VOWEL SYSTEMS

Louis ten Bosch

1. INTRODUCTION

In this paper we will deal with natural vowel systems and their structure in phonetic and, to some extent, phonological sense. We will discuss some of the literature with respect to vowel systems. In the last section we aim to give an introduction to the method that we are developing in ZWO project no 300-161-030: 'A model for the description of the structure of, and dynamics within vowel systems' in order to find underlying generating principles of natural vowel systems and a way to actually generate those systems.

1.1 Restriction of the subject.

The present topic deals with the distribution of vowels in languages, in other words, the structure of vowel systems of languages. Such a system, containing stationary oral vowels as well as diphthongs and nasal vowels, might be considered from different view points, such like phonetical or phonological ones. These views differ as to the level of inspection and abstraction.
Our interest will focus on principles underlying the structure of 'natural' (or 'unmarked') vowel systems in general, and on the relation between those 'natural' vowel systems and 'actual' vowel systems. Natural vowel systems differ from actual vowel systems in containing the 'underlying' vowel phonemes instead of the actual phonetic vowels (which may be allophones of an underlying phoneme). This distinction is relevant with respect to the possibility of describing firstly the structure of vowel systems without using language specific details.
With respect to vowel distributions one may pose two questions related to the abstraction level meant above. One question involves the explanation or prediction of the structure of unmarked vowel systems. For instance, which phonemes can be expected in a vowel system having five elements, regardless of language-specific phonetic phenomena. Another question involves the prediction of the existence and quality of (for instance) allophones while using knowledge of those phonetic phenomena.
In this project an attempt is made to give an answer to both questions. The method will be based upon the use of extra-linguistic physiological and acoustic methods mainly.

Throughout this paper we will always deal with the oral 'subsystem' of a vowel system, consisting of all oral stationary vowels, the schwa included. The vowel space is defined to contain all articulatorily possible realizations of vowellike sounds. A vowel system can therefore be interpreted (and we will use this interpretation) as a finite subset of the vowel space.

1.2 Diversity and regularity in vowel systems

Considering vowel systems of several languages, one observes many typological differences. Of the quantitative differences, we mention the great diversity in the number of elements of a vowel system: some languages have very few vowels (less than four: e.g. some Eskimo and Arabic dialects), others have twelve vowels, or more (Dutch, Frisian). Crothers (1978) shows that many (about one fourth) of his sample languages have five vowels (fig 1).

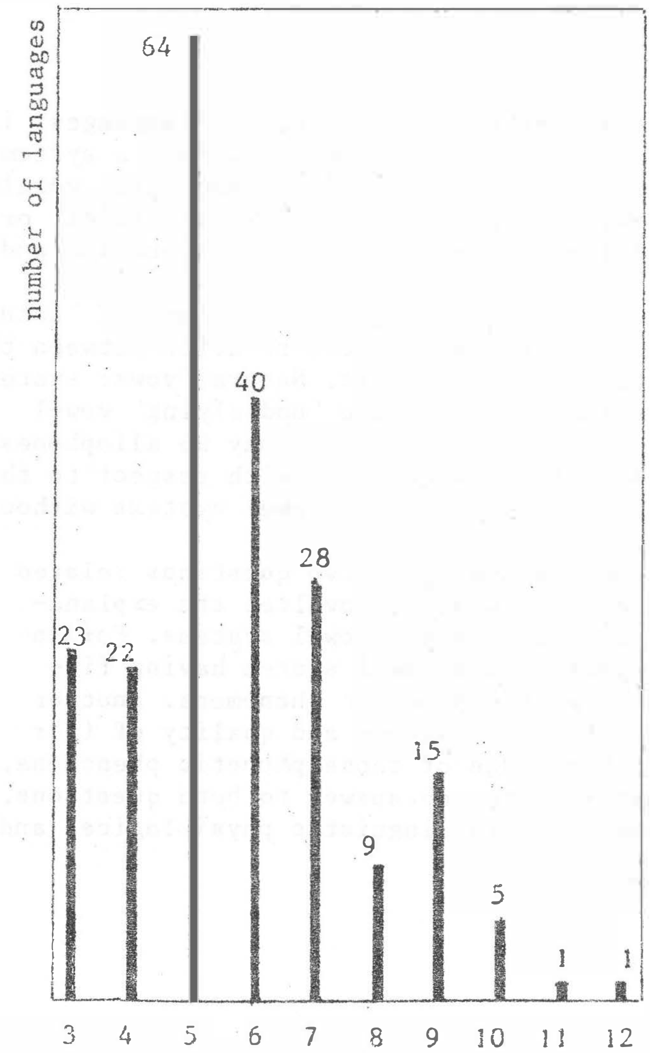


Fig. 1.
Number of languages related to the number of vowels in the vowel system. (Crothers, 1978)

Vowel systems also differ qualitatively, for instance, /y/ belongs to the Dutch, Swedish and German vowel system, but English and the Romance languages Italian, Spanish and Portuguese do not have /y/.
Of course, vowel systems show regularities, sometimes called (experimental) 'laws' or 'universals', as well. For instance: /u/, /i/ and /a/ are elements of nearly all vowel systems. One sometimes constructs a "vowel tree" of a sample of languages: a hierarchical model in which the level of a vowel is determined by its frequency of occurrence in this sample. Crothers (1978) constructed a vowel tree belonging to a sample of 209 languages (shown in fig 2), which suggests a sort of general vowel arrangement. For more peculiarities of vowel systems we refer to Crothers (1978) and Disner (1980, 1983).
We will focus our attention on the following question: Why do vowel systems have these particular properties, in other words: what are the rules (if any do exist) which describe vowel systems?
In the next section we will first deal with ideas and possible answers and methods as found in the literature. After that an introduction to our approach will be given.

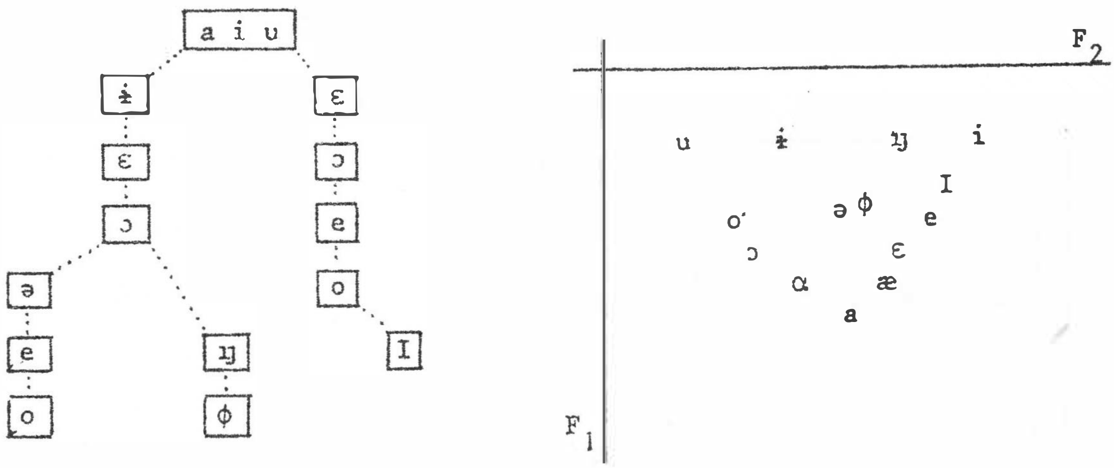
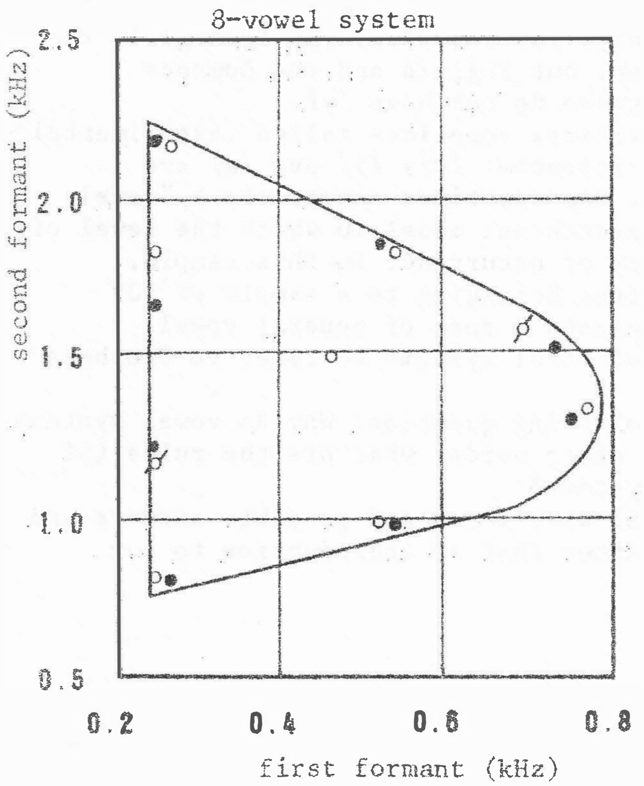
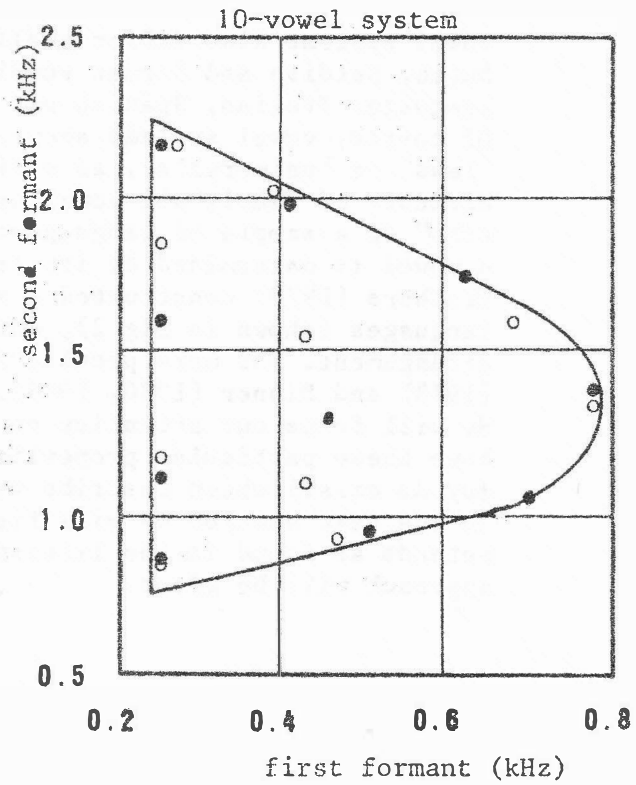


Fig. 2.
Left: Crother's hierarchical vowel tree. At the top the most common vowels are shown, at the bottom the least common ones. Right: a reference vowel triangle.





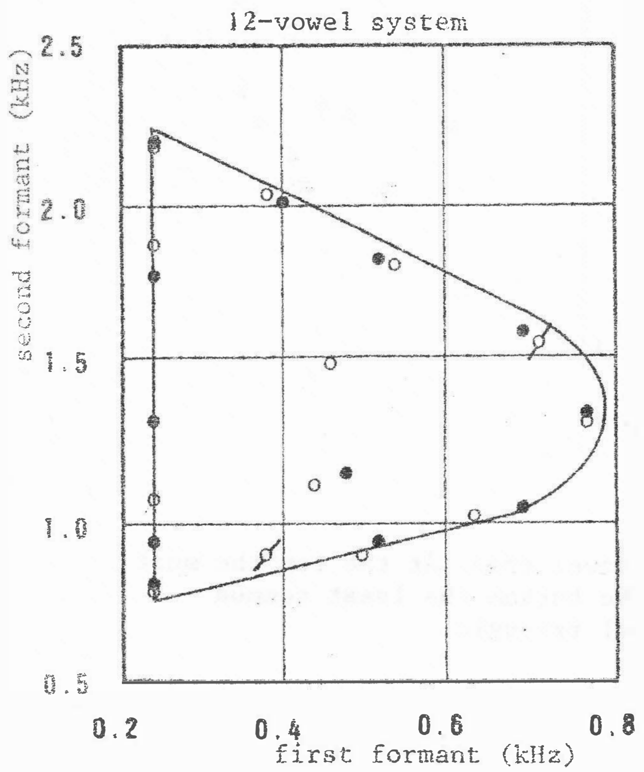


Fig. 4.
Discrepancy between 8-, 10- and 12-vowel systems according to Liljencrants' and Lindblom's model (•) and the data from Trubetzkoy, Hockett and Sedlak (o). ⌀ positions are optional.

1.3 Historical survey. Classifications

In history, much research has been done on vowels, originally mainly descriptive with respect to physiological properties. Already Hellwag (1781) gives a classification of vowels, according to the place and degree of articulatory constriction.
Since Hellwag many attempts have been made to give other classifications of vowels. Trubetzkoy (1929) classifies vowel systems to geometrical properties of vowel configurations: /u, i, a/ would then be triangular, /u, i, e, a/ quadrangular, and so on, in a space with jaw opening (height) and tongue body position (backness) as coordinates. Even 'linear' (one-dimensional) systems appeared in his classification (fig. 3).
In more recent years, when appropriate instruments (spectrograph, recording apparatus, computer) became available, a differentiation by acoustic, rather than physiological properties became possible, but also purely phonological classifications have been perfected more and more (cf. Chomsky and Halle, 1968; Ladefoged, 1971). For surveys we refer to Ungeheuer (1962) and to Jakobson and Waugh (1979).

While considering several vowel systems one observes that its elements in general never lie near to each other in formant space, provided that there is no other type of opposition (like duration, nasality) between them. Many vowel systems are therefore likely to obey rules with respect to distribution in the formant space. In this paper we will deal with some ideas from the literature concerning this vowel distribution, and with a global set up for our approach. This approach is based upon a method used by Liljencrants and Lindblom and uses two extra-linguistic presuppositions.
In 1972, Lindblom and Liljencrants investigated for the first time numerically the behaviour of a vowel system with respect to a particular (physically inspired) 'dispersion rule'. For N from 3 up to 12, they considered a set of N points, all to be placed within a (fixed) bounded formant space of dimension 2. They defined the **acoustic contrast** between two vowels as the squared (euclidean) distance between these vowels in the formant space. An appropriate vowel configuration was found by shifting all the vowels more or less systematically and in that way maximizing the acoustic contrast K of that configuration, K being an expression of the contrasts between all vowel pairs in the configuration in question.

Acoustic contrast
The principle of maximal contrast was introduced by Passy in 1890, and used, since that time, especially in a typological sense. Liljencrants and Lindblom tried to surpass the purely typological level by looking for an underlying generating acoustic principle. (Chomsky and Halle (1968) and many others did the same in the phonological sense.)
In recent years several improvements and adaptations have been proposed to refine this acoustic contrast principle (e.g. Lindblom, 1975; Crothers, 1978; Lindblom, 1981; Disner, 1983). Some of these alternatives still only deal with distribution on the basis of acoustic contrast, other ones incorporate articulatory properties of vowels. Articulatory constraints do

a priori influence the position of vowels in vowel space relative to each other: sound realizations are limited by physiological constraints. Liljencrants and Lindblom suggested in the final discussion section of their paper that the vowel system of a language is determined by: (a) the number of monophthongs in the vowel system, and (b) the premiss that the overall acoustic contrast is maximal while the overall differences in articulatory positions are minimal.

1.4 Goals of this project

Following the suggestion of Liljencrants and Lindblom (1972) we consider a combination of acoustic and articulatory properties of monophthongs, in order to find extra-linguistic generating principles of natural vowel systems. We will however not fully adopt the notion of "articulatory contrast" (see below: assumption 1 and paragraph 1.6). This research should lead to a model that predicts the structure of a vowel system from the number of its elements. This prediction is not likely to be accurate in case of all existing languages - the interpretation of results must be found in a more probabilistic sense. The model can only generate vowel systems up to that abstraction level that is not involving language-specific details.
To give a classification of natural vowel systems at a phonemic level to the number of the present vowel phonemes, and to verify the agreement of the probabilistic model results in case of a few particular languages (at this stage taking language-specific factors into account) will be the first goal of this project. We will make use of a vowel dispersion principle, modified and extended by implementation of articulatory principles (cf. Liljencrants and Lindblom, 1972; Disner, 1983).
The investigation how vowel transitions can be described while using a vowel distribution model is the second goal of the present project. Vowel transitions (e.g. diphthongs) can be seen as paths in the vowel space. Possibly diphthongs may be considered as solutions of problems with respect to searching 'shortest' or 'easiest' paths in the vowel space (cf. Almeida et al., 1977; Goldstein, 1983).

1.5 Setting of the problem. Assumptions.

In order to give an idea of rules acting on and properties of vowel systems we consider first a few examples. These examples often lie between phonetics and phonology. Then we will put down two assumptions from which this project starts.

I. Vowel harmony.
The term vowel harmony (or synharmonism) is given to the phenomenon that vowels appearing in adjacent syllables (more exactly: in the same phonological word) may share certain phonological features, in other words, may (tend to) agree in articulation place, degree of constriction (especially in front/back position (Fischer-Jørgensen, 1983)). Well known

examples are shown in German: all the Umlaut effects can be seen as consequences of vowel harmony (e.g. Jahr → jährlich, gut → gütlich: /a/ → /e/ and /u/ → /y/ as consequence of the presence of the /i/ in the suffix "lich"). Similar examples are present, for instance, in most of the Turkish languages, in Finnish, in Korean, in many African languages (cf. Jakobson and Waugh, 1979).

II. Vowel shift (and permutation).
Other examples show that in history, vowels may shift or permutate their positions in a vocabulary, and, at the same time, in the vowel set. Vowel shifts were common in Western Germanic: Old and Middle English, Dutch and German in the Middle Ages, and in many other languages; in some cases as a result of "interference" between more vowel systems (cf. Brosnahan, 1957; Prins, 1966; Schane, 1973; Jakobson and Waugh, 1979; Hoppenbrouwers, 1982; Ohala, 1983; Van Zanten and Van Heuven, 1984).

III. Vowel reduction.
Well known examples of vowel reduction are present in the sequences (Eng) competing competition, (Fr) fatigue → fatiguē, (Du) profeet → profetes, in which the underlined vowel at the left side of "→" is reduced to the underlined vowel at the right side of "→". Reduced vowels are more centralized than their unreduced originals.

Assumption 1
These examples show the independence of vowels in their behaviour as "entities" in a vowel system (they still may shift from original positions) on the one side and their dependence of their position in the vowel space with respect to the whole vowel system (vowels may be pushed away by other adjacent vowels) on the other side. They further show that vowel systems are not necessarily 'stable' or fixed but support the idea that vowel systems are subject to different tendencies. Some of these tendencies will be imposed by non-linguistic principles: the physiological principle of least articulatory effort in running speech and the perceptual principle of maximal (or sufficient) contrast. With respect to vowel phonemes one may expect that still the second principle holds: the principle of maximal or sufficient acoustic contrast is a paradigmatic one. But, although a priori no principle of minimal articulatory effort exist for underlying vowel phonemes (because this principle is originating from running speech constraints), articulatory limitations must be imposed on vowel phonemes too. This idea leads to the first assumption: the articulatory effort per vowel in a vowel system is minimal.
The speech-originated principle of minimal action is recognized by numerous authors. Zipf introduced, in the thirties and forties, a psycho-biological theory of minimal articulatory action (cf. Zipf, 1949). This principle (or an adapted version) is often used with respect to coarticulation or assimilation effects in running speech (Ohala, 1983).
Assumption 1 specifies the principle of 'minimal difference between articulation positions of vowels pairwise', which is reduced from the principle of 'minimal articulatory effort' (Lindblom: articulatory 'synergy'), to the principle of 'minimal articulatory effort per vowel'.

Assumption 2
On the other hand, the perception of vowels and perceptual distinctions between them can to a high extent be related to the formant positions of vowels and differences between these positions, or to differences between vowel (bandfilter) spectra (Plomp, 1970; Klein, Plomp and Pols, 1973). Example II shows that vowels, if situated near to each other in the formant space, may exchange their positions in the vocabulary - and therefore also may be regarded as separate objects with respect to vowel distribution rules. A distribution principle controlling natural vowel systems deals doubtlessly with at least distribution with respect to formant frequencies or spectra in order to guarantee a clear perception (assumption 2). Assumption 2 implies a relation between perceptual and acoustic differences and requests (a) quantification of the acoustic distance between any two vowels, (b) an overall measure for vowel dispersion in a vowel system.

One can expect that while 'minimizing the articulatory effort', speech utterances will more and more 'converge', which results in a (still to define) 'contraction' in the vowel space. This contraction will compete with the dispersion, the 'divergence' of the vowels, a consequence of conditions implied by a satisfactory perception (a perception characterized by minimal confusion). In other words, assumptions 1 and 2 deal with opposite tendencies of vowel configurations in vowel space and therefore involve a sort of balance between them. We, in addition, should assume that compromise solutions do indeed exist after adjustment of the involved balance (a sort of paradigm statement, a meta-assumption).

1.5 Discussion paragraph

(a) An alternative assumption 1 may deal with the hypothesis that the articulatory changes in all (V,C)-strings, as actually realized in running speech, are to be minimized. The V's are assumed to originate from the ideally predicted vowel 'archi' phonemes and are to be 'properly embedded' between (clusters of) consonants, that means: embedded such that the articulatory differences $C_b \rightarrow V$ and $V \rightarrow C_a$, with C_b and C_a defined as the neighbouring consonants just before and after V, are minimized. This approach involves more theoretical and practical difficulties than the present one. According to Lindblom (1981), the consonant system is important because its structure implies a mean articulation place and a mean constriction degree which influences the (mean) articulation place of vowels. In this project, however, we will not deal with consonant structures.

(b) In our approach, the articulatory difference with respect to the schwa is more important than other articulatory differences in the vowel system. Such a supposition might explain the phenomenon of vowel contrast reduction in running speech with respect to the schwa (cf. F. Koopmans-Van Beinum, 1980, 1983). This implies in our approach that two, distinctly argumented

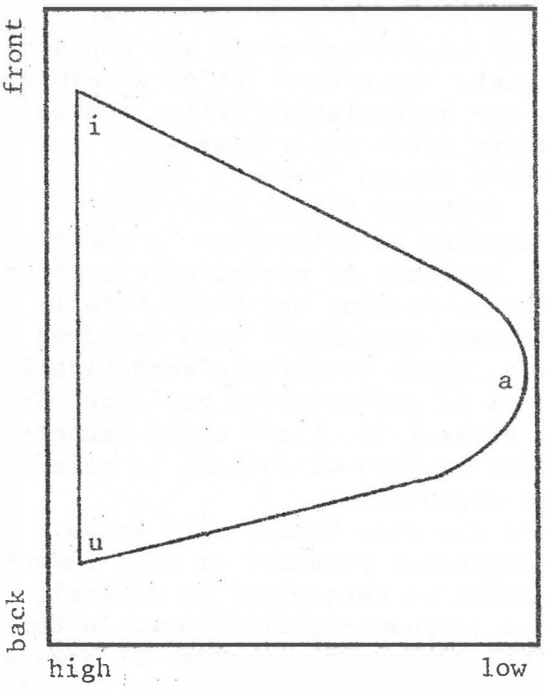
distribution rules should be used: one, perceptually based, ruling an overall (vowel, vowel)-dispersion, and a second, articulatorily based, ruling only the (vowel, schwa)-contraction. The principle of minimal overall articulatory contrast (Lindblom and Liljencrants) is now replaced by an minimal overall [tense]/[lax] contrast, 'tenseness' being an expression in articulatory parameters. In this way articulatory effort can be considered as a special case of articulatory difference, namely, by defining the effort as the difference with a chosen 'neutral' tube.
There is some support for this vision. In Crothers (1978) only "fully phonemic vowels" are considered (a full phoneme is defined as 'a phoneme with a severe distributional restriction' in terms of phonological, morphological or lexical environment). According to Stevens and House (1963), tense vowels "have substantially lower formant variances' than lax ones in the coarticulation process. In other words: tense vowels are acoustically more invariant than lax ones, both compared to articulatory variance. This probably indicates that the emergence of [tense] and [lax] typed members of vowel systems, and [long] and [short] typed members with them, is related to the balance dealt with in the paradigm statement.
The definition of the 'schwa' remains, for the time being, very vague. Whether it has to be considered as an independent phoneme, or as a reduced vowel originated from unreduced vowels cannot be determined in general; this question will probably appear to be a language-specific one. In this project the schwa has only a meaning in the articulatory sense: it stands for the acoustic output of the 'neutral tube', which is reference tube in the calculation of the articulatory differences (see paragraph 2.3).
For commentary on relations between vowel features and vowel systems we refer to Fischer-Jørgensen (1983).

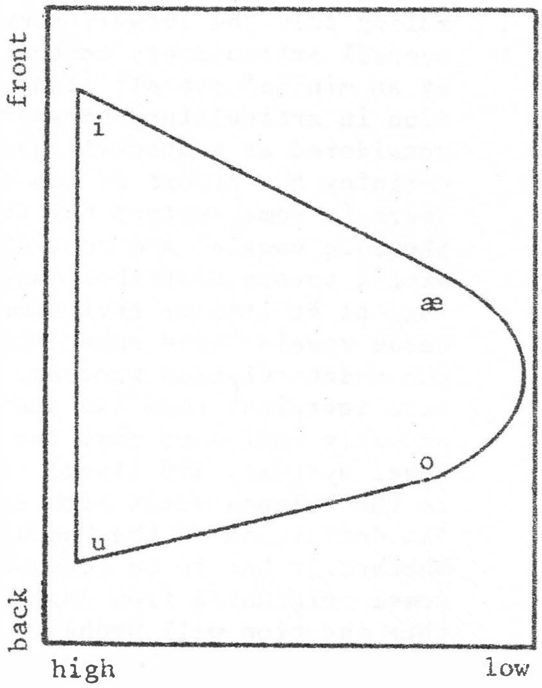
(c) We will not deal with nasalized vowels, mainly because there are less acoustic data of them, and secondly because it isn't quite clear how to handle them in an articulatory way. The acoustic impact of the nose cavities is a more difficult subject. One of the findings in Crothers (1978) (the number of nasalized vowels is roughly equal to four tenth of the total number of vowels in many natural vowel systems) may be explained by the fact that the acoustic differences between nasalized vowels are smaller than between the non-nasalized counterparts, while the oral articulatory difference between nasalized and non-nasalized vowels is negligible.

2. LITERATURE SURVEY

In this section we will give a survey of the literature dealing with the relations between articulation and acoustics. This survey is specified to:

(2.1) the relation between acoustic and articulatory properties of vowel systems,





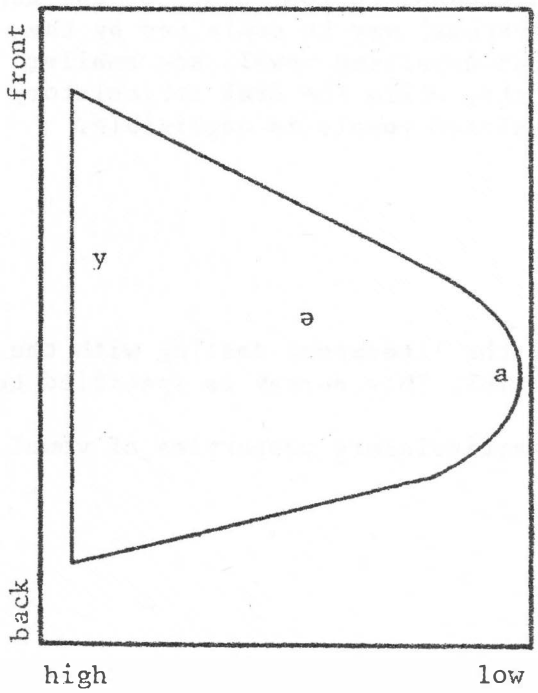


Fig. 3
Typology of vowel systems according to Trubetzkoy (1929).
Top left: triangular system.
Top right: quadrangular system.
Bottom left: one-dimensional ('linear') system.

(2.2) mathematical models of the vocal tract,
(2.3) articulatory effort,
(2.4) acoustic difference.

2.1 The relation between acoustic and articulatory properties of vowel systems

Liljencrants and Lindblom (1972) consider a vowel dispersion model, in which a (physically inspired) acoustic contrast is maximized. The vowel space is defined as the area bordered by a straight line and a parabola in the formant space, appropriately fitted to experimental data. Liljencrants and Lindblom evaluate the acoustic contrast K of a certain vowel system in the formant space in terms of a summation over all intervowel distances:

$$K = \Sigma \; d^{-2}(v_1, v_2) \qquad (1)$$

in which $d(v_1, v_2)$ is the (euclidean) distance between the vowels v_i in the formant space. The summation is taken over all vowel pairs in the vowel system. (They identify vowels v_i with the points in the formant space of which the coordinates are expressed in mels in order to meet with a perceptually inspired acoustic distance. The first coordinate relies to the first formant frequency, the second coordinate to a weighted average of the second and third formant frequency.)
For N = 3, 4, ..., 12 they compute only one N-vowel system, for which K is minimal (or, in other words, the 'overall" intervowel acoustic distance maximal). The numerical results are compared with natural N-vowel systems, as recorded by Trubetzkoy (1929), Hockett (1955) and Sedlak (1969). Fig. 4 shows an example of this comparison.
By their method, important methodological and numerical limitations are imposed on the solutions. Most of them were already acknowledged by Liljencrants and Lindblom in their discussion paragraph:
(a) the use of a fixed boundary of the vowel space in their model may influence the found configurations very unfavourably;
(b) the dimensionality of the formant space is limited to 2;
(c) actual existing languages show that the number of existing N-vowel systems increases with N (cf. Disner, 1980). Liljencrants and Lindblom simply calculate one solution for N = 3, 4, ..., 12, but it is not clear whether that particular N-point configuration is the unique minimizing solution of their problem, or not (and even whether uniqueness exists, or not);
(d) their results do not well account for the emergence of the so called "interior" vowels (cf. Crothers, 1978). In figure 5 is denoted the number of interior vowels versus the total number of vowels in a mean vowel system, according to the language sample in SPAP (the Stanford Phonological Archiving Project), as well as the model of Liljencrants and Lindblom

(1972) and the model of Crothers (1978). One may observe the differences between the models with respect to the interior vowels for N = 5, 6 and 7.

Predictions of vowel systems on the basis of perceptual contrast often lead to systems with relatively many 'high vowels', i.e. vowels with low F_1 (Lindblom, 1975; cf. Terbeek, 1977). Such systems are very uncommon. Some improvement is attained by re-scaling of the axes of the formant space (cf. Lindblom in his later work; Ladefoged, 1975; Chomsky and Halle, 1968). Following the discussion in Disner (1983): "(...) the proper re-scaling factors, based on a thorough understanding of perceptual mechanisms, remain to be discovered'.
Disner (1980), evaluating several types of vowel systems, evenly spaced as well as "lacking" one or more vowels, claims that the great majority of the defective natural vowel systems obeys a vowel dispersion rule with respect to the remaining vowel space.

Lindblom (1975) suggests an alternative interpretation of the terms in formula (1), namely to consider them as confusion probabilities. (This idea is based upon data from Nooteboom (1968) and Pols, Tromp and Plomp (1973). They find a relationship between intervowel confusion probabilities and the acoustic distance between vowels.) Lindbloms modification leads to an alternative, namely to replace maximalization of the overall acoustic contrast by minimalization of the (overall) confusion, or by diagonalization of confusion matrices. In this case the concept of acoustic contrast may, of course, be asymmetrical (this means that $d_F(v_1, v_2)$ and $d_F(v_2, v_1)$, where d_F stands for the acoustic distance between v_1 and v_2, may be unequal).
An alternative is to use a theory of information transfer, rather then the more static principle concerning dispersion of vowel formant positions. Appropriate vowel systems are those systems that satisfy the condition of containing maximal information in the sense of the communication theory. For a historical introduction to this theory we refer to Shannon (1949).

In fact Lindblom (1975) uses another type of generating model for vowel systems: from a given N-vowel system a "following" (N+1)-vowel system is constructed by looking at favourite zones for the (N+1)th vowel, but still his method only uses acoustic contrast and has the disadvantage of being less flexible. Our model is profoundly different from his one, not using generating concepts of Lindbloms type. If one tries to predict vowel hierarchy, this type of generating concepts may be very useful.

In Lindblom (1975, 1979), the strict principle of "maximal" acoustic contrast is left in favour of a more flexible principle of "sufficient" acoustic contrast. As observed in Disner (1983), such a principle tends to be unfalsifiable, especially if it is not specified further. The problem of the somewhat greater tolerance of vowel positions in smaller vowel systems plays here an important role.

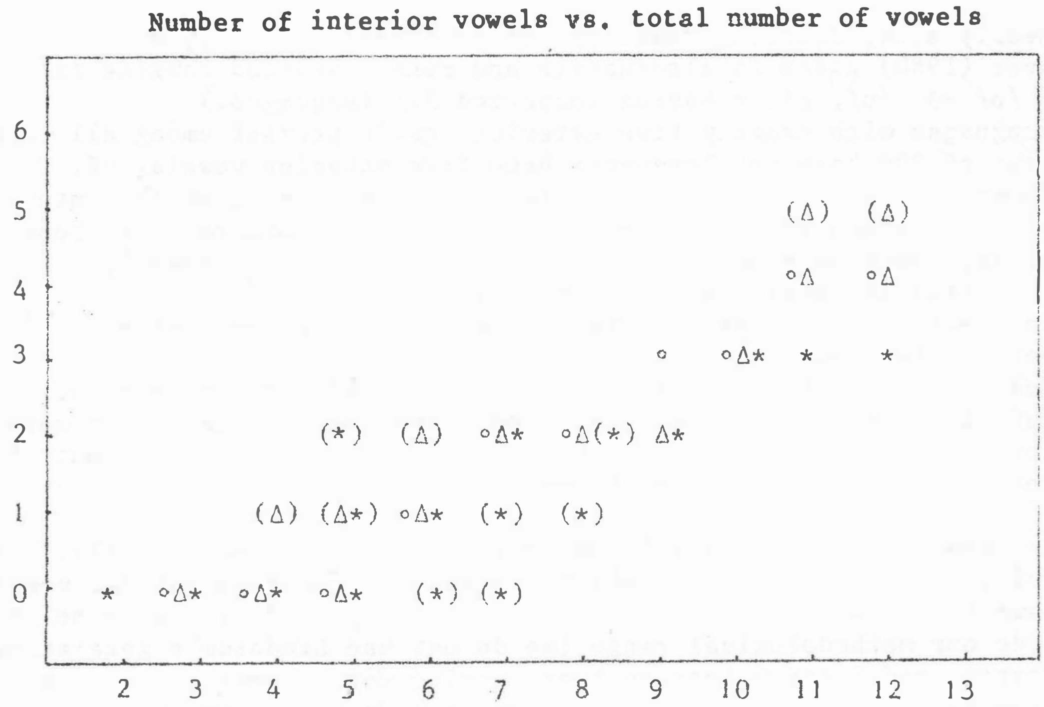


Fig. 5.
The number of interior vowels is plotted versus the total number of vowels, according to the SPAP language sample (*), Crother's model (Δ), and Liljencrants' and Lindblom's model (∘).

Crothers (1978) deals with a possible classification of natural vowel systems using the distinction between peripheral (exterior) and interior vowels (cf. fig. 5). His paper, based on work of Trubetzkoy (1929) and Hockett (1955), gives a more phonological approach to vowel systems, resulting in a list of 'universals' (experimental data), satisfied by (nearly) all languages in the Stanford sample SPAP, the Stanford Phonological Archiving Project. Some of them are (we only mention the main ones):

1. Nearly every existing vowel system contains /a/, /i/, /u/.
(Disner (1980) gives an alternative and more specified ranking /i/, /a/ → /e/, /o/ → /u/, after having inspected 317 languages.)
2. Languages with exactly five exterior vowels prevail among all languages (55 out of 209 "sample" languages have five exterior vowels, cf. fig. 1).
3. About one fourth of the sample languages has more than one interior vowel. (The schwa is only taken into account by Crothers if it represents an "independent phoneme", rather than 'only a neutral' vowel).
4. A particular vowel hierarchy, cf. fig 2.
5. In a natural vowel system, the number of nasals never exceeds the number of oral vowels.
6. Nearly half of the sample languages has long/short opposition.
7. Defective systems ('unbalanced' vowel systems missing one or more exterior vowels, such as /a, i, u, o/ or /a, e, i, u/) make about 15 percent of all vowel systems (Disner, 1980).

These experimental data still remain to be founded theoretically, but are useful to check with a model which pretends to generate natural vowel systems (perhaps we must exclude the points 4, 5 and 6: 4 because it falls outside our methodological range (we do not use Lindblom's generating concepts), and 5 and 6 because they involve more phonetic features than are covered by the articulatory or acoustic concepts used by us so far).
One should, of course, be careful in deriving conclusions from a comparison between numerical results of a model and global (field) data in literature: the specifications of vowel systems in natural languages, as described in phonological data libraries such as the UCLA Phonological Standard Inventory Database (UPSID) and SPAP are, acoustically, rather poor. However, phonological/phonetic data as recorded in Maddieson (1984) and Ruhlen (1976) will be useful to find systematic patterns in vowel systems.

Crothers (1978) introduces an adapted version of the Liljencrants and Lindblom model: a 'coin' model, in which vowels inside the vowel space are considered as having hard cores, with minimized overall perimeter of the obtained configuration. This method shows some improvement, especially with respect to the emerging of interior vowels, compared with the method of Liljencrants and Lindblom (fig. 5).

Quantal Theory
The 'Quantal Theory' in its original form (Stevens, 1972) states that all phonemes can be principally found in those articulatory regions ("pla-

teaus") in which the acoustic differences, as noticed by the auditory system, are relatively small compared to the articulatory differences. Stevens exemplifies this statement by means of a (two-)tube model and considering changes of the formant frequencies resulting from perturbations of the tube shape. The Quantal Theory predicts fixed positions in the vowel space for the vertex vowels /i/, /a/ and /u/, which are therefore named 'quantal'. Since 1982 this theory has been modified to one, stating 'that some non-quantal vowels are more or less evenly distributed between the quantal vowels in a vowel system' (Disner, 1983). A modification of the original theory is presented by Pisoni (1976); the latter states that the vowels /u/, /i/ and /a/ are articulatory more stable than other vowels, just because of their specific articulatory profile.

One may suggest that the Quantal Theory can figure as theoretical background in our project as well. However, it will be difficult to deal with its statements (which remain rather vague) and therefore difficult to operationalize Stevens' concept of 'relative' stability of phonemes. Furthermore, the claim of acoustic stability of the vertex vowels lacks convincing proofs (Disner, 1983). For a discussion on the link between articulatory variability, articulatory constraints and the imposed vowel shifts we refer to Goldstein (1983).

Let us consider an actual vowel system in which the long/short opposition plays a role in order to see what sort of problems we may meet. As an example we consider the Dutch monophthong system: /a, e, o, ø, i, y, u, œ, ɛ, ɔ, I, ɑ/ and the schwa.
In this system, the vowel pairs /o/-/ɔ/, /ø/-/œ/, /a/-/ɑ/, /e/-/I/ are generally recognized as long-short opponents. The phonological descriptions of this particular system use oppositions like tense/lax, long/short, to be able to classify vowels as eventually being each others counterparts (cf. Booij, 1981). In general, the perceptual difference between vowels v_1 and v_2 is of course not only relied to the difference between formant positions of v_1 and v_2, but is also dependent on the duration, tenseness or other phonological features of v_1 and v_2. Because of this, not all the acoustic differences between vowels (as members of a vowel system) should be weighted equally. It may therefore be convenient to limit the impact of a vowel distribution rule to vowel pairs or vowel **classes**, after having introduced an equivalence relation on a vowel system.

2.2 Mathematical models of the vocal tract.

We will review our vocal tract model concerning the mathematical relation between articulatory and acoustic/perceptual concepts.
The voice production mechanism can be modelled as consisting of a power source (lungs, muscles), a vibrator (the glottis), a resonator (mainly oral and nasal cavities), and a radiator (opening at the end of the vocal tract). All of them are physically coupled. In this project we mainly deal

with the resonator part.

Glottal spectrum (source spectrum).
The glottal spectrum may be considered as input spectrum of the resonator, the n-tube. The overall pressure spectrum envelope has an exponential decay of about 12 dB/octave (cf. Flanagan, 1958; Boves, 1984). The glottal incitation of the tube may be translated mathematically into 'linear coupling'.

Resonator.
In this project the resonator (vocal tract) is represented by a n-tube (a tube consisting of n coupled straight cylinders of equal length), rather than by electrical circuits, or other models. This choice makes it clearer how to define articulatory differences as a function of the tube shape variation, and makes it easier to put a link to another project of which the n-tube was the central theme (ZWO project no. 17-21-08, Van Dijk/Bonder). Moreover, the n-tube is a generally accepted model in literature. The tube, with or without damping (and sometimes considered together with extra cavities, modelling the nose), is well suited for straightforward calculations (Flanagan, 1958; Landau and Lifschitz, 1959; Flanagan, 1972; Atal et al., 1978; Butler and Wakita, 1982; Bonder, 1983a; Milenkovic, 1984).

The tube transformation and the output spectrum.
The tube transformation formula describes the relation between the source spectrum and the output spectrum of the tube in terms of physical quantifiers determining the tube (shape, length, damping) (the tube parameters). The tube transformation can be expressed in terms of the tube parameters by means of the Webster differential equation (Ungeheuer, 1962). The inverse transformation, i.e. the calculation of tube parameters from the tube transformation, is far more difficult to describe and is often treated numerically. Some of the literature used in this project deals with the computational aspects of the inverse tube transformation when using large computer facilities (cf. Atal et al., 1978).
The tube output spectrum depends on the tube input (glottal output) spectrum, the tube transformation, and the radiation at the lips and the nose. For the time being, only the tube transformation is emphasized.

Internal spectrum (cochlear spectrum).
For completion we mention the 'inner ear transformation' with the incoming external sound spectrum as its input and the '(internal) cochlear spectrum' as its output. We deal with simple transformations simulating the far more complicated inner ear transformation (by using logarithmic scaling, for instance), but ignore aspects concerning band sensitivity or spectral resolution of the ear. It seems plausible that perceptual differences are strongly connected with internal spectral differences.

While articulatory differences are measured from the shapes of different n-tubes, acoustic differences are measured from the behaviour of the tube transformation. In this way we introduce an eigenvalue problem rather than a resonance problem.

2.3 Articulatory effort. Our choice.

Several methods are available to quantify the concept of articulatory difference, which, as said above, implies the quantification of articulatory effort. Differences in vowellike sounds are mainly due to two parameters: the position of the tongue body and the protrusion at the lips (Booij, 1981). Already Lindblom and Sundberg (1969, 1971) deal with the lower jaw/tongue body protagonist principle, which implies a correlation between jaw opening and degree of constriction in the vocal tract by the tongue. Lindblom (1975) and Harshman et al. (1977) both deal with tongue shapes; the former defines a type of positional contrast of the tongue body (a polar model), the latter deals with tongue shapes using principal component analysis. The polar articulation model may become part of a useful alternative for the present qualification of differences between vocal tract shapes. Lindblom (1981) suggests an improved version of his 1975 model, relating the mean articulation place of consonants with that of vowels. Charpentier (1984) restricts the number of articulatory parameters by using parameters which specify the tube shape globally, instead of independent segment diameters of a tube.
Our definition of articulatory contrast will be directly related to differences between the shape of two n-tubes. Bonder (1983a) defines the articulatory contrast of a n-tube $S = (S_1, S_2, \ldots, S_n)$ with a n-tube $S' = (S'_1, S'_2, \ldots, S'_n)$ to be

$$d_A^2(S, S') = \sum_{i=1}^{n-1} ((S_{i+1}/S_i) - (S'_{i+1}/S'_i))^2. \qquad (2)$$

This formula figures as starting point in our model. The formula relates indirectly the minimization of tube shape differences with respect to the straight tube with the minimization of the "vocal tract opening degree" S_n/S_1. Alternatives are implied by some ideas found in Lindblom (1975) or Fujimura (1984); they promote consideration of separate positions of articulatory organs. These and Charpentier's ideas can for instance be met with by introducing a weighted sum in formula (2).

2.4 Acoustic difference. Our choice.

The acoustic difference of vowels could be evaluated in two basically different ways:
- by using the formant positions of vowels,
- by using whole spectra of vowels.

Gray and Markel (1976) give a summary of possible expressions for the difference between vowel spectra. Some of them are based on assumptions concerning the frequency distributions in speech signals. For the time being, however, we use a definition involving the second type:

$$d_F{}^2(v_1, v_2) = \sum_{i=1}^{2} (\log F_{i,v_1} - \log F_{i,v_2})^2 \qquad (3)$$

in which log denotes the natural logarithm, and F_{i,v_j} stands for the frequency of the i^{th} formant of the j^{th} vowel.
Compared with other types of differences this choice will make it slightly easier to relate model results with available data in literature, because many phonetic vowel data (if they do exist) are given in terms of formant frequencies only. The logarithm is used to simulate the log-like frequency perception. In the course of this project we will use more advanced, spectral instead of formant expressions to define the acoustic intervowel distance, if necessary. Then also specific knowledge of the tube's input spectrum (the glottal spectrum) is needed.

3. SURVEY OF POSSIBLE METHODS IN THIS PROJECT

This chapter is a rather preliminary one. In this phase of the project several possible methods will be investigated. We will therefore only sketch the outlines of a solution.

In order to construct a computer model that links the acoustic and articulatory properties of natural vowel systems, one has two options:
1. to deal firstly with the set of possible tubes, and then with the calculation involving the tube transformation formula and the eigenfrequencies of the tube (Websters equation),
2. to deal firstly with the formant space, and then with the calculation involving the tube parameters.

Option 1 has the advantage of being theoretically very clear. On the other hand, it may still involve numerical difficulties and its implementation will be rather complicated.
Option 2 has the advantage of laying an accent on the acoustic output of the tube but needs additional constraints.

If n, the number of tube segments, is small ($n \leq 4$), the differences between both options are likely to be quite insignificant. For the time being we choose option 2 in order to get quickly an idea of the behaviour of the model. This means that vowel systems will be chosen at first. Several methods are available to search for (sub)optimal vowel systems or classes of such systems:

a. One could simply consider a great number of vowel systems at random and evaluate them (monte carlo method),
b. One could start from one vowel system and then shift the vowels in an appropriate way such that Q, a still to be specified 'quality factor' that has to be optimized, increases during this process,
c. One may combine these methods and use methods such like 'stochastic cooling'. This means that, in the starting phase of the search, several randomly chosen vowel systems are evaluated, but that during the searching process the dependence of the choice of vowel systems increases. At last the process will be highly constructive. Such a method is used by Bridle and Moore (1983).
Method c. combines the advantages of the methods a. and b. (For a systematic comment on these and other search algorithms we refer to Knuth, 1973.)

3.1 The mathematical and methodological model

From now on, d (with subscript A or F, indicating articulatory or acoustic (formant)) denotes a distance or "difference" and Q denotes the final ("quality") parameter of a vowel system (to be optimized).

The mapping from the set of (physiologically appropriate) tubes T to the formant space F is determined by the tube transformation formula and can be considered as a representation of the mapping from T to the vowel space. The former mapping plays an important role in calculations involving acoustic properties of tubes (cf. Ungeheuer, 1962; Flanagan, 1972; Markel and Gray, 1976; Atal et al., 1978; Butler and Wakita, 1982; Bonder, 1983a; Charpentier, 1984). In general, the mapping is not one-to-one. Uniqueness can be saved by considering additional conditions such like (4) or (5) (Bonder, 1983):

"d_A([t], neutral tube) is minimal" (4)

"opening degree of [t] is minimal" (5)

These conditions are examples of constraints which restrict the number of degrees of freedom in the set T of tubes [t] that are all mapped onto the same point in the formant space.

3.2 Programming phase

While programming the vowel system generation model, we intend to make the following steps:

Step I.
Firstly, several parameters that are important with respect to the followed method have to be chosen: the number of tube segments (n), the

dimensionality of the formant space, the boundary of the formant space, the expressions d_A, d_F and the quality factor Q. Q is some function which involves unordered vowel pairs (v_1, v_2) and tube pairs (S_1, S_2), like

$$Q = \sum_{v_i} d_F(v_1, v_2) / \sum_{v_i} d_A(S_1, S_2)$$

A few expressions for Q have been evaluated now. Until now the most appropriate one is unknown.

Remarks

- if n is chosen equal to 4, the tube transformation takes a simple form (Bonder, 1983a). In this programming phase n = 4 will be used. Eventually we will test the program for higher values of n and simultaneously restrict the number of independent shape parameters to about six;
- the dimensionality of the formant space (the number of formant frequencies involved in the calculations), is now chosen equal to 2 to make an appropriate link to Bonder (1983a);
- the boundary of the formant space is induced by the boundary of the set of appropriate tubes (the boundaries are mutually dependent),
- extra conditions can be formulated in order to guarantee the existence of the inverse tube transformation, such as those given by expressions (4) or (5).

Step II.

Secondly, N, the number of vowels in vowel systems, is chosen, after having defined a grid in the formant space. These N points ('vowels') are placed at random on grid positions.
d_F and d_A are computed for all vowel pairs figuring in the chosen vowel configuration, which implies an expression for d_F, d_A and Q with respect to the vowel system itself. This is done for a 'large number' of N-vowel systems (monte carlo method). This large number (being in the order of 10000 for N = 4 but increasing with N) should represent a reliable sample of all possible vowel configurations in the formant space using the defined grid in that space. In this way several optimal and sub-optimal vowel systems will be found (optimal with respect to Q). This 'sample' method will be followed by a more constructive search method, such as a tree structure analysis (branch and bound method) or a combination of several methods. In this way it is possible to examine more extensivily a relatively small number of vowel systems that were labelled as "interesting" by the monte carlo method. On the other hand, it remains possible to keep trace of the sub-optimal vowel systems.

In this search method Q is used as a parameter that has to be optimized while searching, in other words, Q induces an optimality strategy. Favourable configurations are those which satisfy certain conditions on Q, d_F and/or d_A.

These conditions are to be found by examining the results of the present model when the expressions for d_A, d_F, Q are changed.

Final remark
The model that we intend to use in our approach is not established yet. Ultimately, the expressions for d_F, d_A and Q may need substantial changes. Of course, one could try to enlarge the number of the sample of N-point configurations, involve more than two formant frequencies, or take n (the number of tube segments) larger than 4. Substantial improvements might be attained by slight modifications in the optimalization procedure. During the preliminary phase of this project, all these possibilities must have our attention.

4. ACKNOWLEDGEMENTS

This investigation is supported by the Netherlands Organization for the Advancement of Pure Research (Z.W.O.), project no. 300-161-030.
I wish to thank especially Louis Pols for his inspiration and many valuable suggestions, Louis Bonder for his helpful comments and both of them for reading earlier versions of the manuscript. Furthermore I am indebted to Florien Koopmans-Van Beinum, Tjeerd de Graaf, Jan van Dijk and Geert Booij for their commentary given in several discussions.

5. REFERENCES

Almeida, A., Drommel, R.H. and Fiedler, Th. (1977). A mathematical approach to the acoustics of diphthongs. IP Köln Berichte 7, 73-83.

Atal, B.S., Chang, J.J., Mathews, M.V. and Tukey, J.W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. J. Acoust. Soc. Amer. 63, 1535-1555.

Bonder, L.J. (1983a). The n-tube formula and some of its consequences. Acustica 52, 216-222.

Bonder, L.J. (1984). Between formant space and articulation space. Proc. of the Tenth Intern. Congress of Phon. Sciences Utrecht. Foris Publications, Dordrecht. 347-353.

Booij, G.E. (1982). Generatieve fonologie van het Nederlands. Het Spectrum, Utrecht.

Boves, L. (1984). The phonetic basis of perceptual ratings of running speech. Diss., University of Nijmegen, the Netherlands.

Bridle, J.S. and Moore, R.K. (1984). Boltzmann machines for speech pattern processing. Proc. IOA, Vol. 6, Part 4, 315-322.

Brosnahan, L.F. (1957). The sounds of language. An inquiry into the role of genetic factors in the development of sound systems. W. Hefter and Sons, Cambridge.

Butler, S.J. and Wakita, H. (1982). Articulatory constraints on vocal tract area functions and their acoustic implications. Paper presented at the 104th meeting of the Acoustic Society of America, Orlando, 1982. J. Acoust. Soc. Amer. 72, suppl. 1, RR5.

Charpentier, F. (1984). Determination of the vocal tract shape from the formants by analysis of the articulatory-to-acoustic nonlinearities. Speech Communication 3, 291-308.

Chomsky, N. and Halle, M. (1968). The sound pattern of English. Harper and Row, New York.

Crothers, J. (1978). Typology and universals of vowel systems; In Universals of human language; ed. J.H. Greenberg; vol. 2; Stanford University Press, Calif., 93-152.

Delattre, P.C. (1969). An acoustic and articulatory study of vowel reduction in four languages. International Review of Applied Linguistics in Language Teaching 7, 295-325.

Disner, S.F. (1980). Insights on vowel spacing: results of a language survey. UCLA WPP 50, University of California, LA. 70-92.

Disner, S.F. (1983). Vowel quality. The relation between universal and language specific factors. UCLA WPP 58, University of California, LA.

Fischer-Jørgensen, E. (1983). Vowel features and their explanatory power in phonology. Proc. of the Tenth Intern. Congress of Phon. Sciences, Utrecht. Foris Publications, Dor. 259-265.

Flanagan (1972). Speech analysis, synthesis and perception. Springer-Verlag, New York.

Goldstein, L. (1983). Vowel shifts and articulatory-acoustic relations. Proc. of the Tenth Intern. Congress of

Phon. Sciences, Utrecht. Foris Publications, Dordrecht. 267-273.

Gray, H.G. and Markel, J.D. (1976). Distance measures for speech processing. IEEE transac. ASSP 24. 380-391.

Groot, A.W. de (1929). Zum phonologischen System des Nordniederländischen. Donum Natalicium Schrijnen, Chartres.

Harshman, R., Ladefoged, P. and Goldstein, L. (1977). Factor analysis of tongue shapes. J. Acoust. Soc. Amer. 62, 693-707.

Hellwag, C. (1781). Dissertatio inauguralis physiologico medica de formatione loquelae; Tübingen. - Ned. vert.: Publ. 10, IFA, Universiteit van Amsterdam, 1967.

Hockett, C.F. (1955). A manual of phonology. IJAL memoir 11. Indiana University, Bloomington.

Hoppenbrouwers, C.A.J. (1982). Language change. A study of phonemic and analogical change with particular reference to S.E. Dutch dialects. Diss., Universiteit van Groningen, The Netherlands.

Jakobson, R. and Waugh, L. (1979). The Sound Shape of Language. Indiana University Press.

Knuth, D.E. (1973). Fundamental algorithms. The art of computer programming. Reading, Massachusetts.

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

Koopmans-Van Beinum, F.J. (1983). Systematics in vowel systems. In: Sound structures, studies for Antonie Cohen. Van den Broecke et al., eds. Foris Publications, Dordrecht. 159-171.

Ladefoged, P. (1971). Preliminaries to linguistic phonetics. University of Chicago Press, Chicago.

Ladefoged, P. (1975). A course in phonetics: Harcourt Brace Jovanovich, Inc., New York.

Landau, L.D. and Lifschitz, E.M. (1959). Fluid mechanics. Course of theoretical physics, Vol. 6.

Liljencrants, J. and Lindblom, B.E.F. (1972). Numerical simulation of vowel quality systems: the role of perceptual contrasts.

Language 48, 839-862.

Lindblom, B.E.F. (1975). Experiments in sound structure. Paper read at the 8th Intern. Congress of Phon. Sciences, Leeds.

Lindblom, B.E.F. (1981). A deductive account of vowel features. SCAMP Working Paper no. 18/81; Proc. Symp. Acoust. Phon. Mod., Williamstown, Mass.

Lindblom, B.E.F. and Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. J. Acoust. Soc. Amer. 50, 1166-1179.

Maddieson, I. (1984). Patterns of sounds. Cambridge studies in speech science and communication. Cambridge University Press.

Markel, J.D. and Gray, A.H. Jr. (1976). Linear prediction of speech. Springer-Verlag, NY.

Nooteboom, S.G. (1968). Perceptual confusions among Dutch vowels presented in noise. IPO Annual Progress Report 3, 68-71.

Ohala, J.J. (1983). The direction of sound change. Proc. of the tenth Intern. Congress of Phon. Sciences, Utrecht. Foris Publications, Dordrecht. 253-258.

Passy, P. (1890). Etudes sur les changements phonétiques et leurs caractère généraux. Paris.

Pisoni, D.B. (1976). Vowel mimicry and the quantal theory of speech: A first report. In: Research on Speech Perception, Progress Report no 3. Indiana University. 81-119.

Pols, L.C.W., Tromp, H.R.C. and Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. J. Acoust. Soc. Amer. 53, 1093-1101.

Prins, A.A. (1966). A synopsis of the history of English tonic vowels. Universitaire Pers, Leiden.

Ruhlen, M. (1976). A guide to the languages of the world. Language universals project. Stanford University.

Schane, S.A. (1973). Generative Phonology. Englewood Cliffs, N.J.

Sedlak, P. (1969). Typological considerations of vowel quality systems. Stanford University Working Papers on Language

Universals 1, 1-40.

Shannon, C.E. and Weaver, W.(1949). The mathematical theory of communication. Univ. of Illinois Press, second edition, 1963.

Stevens, K.N. (1972). The quantal nature of speech: Evidence from articulatory acoustic data. Human Communication: A unified view. Eds.: E.E. David Jr. and P.B. Denes. McGraw Hill Book Company. 51-66.

Stevens, K.N. and House, A.S. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. Journal of Speech and Hearing Research 6, 111-128.

Terbeek, D. (1977). A cross-language multidimensional scaling study of vowel perception. UCLA WPP 37, University of California, LA.

Trubetzkoy, N.S. (1929). Zur allgemeinen Theorie der phonologischen Vokalsysteme. Travaux du Cercle Linguistique de Prague 1, 7-36.

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

UPSID (1981). UCLA Phonological Segment Inventory Database. UCLA WPP 53, University of California, LA.

Zanten, E.A. van and Heuven, V.J. van (1984). A phonetic analysis of the Indonesian vowel system, a preliminary acoustic study. NUSA, linguistic studies in Indonesian and languages in Indonesia, 15, 70-80.

Zipf, G.K. (1949). Human Behaviour and the Principle of Least Effort. The Riverside Press, Cambridge, Mass.