# ABSTRACTS

## ON THE STRUCTURE OF VOWEL SYSTEMS, ASPECTS OF AN EXTENDED VOWEL MODEL USING EFFORT AND CONTRAST

*Louis F.M. ten Bosch*

Abstract Ph.D. thesis (defended on October 4th, 1991)

Vowels and consonants are the well-known speech units used in spoken language. In speech, vowels and consonants are indispensable; they play their own role in the construction of larger units such as syllables and words, and they obey their own rules. In this thesis, we consider the set of *vowels* more closely.

The set of vowels (or 'vowel system') is not the same set for every language. On the contrary, vowel systems show a large variety across languages. The variation is apparent with respect to the number of vowels as well as the timbre of vowels. In the literature, the number of vowels is reported to range from 3 up to 24. Moreover, languages differ with respect to the existence of vowel types: Some languages have short vowels only, other languages have short and long vowels, or diphthongs. Vowel systems may vary from 'easy' to 'difficult'.

Despite this variation, a broad inspection of all vowel systems shows much regularity. Some vowels occur more frequently than others: these seem to be 'basic' to any vowel system. Furthermore, vowels preserve a certain 'distance' from each other as if they combat their position in a competitive environment.

The problem addressed in this thesis is, how we can explain the structure of vowel systems by using properties of vowels that are related to their production (articulation) and perception. When we consider this question more closely, we observe that we are dealing with a relation between two types of description. The first type is phonological, i.e. deals with the linguistic function of vowels. For example, the phonological difference in Dutch between the vowels in 'poot' and in 'pot' is related to the fact that these words have in Dutch a different meaning. The words 'poot' and 'pot' are called a 'minimal pair' corresponding to the opposition between their vowels. It is possible to find such minimal pairs for any vowel pair in Dutch.

Besides the phonological representation of speech sounds, we can consider the phonetic representation. From the phonetic point of view, vowels (and consonants also) are described according to the way in which they are produced (which gives a description in articulatory terms of the vocal tract) or perceived (which yields a description in auditory or acoustic or perceptual terms, often in terms of 'formants'). At the phonetic level, differences between vowels can be described along a continuum, whereas the phonological difference between vowels are discrete. In the phonological description, the differences between the exact realizations of a vowel do not count. These differences, however, play a role in the phonetic description, where the vowel sounds may differ even between repeated utterances by one single speaker.

Phonological information about speech sounds ('phonemes') for a large sample of languages is included in phonological databases, such as the Stanford Phonology Archive (SPA) and the UCLA Phonological Segment Inventory Database (UPSID). These database contain descriptions of about 690 and 450 languages, respectively. Such databases can be used to find general phonological patterns in vowel systems.

In the past, it has often been attempted to set up phonetic models that describe the phonological structure of vowel systems in phonetic terms as much as possible. In other words, it was attempted to supply the phonological structure with an underlying phonetic structure. In these models, two principles are often used that are related to the process of vowel production and to the process of vowel perception. The articulatory principle can be paraphrased as the 'minimization of articulatory effort'. This principle is used to model a tendency of speakers to reduce their articulatory effort, in accord with the specific demands of the situation. Examples are: loose pronunciation in informal situations, reduction of vowels in unstressed syllables, inaccurate articulation at high speech rates, and simplification of difficult consonant clusters.

On the other hand, the perceptual principle can be paraphrased as 'sufficient vowel contrast'. This principle states that vowels should preserve sufficient contrast. It models the constraint on the listener concerning minimzation of confusion errors in speech. No language contains vowels that are phonetically very close.

These concepts are found to play a prominent role in the description of the structure of vowel systems, but the various models that have been designed all differ in detail as to how these concepts are to be quantified. We will refer to these models as 'vowel models'. A basic idea of many vowel models is that vowel systems tend to optimize their internal acoustic contrast and, at the same time, tend to minimize the required effort. (A simple model may clarify the procedure. Vowels are assumed to be representable as floating magnets in a water tank: the equilibrium position of the magnets would indicate the optimal position of vowels in the vowel space. The boundary of the water surface is given by articulatory constraints. The magnets move due to the perceptual claim of contrast enhancement.)

In these models, both principles are quantified and combined in one way or another; by an optimization technique, 'optimal systems' are found that optimally fulfil both principles. These optimal (phonetic) systems are then compared to phonological systems in natural languages. On the basis of that comparison, the principles used are rejected or are modified and improved.

A basic assumption in all these models is that the optimal vowel systems, as found by the optimization algorithm, can indeed be considered as 'optimal' prototypes for the systems in natural languages.

In this thesis we constructed a vowel model that is partially based on other vowel models proposed in the literature, but it deviates on essential points. Chapters 1 and 2 serve as a set-up for the model. The effort function, which assigns effort values to acoustical positions of vowels, is dealt with in chapter 3. In this chapter, we showed how to obtain an effort expression on the basis of acoustic information. In order to define effort, the vocal tract was modelled by an *n*-tube approximation. That approximation yields a simplified model for the human speech production apparatus. On the basis of an *n*-tube model, it is possible to calculate an effort value corresponding to any feasible formant position. The central problem in modelling the effort is how to get from a formant solution to the corresponding n-tube (known as the *inverse problem*). This problem is always solvable in a numerical way, as was shown by Atal and colleagues in 1978. In this chapter, we used a method, partly based on an approach described by L.J. Bonder in 1983, by which we can obtain analytical solutions in simplified cases that are nevertheless of considerable practical importance. Our approach, which will not be explained here as it is rather technical, allows an direct interpretation of the boundary of the vowel space, as found in languages, in terms of the effort function.

In chapter 4, the contrast between vowels was dealt with. First, inter-vowel distance was defined on the basis of the distance between formant positions. After that, a system contrast was defined on the basis of all inter-vowel distances in the system. By definition, optimal vowel systems globally optimize the system contrast.

This approach is, sometimes tacitly, followed in many vowel models. In this chapter, however, we have further elaborated this principle. A new approach to the inter-vowel distance was considered, and a new system contrast measure was defined that is to be optimized in order to obtain optimal model systems. We show that if these implementations are well-chosen, the optimization results show a satisfactory match with the phonological data. Our proposal for system contrast, which was denoted by $Q_5$, has the advantage of being much more interpretable than the commonly applied function $Q_2$. This interpretability question is relevant, as a complicated contrast function still does not give much insight into the actual structure of vowel systems. The function $Q_5$ has a much simpler structure than $Q_2$: it involves only one term (it denotes the minimum over a set of distances), while $Q_2$ involves a sum of squared inverses of distances.

We argue that the form of $Q_2$ does not reflect any essential physical property in the world of vowels; moreover, it makes it more difficult to relate the system contrast with any notion concerning confusion probabilities.

In chapter 5 we attempted to incorporate long vowels and diphthongs into our vowel model, thereby obtaining an 'extended' model. In this model, the acoustic distance (defined in chapter 4) is extended with a durational argument in order to be able to take durational oppositions into account. In the literature, insufficient data were available for an accurate specification of the articulatory and perceptual principle. We therefore attempt to handle the extension from a more logical point of view, aiming at the description of logical constraints that are to be met by the extended acoustic distance. In numerical experiments, we obtain results that are satisfactorily coherent with the available data concerning long vowels and diphthongs. It is suggested that the combination of durational oppositions (long versus short vowels) and dynamic oppositions (stable vowels versus diphthongs) reaches the explanatory limits of the vowel model in general.

In chapter 6, we briefly summarize extensions to the present model, and two other vowel models: the Quantal Theory of Speech (QT) and the Theory of Adaptive Dispersion (TAD). The Quantal Theory is a major opponent to our model, as this model suggests that vowel systems (actually, phoneme systems) are structured according to a totally different principle (Stevens, 1972, 1989). The main difference between QT and our model lies in the different combination of effort and contrast principles into one unified target principle.

There are no essential differences between our model and TAD. In fact, TAD is a general phoneme model of with our vowels represents a submodel. Such a general model, however, is difficult to implement, as it involves several speaker dependent, listener dependent and social factors.

As a final conclusion, we observe that the dispersion formalism enables us to explain the general structure of vowel systems to a large extent, however, not in specific detail without going into peculiarities of a particular language. We cannot expect to be able to describe a specific vowel system on the basis of general principles: the *third simple* vowel, a diphthong, is already clearly perceived in a language-dependent manner.

Future research must show whether more adequate optimization functions can be found, how dynamic effects and context are to be taken into account and how the theory can be brought into line with opposing theories such as the Quantal Theory of Speech.

## THE ROLE OF FOCUS WORDS IN NATURAL AND IN SYNTHETIC CONTINUOUS SPEECH: ACOUSTIC ASPECTS

*Florien J. Koopmans-van Beinum*

Abstract of paper to be published in:
Journal of Speech Communications
Special Issue on Phonetics and Phonology of Speaking Styles: Reduction and elaboration in speech communication
Guest editors Joaquim Llisterri and Dolors Poch
ETRW, Barcelona 1991.

In everyday communicative situations not all parts of the spoken message are pronounced equally clear. Especially words bearing a high load of semantic information are put in focus by the speaker. The question of how this is realized in natural spontaneous and read speech, and whether resulting knowledge can be applied in synthetic speech to improve naturalness and acceptability, is subject of this study.

By introducing a 'peak-and-level' model we examined spectral and temporal aspects in focus and non-focus words from spontaneous speech material and from the same texts, read out after orthographic transcription.

Audio recordings were made of a professional male speaker, whose voice and pronunciation also served as a model for the diphone-based component of the Dutch national speech synthesis program. For a number of acoustic parameters it can be concluded that there is a clear difference, both in 'peak values' and in 'level values', between the two natural speech styles, but that the peak values display comparable contrasts to the level values in both styles.

The results of our measurements in natural speech were compared to the data of the same texts synthesized by the Dutch diphone text-to-speech system. In a pilot experiment, varying temporal aspects in the synthesized speech, listeners were asked to judge the naturalness and intelligibility in order to determine the starting-point for future evaluation of text-to-speech synthesis including peak-and-level contrasts.

# TEMPORAL ASPECTS OF THE VOICED-VOICELESS DISTINCTION IN SPEECH DEVELOPMENT OF YOUNG DUTCH CHILDREN

*Cecile T.L. Kuijpers*

Abstract of paper to be published in:
Journal of Phonetics

Several temporal phenomena have been examined in the speech of two groups of Dutch children. One group consisted of four-year-old children, the other consisted of six-year-old children and in both age groups six subjects participated. Intervocalic closure and burst durations of voiced and voiceless stops, as well as preceding vowel durations, were compared to study developmental patterns. Although the younger children produce longer segmental durations, relative differences in voiced and voiceless closure duration and burst duration seem to correspond between four- and six-year-old children. In the same way, relative durational differences between phonologically short and long vowels are produced in an adult-like way by these children. However, the temporal adjustment between vowel and consonant in the VC sequence displays a developmental trend. The data show that coordination of vowel and closure duration in the VC sequence with voiced context has been acquired at the age of four. However, the coordination of vowel and closure duration in the VC sequence with voiceless context, i.e. a relative shortening of the vowel, has not been acquired by the younger children. At the age of six the children seem to have obtained already this vocalic adjustment rule. The durational characteristics in speech of young children can be interpreted as developing from a 'syllable-independent' mechanism towards a 'syllable-integrated' mechanism with increase of consonantal influence across syllable boundary.