068	729	2162	618	2029

Table 1. The formant frequency data from Koopmans-van Beinum (1980). Units are Hz. Speaker 1 is a trained male speaker and speaker 2 male, untrained. Speaker 6 is female and trained whereas speaker 9 is female, untrained.

Sp. 1	F_1^r	F_2^r	D_1	D_2
1	445	1522	0.31	0.36
2	429	1466	0.32	0.36
3	430	1407	0.26	0.32
4	424	1476	0.23	0.31
5	442	1483	0.25	0.29
6	436	1500	0.17	0.24
7	434	1478	0.24	0.29
8	419	1482	0.19	0.23

Sp. 2	F_1^r	F_2^r	D_1	D_2
1	446	1374	0.31	0.32
2	446	1436	0.27	0.30
3	425	1415	0.24	0.34
4	400	1438	0.17	0.29
5	425	1397	0.20	0.24
6	408	1445	0.15	0.19
7	424	1433	0.19	0.23
8	413	1434	0.13	0.20

Sp. 6	F_1^r	F_2^r	D_1	D_2
1	500	1630	0.33	0.40
2	480	1604	0.31	0.35
3	478	1513	0.30	0.34
4	476	1475	0.25	0.31
5	486	1612	0.28	0.27
6	472	1625	0.26	0.24
7	457	1640	0.21	0.24
8	437	1633	0.15	0.27

Sp. 9	F_1^r	F_2^r	D_1	D_2
1	499	1730	0.42	0.44
2	531	1664	0.42	0.38
3	516	1672	0.32	0.34
4	505	1647	0.26	0.28
5	519	1746	0.28	0.30
6	492	1660	0.22	0.30
7	491	1664	0.27	0.32
8	485	1661	0.19	0.31

Table 2. Scaling data of the four speakers per speech condition. The first colomn of each quadrant gives the number of the speech condition. The second and third colomns contain the coordinates of the reference vowel (5). The numbers in the fourth and fifth colomns are the scaling factors in the direction of the vowel indices.

REFERENCES

Abramowitz, M. & Stegun, I. A. (1964). Handbook of mathematical functions. National Bureau of Standards, Washington.

Bonder, L.J. (1983). The n-tube formula and some of its consequences. Acustica 52, 216-226.

Bonder, L.J. (1983). Changing the shape of lossless n-tubes. Acustica 53, 201-211.

Dunn, H.K. (1950). The calculation of vowel resonances and an electrical vocal tract. J. Acoust. Soc. Am., 22, 740-753.

Fant, G. (1960). Acoustic theory of speech production. Mouton & Co., The Hague.

Flanagan, J.L. (1972). Speech analysis synthesis and perception. Second expanded edition. Springer-Verlag, Berlin.

Koopmans-van Beinum, F.J. (1980). Vowel contrast reduction: An acoustic and perceptual study of Dutch vowels in various speech conditions. Diss., University of Amsterdam.

Lloyd, (1890). Speech sounds: their nature and causation. Phonetische Studie, III, part J.

Paige, A. & Zue, V.W. (1970). Computation of vocal tract area functions. IEEE Trans. AU - 18, 7-18.

Rayleigh, J.W.S. (1894). The theory of sound. Vol. 2. Dover Publications, New York (1945), 476-478.

Ungeheuer, G. (1962). Elemente einer akustischen Theorie der Vokalartikulation. Springer-Verlag, Berlin.

Vieregge, W.H. (1970). Untersuchungen zur akustischen Artikulation der Plosivlaute. S. Karger, Berlin.