STATIC AND DYNAMIC STRUCTURE OF VOWEL SYSTEMS

L.F.M. ten Bosch L.J. Bonder L.C.W. Pols

Institute of Phonetic Sciences, University of Amsterdam

ABSTRACT

A phonetic/phonological model has been developed for describing the structure of natural vowel systems in terms of configurations consisting of N points in the formant space. These configurations (abstract vowel systems) are defined as solutions of an optimalisation algorithm. This search algorithm uses an optimality strategy that is based upon two extra-linguistic principles, one dealing with the articulatory effort, the other with perceptual ease. The model is evaluated by comparing the model results with available phonological data.

INTRODUCTION

The model that we present is developed in order to find basic structure principles underlying the architecture of vowel systems. It uses as a starting-point the dispersion model of Liljencrants and Lindblom (1972). They tried to describe natural vowel systems by maximizing an acoustic distance measure between N points, all of them positioned within a predefined fixed region in the formant space. The novelty of the present model is the extension of the acoustic principle (with respect to vowel dispersion only) with an articulatory minimal effort principle.

In the following three sections, we will gradually unfold the model. Section 1 poses the two basic structure principles we are using. Section 2 describes the model itself: 2.1 deals with the technical translation of the basic principles into an appropriate mathematical formulation and a search algorithm for the abstract vowel systems; 2.2 describes the comparison of these abstract systems with the vowel systems from natural languages; and 2.3 will briefly deal with the implementation of *dynamic* aspects of vowel systems: the long/short-opposition and the diphthongs. In section 3 we will give a summary of the present results. In section 4 we conclude with a discussion.

1. THE PRINCIPLES

We use two principles dealing with the structure of vowel systems which are supposed to be of primary importance:

(a): *minimality of effort* of (static) vowel pronunciation;

(b): *minimality of inter-vowel confusion*.

Vowel systems are said to be 'optimal' if they optimally satisfy both principles simultaneously.

Evidently, the consequences of these principles separately are conflicting: (a) yields minimal overall articulatory vowel distances, whereas (b) leads to maximal inter-vowel distances. In order to be able to handle both principles in an appropriate way, they have been translated into specific mathematical formulae. Some of these formulae directly deal with both the formant position of vowels and the vocal tract area function, other ones are based upon arguments concerning probability and optimalisation techniques (see section 2.1, the search algorithm).

2. THE MODEL

2.1. The Search Algorithm

Each vowel system is represented as a point in a so-called 'state space', in which principles (a) and (b) define an optimality strategy. The search for optimal vowel systems can be considered as looking for stable solutions in this state space. In order to specify the search algorithm, we introduce the following formulae (classified into basic, derived and evaluational ones):

2.1.1. *basic formulae*

These formulae play the most elementary role in the model.

The *inter-vowel acoustic distance* d_F between v1 and v2 is defined as follows:

$$(d_F)^2 = (\log(F_1(v1)) - \log(F_1(v2)))^2 + (\log(F_2(v1)) - \log(F_2(v2)))^2 \qquad (1)$$

Only the relative positions of vowels in a vowel system are relevant. The logarithms of the frequencies are used to meet with the perceptual behaviour of the basilar membrane. This closely relates dF to empirically determined acoustic distance measures involving mel or bark scales.

The expression for the *inter-vowel confusion probability* p(v1, v2) reads:

$$p(v1, v2) = \exp(-\alpha * d_F(v1, v2)) \qquad (2)$$

α being a positive scaling parameter.
Before actually evaluating vowel systems we first introduce the following probabilistic concept. We hypothesize an exponential relation between the inter-vowel confusion probability p and the inter-vowel acoustic distance dF. This relation can be globally verified by inspecting the perceptual vowel confusion matrices in several languages.

We define the *articulatory effort* dA:

$$d_A = \sum (Si - 1)^2 \quad (i = 1, \ldots, 4) \qquad (3)$$

This expression relates the shape of the vocal tract (which is approximated by the straight 4-tube, consisting of 4 segments of equal length with areas Si (cf. [1], [2])) to an articulatory effort value (see figure 1).

2.1.2. *derived system formulae*
In order to be able to define the structure principle for vowel systems as a whole, we introduce the system counterparts of dA and dF.

The expression for the *total articulatory system effort* DA reads:

$$D_A = \max(d_A) \qquad (4)$$

The articulatory effort value of a vowel system is defined as the maximal value of the articulatory effort values of its members.

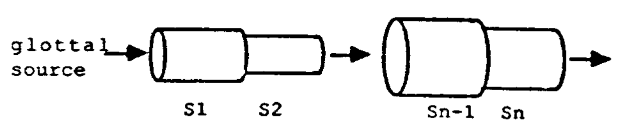


Fig 1. An example of a general n-tube with segment areas Si.

The *total perceptual system discriminality* DF will be

$$D_F = \prod (1 - p(v_i, v_j)) \qquad (1 \leqslant i < j \leqslant N) \qquad (5)$$

1- p(v1, v2) denotes the probability of vowel v1 and vowel v2 not being mutually confused. Therefore DF is a measure for the total discriminality of an N-vowel system. Consequently we have DF = 1 in case of perfect discriminality and DF = 0 in the worst case.

2.1.3. *evaluation formulae*
We have to minimize the articulatory effort DA and to optimize the discriminality measure DF simultaneously. Therefore we introduce the *penalty parameter* Q relating both aspects:

$$Q = (D_A)^2 + S * (D_F - 1)^2 \qquad (6)$$

This type of expressions is well-known from optimality theory and is in fact a natural choice here. Indeed, minimization of Q logically implies minimization of DA towards zero and optimization of DF towards unity simultaneously. The rate of convergence of this process is controlled by the slack variable S (S being a large positive number). Optimal vowel systems are locally found by iteratively improving the position of all vowels in the system while decreasing the value of Q.

2.2. Evaluation Part
The evaluation part of the algorithm described above in fact consists of a measurement of the goodness of fit of the acoustic model output in relation to the more phonologically specified data from language databases ([3], [4]). For the time being we confine the evaluation to vowel systems without dynamic structure (without short/long opposition, without diphthongs). Presently, these latter effects contribute less to a general insight as they are second-order consequences.
In the model a method is implemented for actually effectuating the phonetic/phonological comparison. It is based upon essentially the same probabilistic motivations as already used in formula (5). The result of the comparison is expressed in terms of the *similarity probability* (denoted SP) of the respective abstract phonetic vowel system and a phonological system after having optimally paired each unlabelled v_i in the model system with a vowel v_j in the phonological reference system.

$$SP = \prod \exp(-\alpha * d(v_i, v_j)) \qquad (7)$$

If SP = 1, the similarity is perfect. The model evaluation now consists of the evaluation of all SP values between a model solution containing N vowels and all known phonological N-vowel systems. The present result of this evaluation is plotted in figure 2.

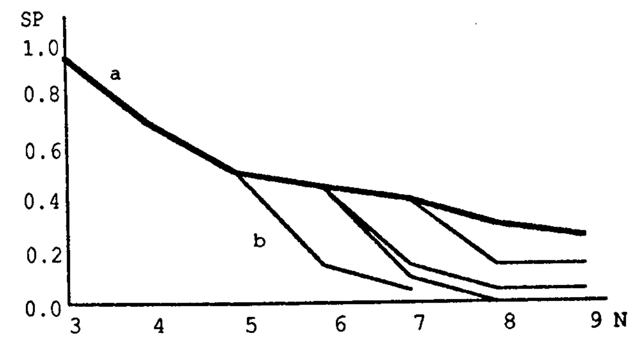


Fig.2. Goodness of fit of the present model in terms of the SP value.
The heavy line a connects all the found maxima, b shows some possible ramifications. N denotes the number of vowels in the model system and the phonological reference system.
One observes the decreasing SP value for increasing values of N. Probably this phenomenon can be traced to
- the declining fit of the model itself
- the increasing number of linguistic possibilities for large N.

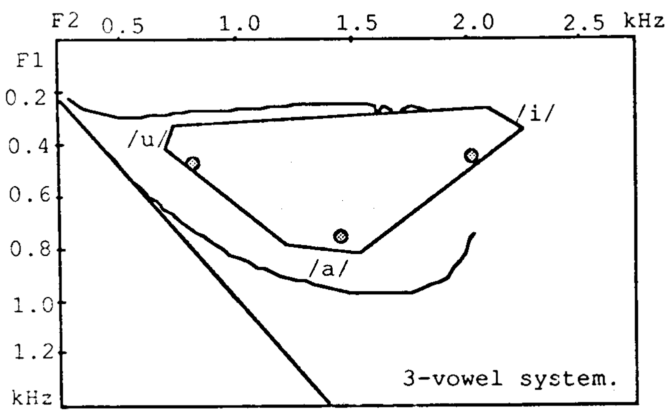
2.3. Dynamics
The description of the dynamic part of vowel systems appears to involve more linguistic details then are contained in the model described above. The model has proved to be inadequate for predicting actual diphtongs and long vowels in a specific language, but it merely defines and bounds the set of physical possibilities out of which a language may select. In order to study these possibilities in more detail we use a *vowel structure matrix* of which the entries represent the long vowels and diphthongs. The short vowels constitute the elements along the two axes. Evidently, long vowels emerge as geminates along the main diagonal and diphthongs off the diagonal. In order to evaluate the entries we considered the acoustic gain relative to the articulatory effort. We give the results of such a calculation in figure 3. One may observe a preference for diphthongs to start in the /a/-region (i.e. to show decreasing first formant frequency).

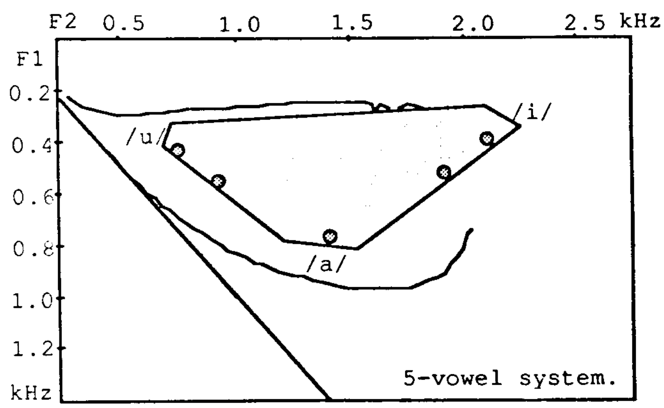
	α	ε	I	ɔ
ɔ	0.5	0.3	0.3	0.6
I	0.6	0.5	0.7	0.4
ε	0.4	0.5	0.4	0.3
α	0.7	0.3	0.5	0.4

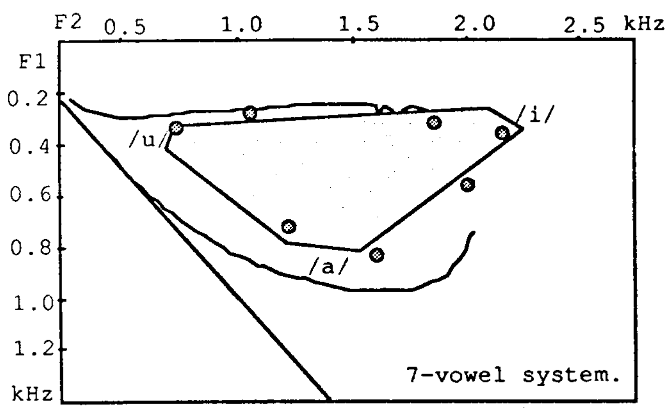
Fig 3. Gain of acoustic contrast in relation to articulatory transitional effort. The transitions are now described as concatenations of two short vowels out of the indicated set of four short vowels. Horizontally, we denote the vowels in initial position and vertically the short vowels in final position are shown. All entries (quotients of acoustic contrast and articulatory effort) have been rescaled to values between 0 and 1. They give an indication of the preference of the corresponding combination of short vowels.
In this four-vowel system the preference for transitions to start in the a-like region of the formant space is demonstrated by the values figuring in the first column relative to those in the other columns. The overall-preference for gemination can be deduced from the values along the diagonal. In general it does not have to be the case that these geminates correspond to actual long vowels such like /a/, /e/ etc. This identification is in fact a phonological item. The quotients have been specified up to only one decimal place in order to express their tentative character. They only have relative significance.

3. RESULTS OF THE MODEL

In the figures 4, 5 and 6 we give the present model solution in case of N = 3, 5, and 7 respectively. The closed contours represent contour lines of the articulatory effort function dA. One observes:
- the preference for the vowel /a/, followed by /i/ and /u/;
- the preference for vowels along the lines /a/-/i/ and /a/-/u/;
- the limitation of the available vowel space without predefining a fixed boundary in the formant space.







Figures 4, 5 and 6 show the model solution in the formant space. For reference the grey area indicate the region which is used by most languages. The staight line denotes the line F1 = F2. The other two lines are contour lines of the articulatory effort function dA, which gives an idea of the theoretically shaped vowel space by using an effort principle (see the text). In case of the 7-vowel system, some of the vowels are positioned outside the grey area, as a consequence of the subtile imperfection of the balance between the two principles (a) and (b) (see the text).

4. DISCUSSION

In our project, we explicitly deal with the model in relation to other recent vowel dispersion theories as well as with recent improvements. The present results have led to the following two suppositions:

a) natural vowel systems may adequately be considered as derivations of specific 'abstract' vowel systems, while
b) the structure of these abstract vowel systems is defined by two extra-linguistic principles:
   - reduction of perceptual vowel confusion probability and
   - reduction of articulatory effort.

The present model certainly does not pretend to be the final answer to the question of the structure of vowel systems in general but it may stimulate a further fundamental approach to the subject. In our presentation we will briefly mention some of the parallels with recent phonological theories, e.g. [5]. Our model does not predict all linguistic details of vowel systems as it is not based upon such linguistic or other language-sensitive principles. However, some important tendencies are clearly demonstrable: tendencies in the appearance and behaviour of vowel systems are described by combining a few, indeed simple arguments concerning articulation and perception. The main question will be the search for a convincing theory relating vowel systems as they are actually observed on the one hand to the results of a stipulative or normative model at the other hand.

ACKNOWLEDGEMENTS

This project is sponsored by the Netherlands Organization for the Advancement of Pure Research ZWO (project no. 300-161-030).

REFERENCES

[1] Dunn, H.K. (1950). The calculation of vowel resonances, and an electrical vocal tract. J. Acoust. Soc. Amer., vol. 22., pp. 740-753.
[2] Bonder, L.J. (1983). The n-tube formula and some of its consequences. Acustica vol. 52, pp. 216-226.
[3] Crothers, J. (1978). Typology and universals of vowel systems. In: Universals of human language. Ed. J.H. Greenberg. Vol. 2. Stanford University Press, Calif., pp. 93-152.
[4] Maddieson, I. (1984). Patterns of sounds. Cambridge studies in speech science and communication. Cambridge University Press.
[5] Kaye, J., Lowenstamm, J. and Vergnaud, J.-R. (1985). Vowel systems. Paper presented at the 1985 GLOW colloquium, Brussels.