A PREDICTION METHOD FOR MODAL n-VOWEL SYSTEMS

L.J. Bonder

1 INTRODUCTION

A number of researchers has made attempts to discover so-called universals of vowel systems. One of the early discovered rules which vowel systems seem to obey is the principle of maximal contrast.
Liljencrants and Lindblom (1972) were the first to implement such a rule in a computational model for the prediction of vowel systems. Since then this approach has been improved by using more sophisticated models of speech and hearing. A more thorough discussion is given by Lindblom (1986).

In this paper we will show by means of a simple vocal tract model that the set of the modal n-vowel systems (a modal n-vowel system is the most frequent system of the collection of vowel systems that consist of n vowels) can be thought of as hierarchically ordered, with increasing n, in an acoustic-phonetic way. We have been able to construct an algorithm which reflects this hierarchy (paragraph 9). The algorithm is primarily based on the idea that a modal (n+1)-vowel system can be constructed **directly** from an n-vowel system by addition of one optimally contrasting vowel. The decision rule of optimal contrast consists of two elements: maximal acoustic and/or articulatory contrast, and the avoidance of vowel confusion. This will be described in paragraphs 7 and 8.
In paragraph 10 we will present the results of our vowel prediction. The results will be tested against the reference vowel system data of Crothers (1978), cf. paragraph 5, and discussed in paragraphs 11 and 12, respectively.

2 THE SPEECH PRODUCTION MODEL

The relation between formant frequencies and global articulation will be described by means of the lossless tube model that has been put forward by Dunn (1950). This model comprises a description of the geometry of the vocal tract, and of the propagation of sound through the tract.
The shape of the vocal tract is modelled as a concatenation of n cylindrical tubes of equal length but different cross-sectional areas. The cross-sectional areas will be denoted by S_i ($i=1,\ldots,n$). Further, we define the area ratios k_i as

$$k_i = S_i/S_{i+1} \qquad (i = 1,2,\ldots,n-1)$$

Cf. Figure 1. Description of an n-tube in terms of its n-1 parameters k_i enables us to view such a tube as a point in the (n-1)-dimensional space spanned by the k_i. This space will be called **articulation space.**

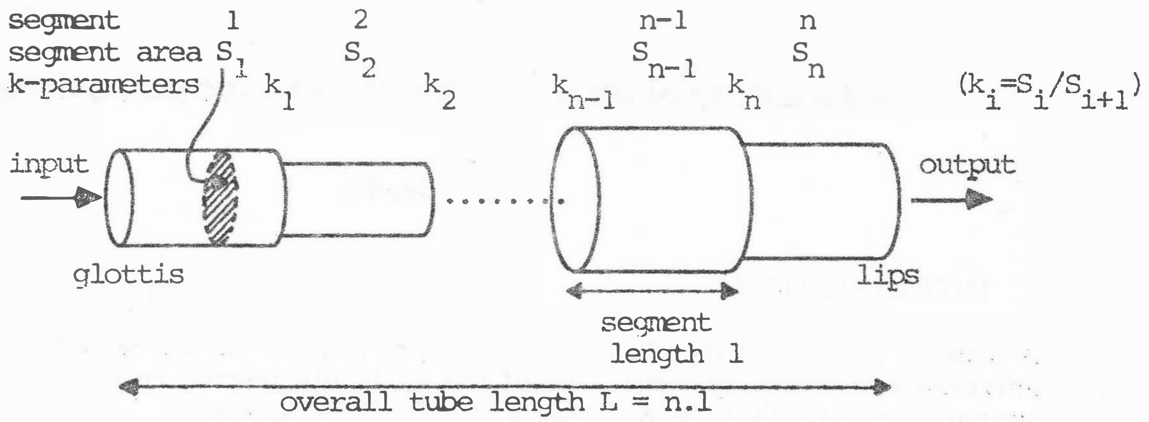


Figure 1
The n-tube model of the vocal tract.

The transmission of sound through the vocal tract is decribed by means of the one-dimensional wave equation. The speech production model assumes that no energy losses occur inside the tract or at the glottis or the lips (Bonder, 1983).

Using this speech production model, we obtain a mathematical formula, the so-called **n-tube formula** (Bonder, 1983), in which global articulation shape and formant frequencies are related implicitly.
For 4-tubes this formula reads:

$$k_1k_3 \tan^4\tau F - (k_1+k_2+k_3+k_1k_2+k_2k_3+k_1k_2k_3) \tan^2\tau F + 1 = 0 \qquad (1)$$

where the k_i are the area ratios, c the speed of sound, L the overall length of the tube, and $\tau = 2\pi L/4c$. The solutions of equation (1) are the formant frequencies F_i $(i=1,2,...)$.
The 4-tube formula (1) can be used for the computation of formant frequencies of given tube shapes. But, it can also be used for the inverse computation: tube shapes from given formant frequencies. The inverse 4-tube equations read:

$$k_2 = (-C_2k_1^2 + C_1k_1 - 1)/((1 + k_1)(1 + C_2k_1)) \qquad (2a)$$

$$k_3 = 1/C_2k_1 \qquad (2b)$$

where

$$C_1 = \tan^2\tau F_1 + \tan^2\tau F_2$$

$$C_2 = \tan^2\tau F_1 * \tan^2\tau F_2 \quad .$$

From (2) it follows that the inverse computation is not unique: for each vowel like sound, each value of k_1 gives rise to another value of k_2 and k_3. This means that each vowel like sound corresponds to an infinite set of tube shapes, represented by a continuous fibre in the articulation space (Figure 2).

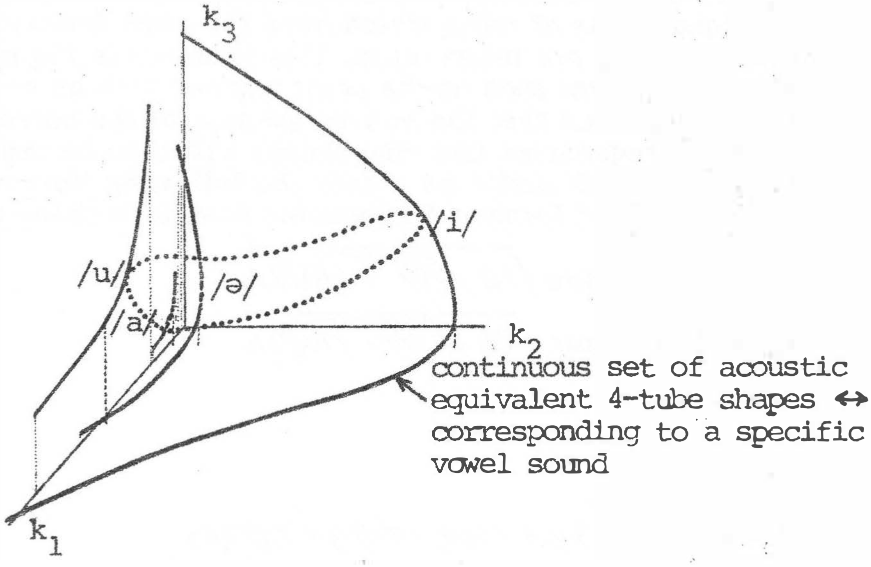

k_3
/i/
/u/
/ə/
/a/
k_2
continuous set of acoustic equivalent 4-tube shapes ↔ corresponding to a specific vowel sound
k_1

Figure 2
The 3-dimensional articulation space.

3 THE MAD MODEL

One way to get rid of the non-uniqueness of the inverse problem is to restrict the tube shapes by choosing from each fibre the tube which has minimal degree of aperture k_4, where

$k_4 = S_4/S_1$.

This model will be referred to as the **MAD model** (Minimal Aperture Degree). For a more thorough discussion of the MAD model we refer to Bonder (to appear).

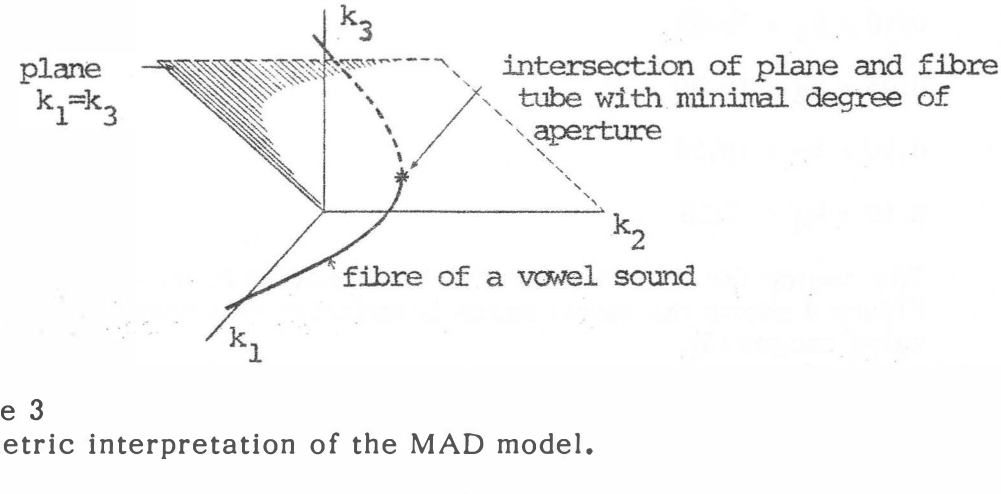

k_3
plane k_1=k_3
intersection of plane and fibre
tube with minimal degree of aperture
k_2
fibre of a vowel sound
k_1

Figure 3
Geometric interpretation of the MAD model.

It turns out that the minimization of k_4 leads to 4-tubes for which $k_1 = k_3$ (Bonder, to appear). Geometrically speaking, the minimization of k_4 means that the fibres of the articulation space, representing the continuous sets of tubes which have the same formant frequencies, and the plane $k_1 = k_3$ are intersected. This is shown in Figure 3. Each intersection yields one point such as the point marked with an asterisk in Figure 3.
It is easily seen that the non-uniqueness of the correspondence between formant frequencies and tube shapes will then be reduced to uniqueness.
Using the MAD model we obtain the following 'forward' equations for the computation of formant frequencies from tube shape parameters:

$$F_1 = (1/\pi) \arctan \sqrt{(B - \sqrt{B^2 - 4A})/2A} \qquad (3a)$$

$$F_2 = (1/\pi) \arctan \sqrt{(B + \sqrt{B^2 - 4A})/2A} \qquad (3b)$$

where

$$A = k_1 k_3$$

$$B = k_1 + k_2 + k_3 + k_1 k_2 + k_2 k_3 + k_1 k_2 k_3$$

and the inverse equations for the compuation of shapes from formant frequencies

$$k_1 = 1/(\tan \tau F_1 * \tan \tau F_2) \qquad (4a)$$

$$k_2 = \tan^2 \tau (F_2 - F_1) \qquad (4b)$$

$$k_3 = 1/(\tan \tau F_1 * \tan \tau F_2) \quad . \qquad (4c)$$

4 BOUNDARY OF THE VOWEL SPACE

The inverse equations can be used to obtain the global shape of the boundary of the vowel space from preset articulatory constraints. These articulatory constraints will be described in terms of the 4-tube shape parameters k_i.
In this paper the vowel space boundary is determined by the following ranges for the articulatory parameters:

$$0.10 < k_1 < 10.00 \qquad (5a)$$

$$0.05 < k_2 < 10.00 \qquad (5b)$$

$$0.10 < k_3 < 10.00 \qquad (5c)$$

$$0.10 < k_4 < 7.50 \quad . \qquad (5d)$$

The choice for these values will be discussed in Bonder (to appear).
Figure 4 shows the vowel space boundaries that result from the parameter value ranges (5).

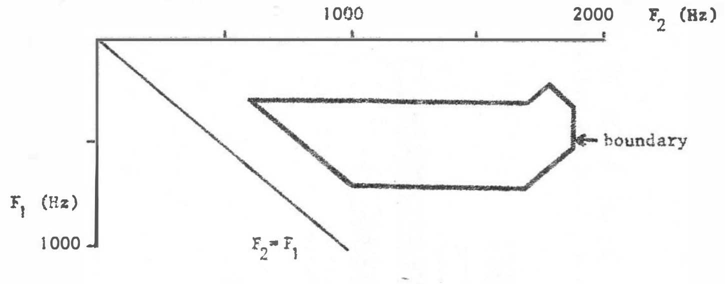


Figure 4
The vowel space as used throughout this paper. The boundaries are highly stylized.

5 VOWEL SYSTEM DATA

Our vowel system prediction and its evaluation will be based on the findings of Crothers (1978).
In his paper, Crothers carried out a statistical analysis of phonological data of 209 languages. Table 1 gives an overview of the modal n-vowel system (n=3,...,9), i.e. a system which has the highest frequency of occurrence of the set of vowel systems consisting of n vowels.

Table 1
Modal n-vowel systems (from Crothers, 1978). The fourth column shows the relative frequency of occurrence of a modal n-vowel system compared to the complete set of n-vowel systems for a specific n as shown in the second column.

number of vowels in system (n)	frequency of occurrence	modal n-vowel set	relative frequency of occurrence (in %)
3	23	i a u	100
4	22	i a u ɛ	59
5	64	i a u ɛ ɔ	86
6	40	i a u ɛ ɔ ü	73
7	28	i a u ü e o ə	50
9	15	i a u ɛ ɔ ü e o ə	47

The systems consisting of 2, 8, 10, 11, or 12 vowels, although appearing in Crothers' data base, have been left out of the overview as they are not very common (each of these systems has a relative frequency of occurrence as compared to the total set of vowel systems of less then 5%). Cf. Figure 5.

On closer examination of the third column of Table 1 we may conclude that, generally speaking from a phonological point-of-view, modal vowel systems can be considered as ordered in a hierarchical way. By this we mean that a modal system consisting of a lower number of vowels can be seen as a subset of a modal system with a higher number of vowels. This

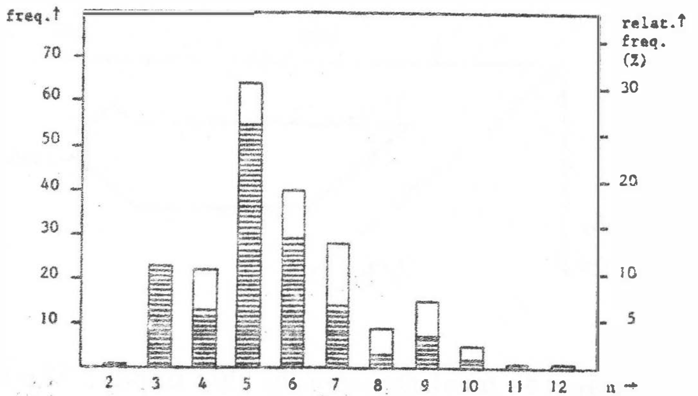


Figure 5
Frequency distribution of vowel systems as a function of the number of vowels in the system (after Crothers, 1978). The shaded parts of the columns refer to modal systems.

is visualized in Figure 6. From an acoustic-phonetic point-of-view, however, the hierarchy of systems might not necessarily be a hundred percent true because of the broadness of the phonemic transcription. As no additional data are available, we will nevertheless suppose that each of the transcriptions refer to exactly one formant frequency pattern.

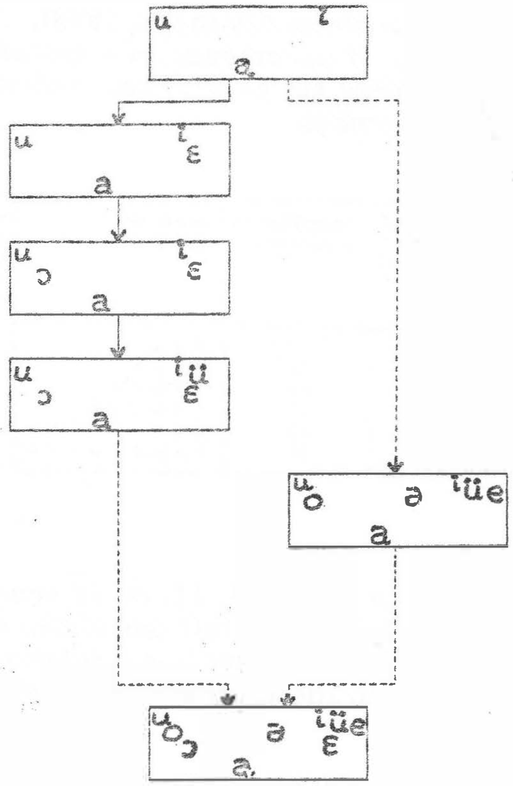


Figure 6
Hierarchy of modal n-vowel systems.

In Figure 6 the n-vowel systems have been rendered in an acoustic-phonetic way, i.e. they have been positioned in the formant space according to their F_1 and F_2 values. To this end we have normalized the data Lindblom (1986) employed for his vowel system simulations. Our normalization procedure is based on the use of the F_3 value of the respective vowels (Fant, 1975). The normalization is performed by means of the transformation

$$\tilde{F}_i = (2500/F_3) * F_i \qquad (i = 1,2) \qquad (6)$$

which corrects the F-values to values that correspond with a vocal tract length of 17.5 cm. The normalized formant frequency values for the vowels of the modal 9-vowel system are shown in Table 2.

Table 2
Normalized formant frequency values of the vowels of the modal 9-vowel system.

vowel	$\tilde{F}_1$ (Hz)	$\tilde{F}_2$ (Hz)
i	205	1760
ü	289	1816
u	311	670
e	324	2010
ə	391	1450
o	406	715
ɛ	449	1845
ɔ	518	814
a	744	1240

We will use the normalized formant frequency values as reference values for the evaluation of our vowel system prediction.
Figure 7 shows how the normalized vowels are positioned with respect to the vowel space boundaries that have been introduced in paragraph 4.

6 DISCRETIZATION OF THE VOWEL SPACE

For computational reasons the vowel space as defined in Figure 4 is discretized into a finite grid of points which have a vertical and horizontal interspace of 100 Hz (Figure 8).
Taking multiples of 100 Hz for the coordinate values of the grid points,

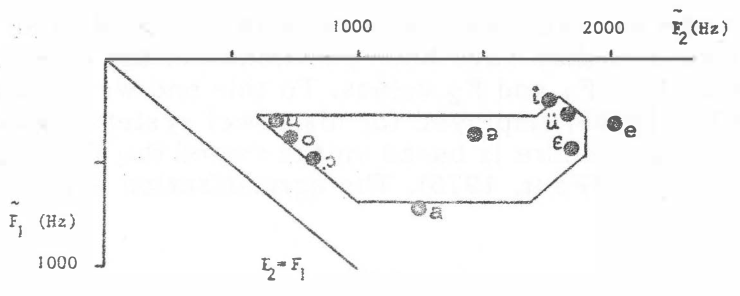


Figure 7
The location of the normalized vowels of the modal 9-vowel system in the formant space.

we thus obtain a set of 58 different vowel like sounds. Each constructed vowel system is a subset of this set of 58 sounds.
Obviously, a consequence of the discretization of the vowel space is that it is not allowed for two vowels to lie too close to each other, acoustically.

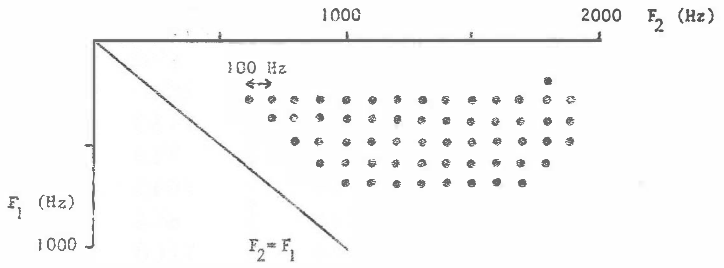


Figure 8
Discretization of the vowel space.

The discretization as used in our model can also be interpreted in terms of confusion between vowels. The probability of confusion between any two vowels v_1 and v_2 can be described as (Ten Bosch, 1986):

$$p(v_1,v_2) = \exp(-\alpha * d_F(v_1,v_2)), \tag{7}$$

where α denotes a weighing factor, and $d_F(v_1,v_2)$ the acoustic contrast between v_1 and v_2. The acoustic contrast d_F may be based on, for example, a linear or a logarithmic distance measure in the formant space. The graph of the probability function (7) is shown in Figure 9a.

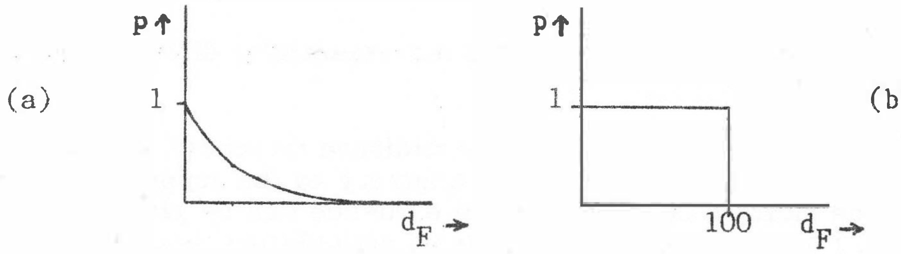


Figure 9
Probability functions of vowel confusion; (a) shows the function $p = \exp(-\alpha * d_F)$ as used by Ten Bosch (1986); (b) shows the probability function $p = \theta(100 - d_F)$ as induced by the discretization of the formant space.

The discretization of the vowel space as discussed earlier in this paragraph implies the nonexistence of pairs of vowel sounds which have an acoustic contrast less than 100 Hz. In terms of confusion we may interpret this as if a contrast smaller than 100 Hz causes too much confusion, while a larger contrast leads to an acceptable (low) degree of confusion. So we may say that in our model a contrast of $d_F < 100$ Hz corresponds to 100% confusion (p=1), and $d_F > 100$ Hz to no confusion (p=0). We might conclude that we are actually using the vowel confusion probability function

$$p(v_1,v_2) = \theta(100 - d_F(v_1,v_2)) \quad . \tag{8}$$

The graph of this function is shown in Figure 9b.
However, for our model we will slightly modify the vowel confusion probability function (8) into one which is also a function of the vowel system density (i.e. the number of vowels in the system). We will describe such a probability function in the next paragraph.

7 VOWEL CONFUSION PROBABILITY

The vowel probability function as defined by (8) is not fully adequate. We may explain this as follows. The vowel confusion probability based on the uniform discretization of the vowel space is independent of the number of vowels in the system. This means that the 'repelling forces' between the vowels are relatively small for small systems. This can be overcome by introducing a minimum threshold distance which is dependent on the number n of vowels in the system.
In our paper we will adopt a threshold for F-distances

$$d_F^{(n)}(v_1,v_2) > A/\sqrt{n} \tag{9}$$

where d_F may be based on a linear or other distance measure, and A is a constant to be determined.
This leads to the vowel confusion probability function

$$p(v_1,v_2) = \theta\,(A/\sqrt{n} - d_F(v_1,v_2)) \quad . \qquad (10)$$

The graph of such a step function is not essentially different from the one in Figure 9b.

In the same way we may introduce a minimum threshold for the tube shape parameters k_i. Although this seems contrary to the supposition that articulation must be optimized, some evidence can be given for such a threshold. For, if we think of vowels as articulatory target positions there obviously must be at least some difference between the targets, otherwise they will be confused.
This leads us to a threshold of minimum articulatory difference between two vowels v_1 and v_2

$$d_k{}^{(n)}(v_1,v_2) > B/\sqrt{n} \qquad (11)$$

where B is a constant to be determined. This leads in a way analoguous to that for the formant space to the vowel confusion probability function in the articulation space:

$$p(v_1,v_2) = \theta\,(B/\sqrt{n} - d_k(v_1,v_2)) \quad . \qquad (12)$$

The graph of such a function is not essentially different from the one in Figure 9b.

8 MEASURES OF CONTRAST

We will investigate three types of contrast measuring. They are:
Type 1: A contrast measure in the formant space based on values of F_1 and F_2 (this measure will be denoted as Q_1).
Type 2: A contrast measure in the articulation space, based on the vocal tract parameters k_1, k_2, k_3 (denoted as Q_2).
Type 3: A contrast measure based on both formant frequencies and vocal tract shapes (Q_3).
The mathematical formulation of these 3 types of contrasts is, as expressed in the intervowel distances

$$d_F(v_i,v_j) = \sqrt{(F_1{}^{(i)} - F_1{}^{(j)})^2 + (F_2{}^{(i)} - F_2{}^{(j)})^2} \qquad (13)$$

$$d_k(v_i,v_j) = \sqrt{(k_1{}^{(i)} - k_1{}^{(j)})^2 + (k_2{}^{(i)} - k_2{}^{(j)})^2 + (k_3{}^{(i)} - k_3{}^{(j)})^2} \qquad (14)$$

given by

$$Q_1 = \sqrt{[\sum_{i,j} d_F{}^{(i,j)}]^2} \qquad (15a)$$

$$Q_2 = \sqrt{[\sum_{i,j} d_F{}^{(i,j)}]^2} \qquad (15b)$$

$$Q_3 = Q_1/Q_2 \quad . \qquad (15c)$$

In addition to these two linear measures we will also apply a logarithmic scale for the formant frequency values. The mathematical formulae for the logarithmic case differ from the ones in the linear case in the sense that F is replaced by log F.

9 THE FVSP ALGORITHM

The complete forward vowel system prediction algorithm (the **FVSP algorithm**) is based on both contrast measuring (paragraph 8) and vowel confusion probability (paragraph 7).
The flow chart of the FVSP algorithm is shown in Figure 10.

The algorithm can be performed under several conditions:

a. We may choose the 'initial' set of 3 vowels; the algorithm is also applicable for 4 or more initial vowels.

b. The type of contrast measuring can be preset: formant contrast, articulatory contrast, or both.

c. We may choose a linear or a logarithmic distance measure in the formant space.

d. Vowel confusion probability (either in the formant space or in the articulation space, or both) may be introduced in order to ensure sufficient contrast.

The operative conditions under which the vowel system predictions are performed will be described in the next paragraph.

10 PREDICTION OF VOWEL SYSTEMS

The FVSP algorithm has been applied with the initial 3-vowel system consisting of /i/, /a/, /u/, in agreement with the vowel system hierarchy which has been put forward by Crothers (1978) (cf. also Figure 5).
The vowel confusion boundaries as defined by (9) and (11) for the formant space and the articulation space, respectively, are specified as follows (where the asterisk denotes that the formant frequencies are scaled to one-hundredth of their actual values):

$$d_F*(v_i,v_j) = 5/\sqrt{n} \qquad (16)$$

for a linear formant space,

$$d_F*(v_i,v_j) = 0.8/\sqrt{n} \qquad (17)$$

for a logarithmic formant space, and

$$d_k(v_1,v_2) = 3/\sqrt{n} \qquad (18)$$

for the articulation space.
All predictions are carried out under the condition that the acoustic contrast between any two vowels must exceed either the threshold (16) (linear case) or the threshold (17) (logarithmic case).
The predictions are performed partly with and partly without an articulatory threshold (18).

Figure 11 shows the prediction made by means of the combined acoustic/articulatory contrast (15c) under the 4 conditions: linear or logarithmic

scale for the formant space, and threshold or no threshold for the articulatory contrast.
The Figures 12 and 13, respectively, show vowel system predictions made with either formant or articulatory contrast, also under the 4 conditions that have been mentioned above.

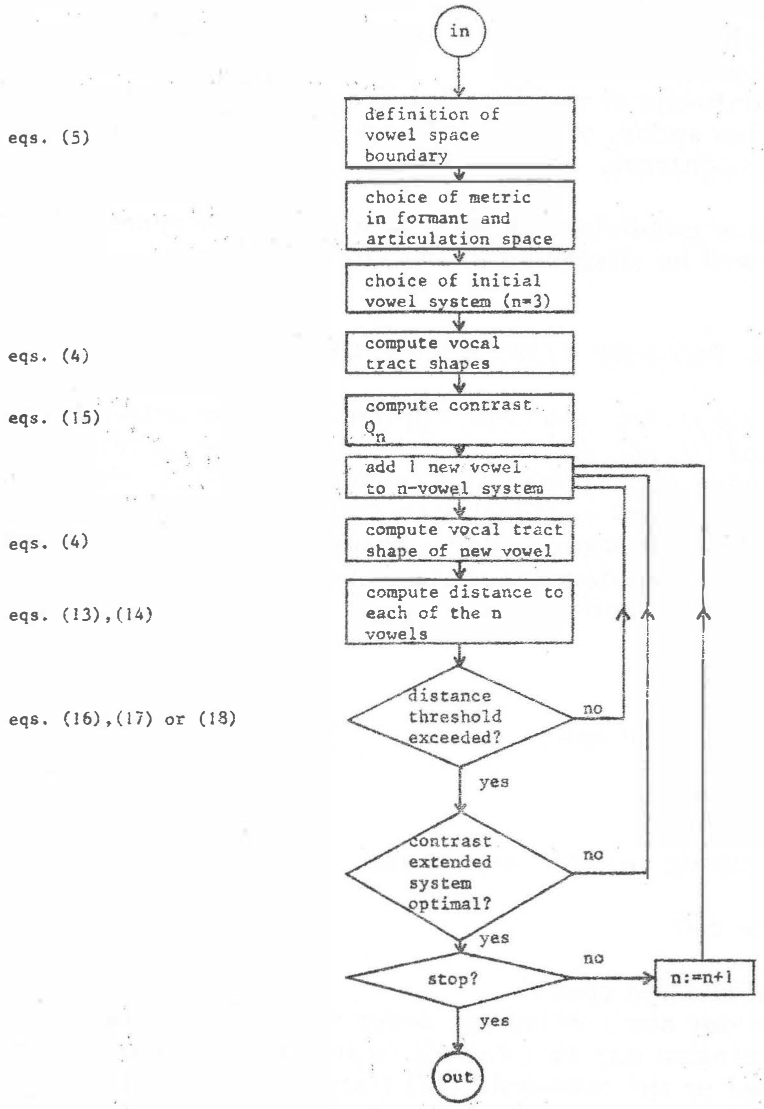

in
eqs. (5)
definition of vowel space boundary
choice of metric in formant and articulation space
choice of initial vowel system (n=3)
eqs. (4)
compute vocal tract shapes
eqs. (15)
compute contrast Q_n
add 1 new vowel to n-vowel system
eqs. (4)
compute vocal tract shape of new vowel
eqs. (13),(14)
compute distance to each of the n vowels
eqs. (16),(17) or (18)
distance threshold exceeded?
no
yes
contrast extended system optimal?
no
yes
stop?
no
n:=n+1
yes
out

Figure 10
Flow chart of the FVSP algorithm. The figure also shows, at each stage, the formulae that have been used.

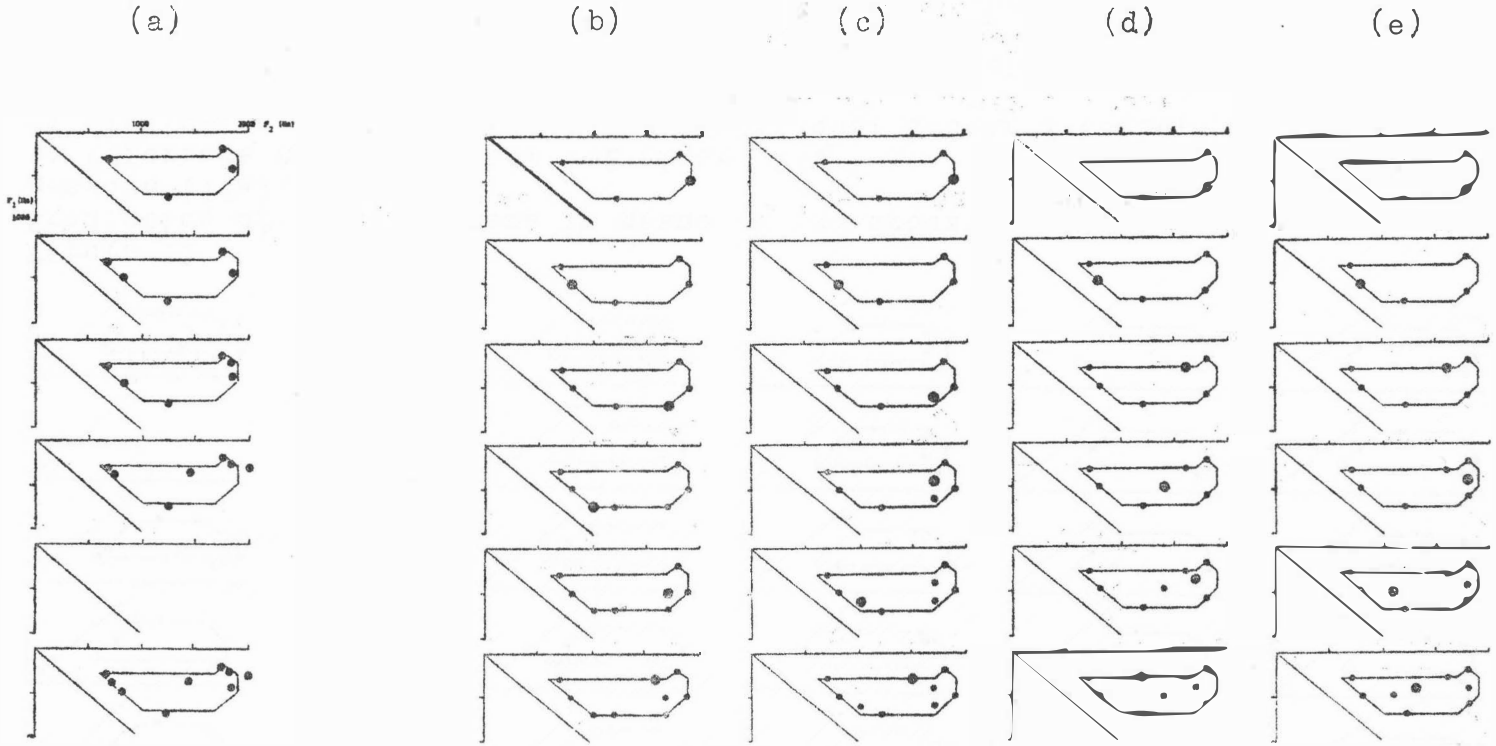


Figure 11
Prediction of vowel systems by means of the combined acoustic and articulatory contrast (15c).
a: reference n-vowel system (cf Figure 7).
b: prediction under the following conditions: linear F scale, no articulatory threshold.
c: conditions: linear F scale, articulatory threshold.
d: conditions: logarithmic F scale, no articulatory threshold.
e: conditions: logarithmic F scale, articulatory threshold.
The solid dots denote the n-th predicted vowel.
All computations are carried out under the restriction of sufficient acoustic inter vowel contrast (either condition (16) or (17)).

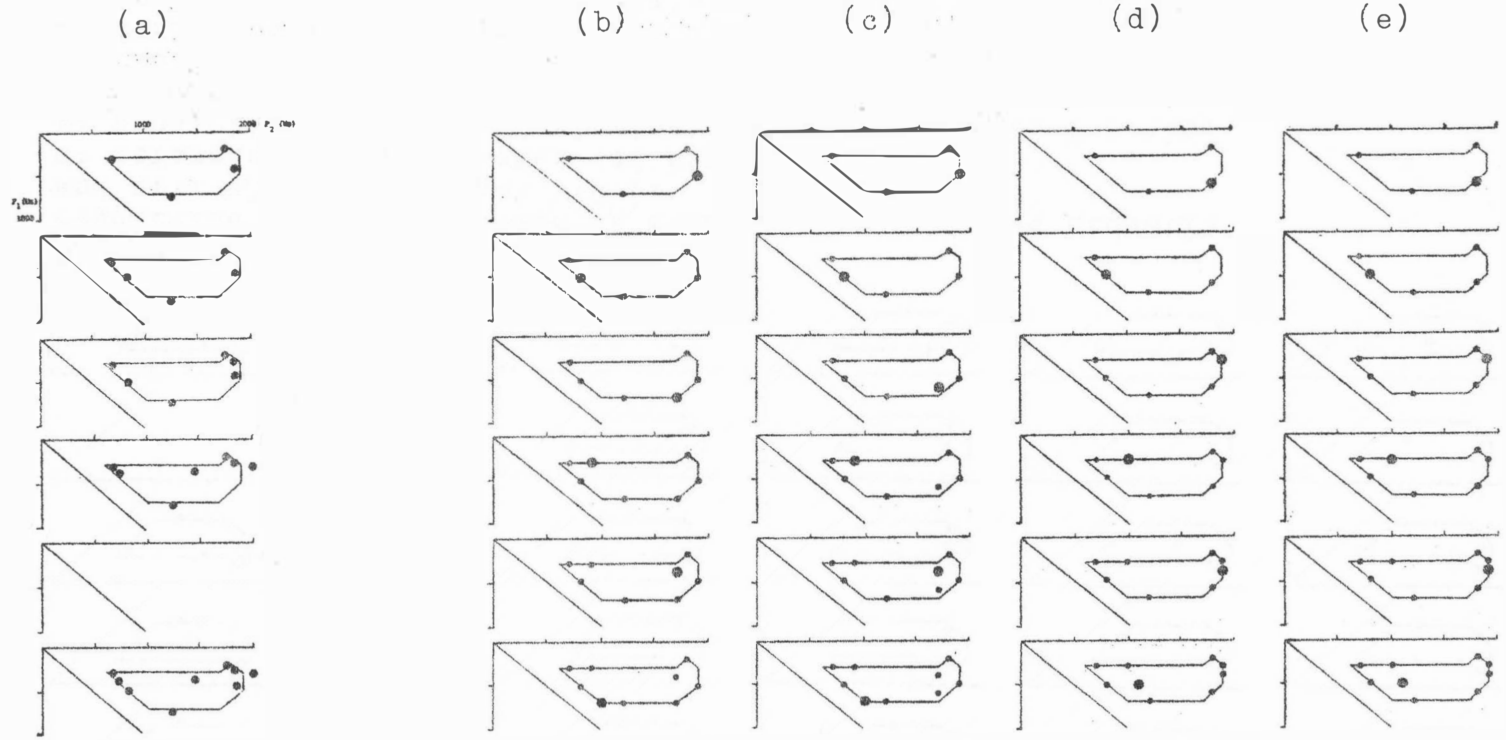


Figure 12
Prediction of vowel systems by means of the acoustic contrast measure (15a).
a: reference n-vowel system (cf Figure 7).
b: prediction under the following conditions: linear F scale, no articulatory threshold.
c: conditions: linear F scale, articulatory threshold.
d: conditions: logarithmic F scale, no articulatory threshold.
e: conditions: logarithmic F scale, articulatory threshold.
The solid dots denote the n-th predicted vowel.
All computations are carried out under the restriction of sufficient acoustic inter vowel contrast (either condition (16) or (17)).

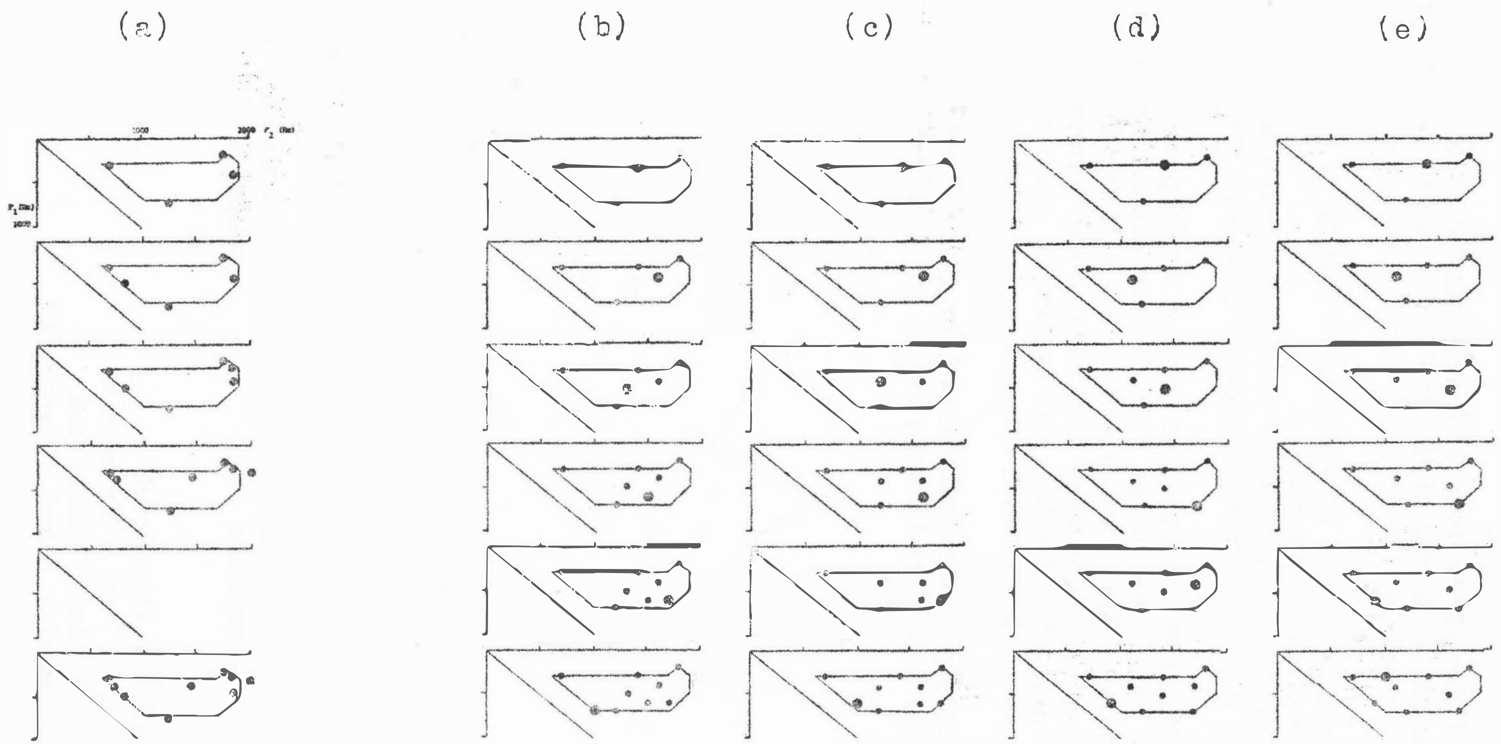
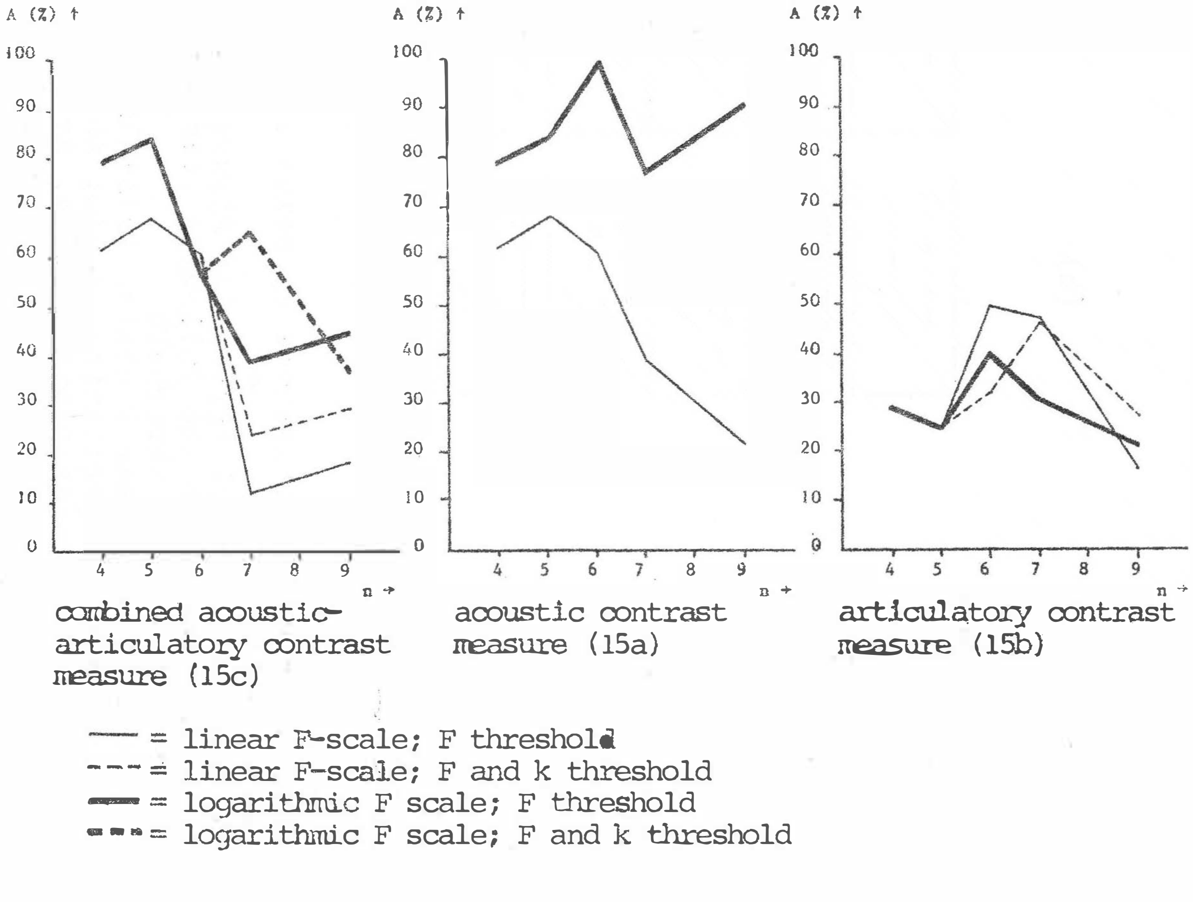


Figure 13
Prediction of vowel systems by means of the articulatory contrast measure (15b).
a: reference n-vowel system (cf Figure 7).
b: prediction under the following conditions: linear F scale, no articulatory threshold.
c: conditions: linear F scale, articulatory threshold.
d: conditions: logarithmic F scale, no articulatory threshold.
e: conditions: logarithmic F scale, articulatory threshold.
The solid dots denote the n-th predicted vowel.
All computations are carried out under the restriction of sufficient acoustic inter vowel contrast (either condition (16) or (17)).

11 GOODNESS OF FIT

For the comparison of the predictions with the original data we will investigate in how far the ordering of the predicted vowels in the formant space fits that of the vowels in the corresponding actual modal n-vowel systems.
For two reasons we prefer testing of ordering to testing of nearness by means of some kind of distance measure.
First, the exact location of the reference vowel sounds in the formant space is not known as the transcriptions used by Crothers are broad transcriptions.
Secondly, the normalization procedure as carried out in paragraph 5, equation (6), is based on the use of the F_3 value for vocal tract length estimation. Although this yields a quite acceptable estimation of the length, in general, it might cause some deviation of the formant frequency value from the 'real' normalized value; this especially occurs when the value F_3 is notably non-harmonic with respect to the values of higher formants.
These two reasons explain why calculation of the distance between a predicted system and a normalized actual one is not preferable in our case.



combined acoustic-articulatory contrast measure (15c)

acoustic contrast measure (15a)

articulatory contrast measure (15b)

—— = linear F-scale; F threshold
- - - = linear F-scale; F and k threshold
▬▬ = logarithmic F scale; F threshold
▪ ▪ ▪ = logarithmic F scale; F and k threshold

Figure 14
The acceptability of the predicted systems as a function of the number of vowels in the system (n).

The goodness of fit of the predicted systems is tested with a two-tailed sign test, under the hypothesis that, as a mean, half of the formant frequency values of the predicted system are lower than those of the reference system, and half of the values higher. For each predicted system we have computed the probability of not being rejected under this hypothesis.
The sign test is carried out in the F_1 and F_2 direction separately. The probability of not being rejected in the F1 and F_2 direction is denoted by A_1 and A_2, respectively.
As a measure for the overall acceptability of a predicted system we have taken the geometric mean of A_1 and A_2:

$$A = \sqrt{A_1 * A_2} \quad . \tag{19}$$

The results of the sign test are shown in Figure 14.

12 DISCUSSION

The results of the two-tailed sign tests (Figure 14) show that it is possible to predict with rather high accuracy the modal n-vowel system directly from the modal (n-1)-vowel system.

The vowel system prediction based on an acoustic distance measure in the logarithmic F-space gives the best results. A special feature of this prediction method is the introduction of a lower boundary for the acoustic distance between any two vowels. The acoustic threshold in the logarithmic formant space (given by (17)) results in greater repelling areas around vowel sounds for greater formant frequency values. Such a threshold fits the experimental findings that vowel sounds- in order to be perceived as different from each other- must have a formant frequency difference that exceeds a certain minimal threshold that is roughly linear with the F values (the so-called critical band). This could explain why 'logarithmic' prediction works better than the 'linear' alternative.

Further, we may conclude from the results that incorporation of an articulatory threshold into the model for prediction may improve the prediction results, especially for higher order systems.
The introduction of an articulatory threshold, although seemingly arbitrary, may be understood in the same way as the introduction of the acoustic threshold: it reflects the avoidance by a speaker of articulatory targets that lie too close to each other.

We may also notice that the prediction of the modal 7-vowel system is relatively bad.
This can be explained by the fact that the modal 7-vowel system does not fit in with the hierarchy of the modal 3- to 6-vowel systems (cf. Figure 6). Thus we could have expected a priori that the prediction of the modal 7-vowel system necessarily deviates from that of other systems.

The deviation of the modal 7-vowel system from the main hierarchy in modal vowel systems leads to a new question: can the non modal vowel systems and the modal 7-vowel system be categorized into phonetically based hierarchies in a way analoguous to that for the modal systems?
As an indication for the existence of other hierarchies we may mention

the fact that the frequency distribution of Crothers' database of 209 languages is **multi**modal (cf. Figure 5). Furthermore, we see that the distribution of the non modal systems (indicated by the non shaded parts of the columns in Figure 5) shows a maximum frequency for systems consisting of 7 vowels, which points maybe to a hierarchy in which 7-vowel systems play a great role.

The main problem in categorizing vowel systems into phonetically interpretable hierarchies , however, lies in the modelling of adequate underlying generating principles for non modal n-vowel systems as they are presumably not just as simple as those for the modal systems (that is what they are **modal** systems for ...?).

ACKNOWLEDGMENTS

We wish to thank L.C.W. Pols and L.F.M. ten Bosch for their valuable comments on an earlier version of this paper.

The research for this paper was supported by the Foundation for Linguistic Research, funded by the Netherlands Organization for the Advancement of Pure Research Z.W.O. under Contract 300-161-030.

REFERENCES

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

Bonder, L.J. (to appear). The MAD model.

Bosch, L.F.M. ten (1986). Architecture of vowel systems. Procs. Inst. Phon. Scs., Univ. of Amsterdam, vol. 10, pp. 55-72.

Crothers, J. (1978). Typology and universals of vowel systems. In: Universals of human language; ed. J.H. Greenberg; pp. 93-152.

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

Fant, G. (1975). Non-uniform vowel normalization. STL-QPSR vol. 2-3, pp. 1-19.

Liljencrants, J. and B. Lindblom (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. Language vol. 48, pp. 839-862.

Lindblom, B. (1986). Phonetic universals. In: Experimental phonology; eds. J.J. Ohala and J.J. Jaeger; pp. 13-44.