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 optimisation 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 optimisation 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 dF between v1 and v2 is defined as follows:

$$(dF)^2 = (\log(F_1(v1)) - \log(F_1(v2)))^2 + (\log(F_2(v1)) - \log(F_2(v2)))^2$$  \hspace{1cm} (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 da to empirically determined acoustic distance measures involving mel or bark scales.

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

$$p(v1, v2) = \exp(-\alpha \cdot d(v1, v2))$$  \hspace{1cm} (2)

\(\alpha\) being a positive scaling parameter.

Before actually evaluating vowel systems, we first introduce the following perceptual model. We hypothesize an exponential relation between the inter-vowel articulatory effort and the inter-vowel acoustic distance. This relation can be globally verified by inspecting the perceptual-vowel confusion matrices in several languages.

We define the articulatory effort as:

$$DA = \sum (S_1 - 1)^2 \quad (t = 1, \ldots, 4)$$  \hspace{1cm} (3)

This expression relates the shape of the vocal tract (which is approximated by a straight 4-tube, consisting of 4 segments of equal length with areas \(S_1, \ldots, S_4\)) to an articulatory effort value (see figure 1).

### 2.1.2. derived system formulae

In order to be able to define the structure of a model for vowel systems as a whole, we introduce the system configuration parameter \(A\).

The expression for the total articulatory system effort reads:

$$DA = \max(da)$$  \hspace{1cm} (4)

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

The total perceptual system discriminability \(Dr\) will be

$$Dr = \prod (1 - p(v_i, v_j)) \quad (1 \leq i < j \leq N)$$  \hspace{1cm} (5)

1 - p(v_i, v_j) denotes the probability of vowel \(v_i\) and vowel \(v_j\) not being mutually confused. Therefore \(Dr\) is a measure for the total discriminability of an N-vowel system. Consequently we have \(Dr = 1\) in case of perfect discriminability and \(Dr = 0\) in the worst case.

### 2.1.3. evaluation formulae

We have to minimize the articulatory effort \(DA\) and to optimize the discriminability measure \(Dr\) simultaneously. Therefore we introduce the penalty parameter \(Q\) relating both aspects:

$$Q = (DA)^2 + s \cdot (Dr - 1)^2$$  \hspace{1cm} (6)

This type of expression 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 \(Dr\) 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 positions of the vowels in the system while decreasing the value of \(Q\).

### 2.2. Evaluation Part

The evaluation 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 phonological reference of the corresponding combinatorial model.

The total perceptual system discriminability \(Dr\) is a measure for the total discriminability of an N-vowel system. Consequently we have \(Dr = 1\) in case of perfect discriminability and \(Dr = 0\) in the worst case.

### 2.3. Dynamics

The description of the dynamic part of the vowel systems appears to involve more linguistic details than are contained in the model described above. The model has proved to be inadequate for predicting actual diphthongs and long vowels in a specific language, but it merely defines and bounds out the set of physical possibilities without 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 low vowels and short diphthongs. The short vowels constitute the elements along the upper diagonal, the long vowels emerge as geminates along the main diagonal and diphthongs are defined by 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 long vowels without predefining a fixed boundary in the formant space.

### 3. RESULTS OF THE MODEL

In the figures 4, 5 and 6 we give the present model solution of \(N = 3, 5, 7\) respectively. The closed contours represent the model solutions 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/-/A/ and /a/-/u/.

*Fig. 1. An example of a general n-tube with segment areas \(S_i\).*

*Fig. 2. Goodness of fit of the present model in terms of the SP value. The heavy line connects all the found maxima, b shows some possible continuations.*

*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 efforts) have been rescaled to values between 0 and 1. They give an indication of the preference of the corresponding combination of short vowels.*

*Fig. 4. The space without predefining a fixed boundary in the formant space to show decreasing first formant frequency.*

3. RESULTS OF THE MODEL

In the figures 4, 5, and 6 we give the present model solutions of N = 3, 5, and 7 respectively. The closed contours represent the model solutions of the articulatory effort function da. One observes the following preferences:

- the preference for vowels along the lines /a/-/A/ and /a/-/u/.
- the limitation of the available vowel space without predefining a fixed boundary in the formant space.

236

Se 12.2.2

Se 12.2.3

237
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 on 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