MODELLING VOWEL SYSTEMS BY EFFORT AND CONTRAST

L.F.M. ten Bosch

Institute of Phonetic Sciences, University of Amsterdam, The Netherlands

ABSTRACT

In the past two decades, several models have been proposed in the literature aiming at the phonetic description of vowel systems. These models are based on principles using constraints from vowel production ('articulatory ease') and/or vowel perception ('perceptual contrast'). In this presentation, we will discuss these theories and will attempt to relate their phonetic bases to more linguistic attributes of vowels.

1. INTRODUCTION

Speech serves as one of the most important means of communication between humans. It results from accurate regulation of the subglottal air pressure and, at the same time, manipulation of the glottal and vocal tract muscles. Phonemes such as vowels and consonants act as linguistic (phonological) units in a language, but at the same time, the corresponding allophones are subject to articulatory and perceptual demands. In phoneme models, the collection of consonants and vowels in a language is assumed to meet rules with respect to articulatory ease, and perceptual contrast and salience. We present an outline of the theories aiming at a structural description of *vowel* systems in relation to articulatory models. We will focus on two aspects of system structure: the internal structure, viz. the manner in which vowels are positioned in the vowel space, and the external structure, apparent in the boundary of the vowel space. Further, we pay attention to how phonological demands on vowel systems can be incorporated in sophisticated vowel models.

2. INTERNAL STRUCTURE

Vowels in language principally serve a linguistic goal. Their existence helps to distinguish words semantically, which is clear in the case of minimal word pairs. Historical linguistics and dialectology show that vowel systems must be considered as systems which are continuously in development, rather than as collections of vowels which are fixed once and for all. Vowels may change e.g. due to accent shifts or Umlaut-effects (as e.g. in Germanic languages), to whims of fashion (some cases of diphthongization), to ease of articulation (vowel reduction). A shift of one particular vowel may induce the shift of many vowels in the system (e.g. the Great Vowel Shift in Middle English).

From a phonological point of view, the static structure of vowel systems is related to the presence of features with an articulatory basis, such as [round], [front], [high]. Every vowel is coded by its specific feature values, and the structure of vowel sets can be represented by 'algebraic' manipulation on the set of feasible feature value combinations.

Phonetically, system dynamics can be modelled by repelling forces between vowels (yielding push chains) or by the tendency to fill system gaps (drag chains). These effects can be understood by assuming principles of 'sufficient perceptual contrast' or 'optimal contrast', respectively (Disner, 1983).

In vowel models, the actual linguistic vowel systems are assumed to optimize perceptual contrast and, in an extension, articulatory ease. Liljencrants & Lindblom (1972) were the first to implement a principle of optimal perceptual contrast in a so-called vowel dispersion model. In a 2D formant space, vowels were positioned such that the system contrast was maximized, by the minimization of the system quality Q:

$$Q = \sum \frac{1}{d(v_i, v_j)^2} \qquad (1)$$

where $d(v_i, v_j)$ denotes the (Euclidean) distance between any two vowels v_i and v_j in the 'perceptual space', and the sum is taken over all distinct vowel pairs ($1 \leq i < j \leq N$). The particular implementation chosen by L&L suffered from the drawback to generate too many high vowels for large N, due to a too large perceptual distance between /i/ and /u/.

One of the basic ingredients in this approach, viz. the perceptual distance between two vowels, has later been modified to more sophisticated submodels for the auditory spectrum (Bladon & Lindblom, 1981; Lindblom, 1986).

In the literature, the 2D inter-vowel perceptual contrast has been subject to further refinement and extension to 3D. The extension to higher-dimensional formant spaces is considered in Schwartz *et al.* (1989) and Ten Bosch (1991). These studies show the great dependency of the resulting model systems on variations in parameters controlling the perceptual distances between vowels. The best perceptual metric for nearby vowels has recently been reported to be the 2D Euclidean metric after bark transformation of F_1 and F_2 (Kewley-Port & Atal, 1989). Since their stimuli, however, were determined by two parameters only, this result must be carefully interpreted, leaving aside the question about the relation between the phonemic distance (that we search) and the phonetic distance (that they measure).

While d has been subject to continuous refinement, the *system contrast* Q, however, has not grown beyond the form

$$Q = \sum \frac{1}{d^2} \qquad (2)$$

d now involving combinations of transformed formant frequencies (Schwartz *et al*, 1989) or spectral differences (Lindblom, 1986)). The problem that we want to address here is that this expression Q is in fact very arbitrary, it being suggested by repelling forces between magnetic monopoles or dipoles, but lacking, in fact, any linguistic or even physical basis. Ten Bosch *et al.* (1987) propose an expression

$$Q = \prod (1 - \exp(-\alpha d)) \qquad (3)$$

the product being taken over all distinct vowel pairs, and α some scaling parameter. Q is to be optimized. The rationale is, that the factor $1 - \exp(-\alpha d)$ ($\equiv \pi(d)$) is interpretable as a probability of two vowels on a distance d not being confused. The system quality Q would then denote the probability of no confusion at all between any two vowels, under the assumption of independence of the probabilities involved. This idea has also been suggested by Lindblom in 1975. Also in this approach, however, a weak argument can be detected, namely that the resulting optimal vowel configurations can (easily) be shown to be dependent on the exact shape of $\pi(d)$ (Ten Bosch, 1991). Moreover, the probability of two vowels being confused is not based upon any *linguistic* consideration at all.

In Ten Bosch (1991), another expression Q has been elaborated:

$$Q = \min_{i \neq j} \{d(v_i, v_j)\} \qquad (4)$$

i.e. the minimum over all distances between distinct vowel pairs. Three advantages can be recognized: (a) the system contrast is related to a 'perceptual bottleneck' in the whole system rather than to global system properties: the bottleneck is then located at the location of the nearest vowels. (b) The influence of the exact shape of the relation between inter-vowel distance and inter-vowel confusion is apparent on exactly one place in the vowel system, rather than being spread out by weighting all inter-vowel distances (as is done in eq. 2). (c) Any sufficiency constraint of the system contrast is directly relatable to the minimal perceptual distance between vowels. The systems, obtained by optimizing eq. 4, are similar (but not equivalent) to the ones, obtained by minimizing eq. 2 (Ten Bosch, 1991). This yields, in my opinion, a strong argument for the latest modified Q (Ockham). Property (a) is particularly useful in numerical simulation of push and drag chains. In Ten Bosch (1991), it is attempted to explain the emergence of *diphthongs* as a consequence of a local high vowel density in the vowel space. Although this model fails to explain diphthongeal properties in detail, gross effects, such as the preference for diphthongs to have a relatively large trajectory, can be clearly demonstrated.

Articulatory constraints were not explicitly dealt with in these models: all calculations were carried out in the acoustic domain. Recent implementations attempted to combine perceptual and articulatory constraints (Bonder, 1986; Ten Bosch, Bonder & Pols, 1987; Ten Bosch, 1991). Other approaches were carried out by Abry, Schwartz, Badin, Boë, Perrier, Guérin (see the references) and colleagues in Grenoble. Stevens (1989) has put forward an elaborated version of the Quantal Theory (cf. Stevens, 1972), in which perceptual and articulatory constraints are combined into *one* principle. In these recent models, other points of view have been adopted (leading to e.g. the notion of focal points, articulatory plateaus, sufficiency instead of optimality), and more elaborated definitions of Q have been introduced (Schwartz *et al.*, 1989).

Ten Bosch *et al.* (1987) propose a vowel system model based on maximal acoustic contrast together with a minimal articula-

tory effort criterion, by minimizing

$$D_A^2 + S \cdot (Q-1)^2$$

where D_A is the total articulatory system effort, Q given by eq. 3, and S a slack variable as used in optimization problems (S being a large positive number). This combination of D_A and Q was left as too many parameters were involved in the optimization sessions. The search for a balance between D_A and Q turned out to be a Pandora's Box. We here leave aside the definition of 'articulatory system effort' and even forget the role of consonantal context in any definition of articulatory ease.

Another important goal is the refinement of the overall articulation-to-acoustics relation. The Quantal Theory (QT; Stevens, 1972, 1989) makes use of the non-uniformity of this mapping. In its pure form, QT states that the articulatory positions of which the acoustic output (to some norm) is less sensitive to articulatory deviations are favourable over other positions (articulatory plateaus). The Quantal Theory predicts, in the case of vowels, the corresponding favoured vowels to likely be a member of a vowel system. The presuppositions of the Quantal Theory, however, still lead to discussion and have been questioned by many authors (cf. Journal of Phonetics, vol. 17), whereas the results are not convincing (cf. e.g. Ladefoged & Lindau, 1988; Ten Bosch & Pols, 1989). It is generally believed, however, that the speech signal inherits 'quantal' phonetic properties as a consequence of non-linearities of the articulation-acoustics mapping and probably, the categorical perception of speech sounds. If quantality exists, it is probably a result of close approximations of formant frequencies (Stevens, 1989; Badin *et al.*, 1990; Schwartz *et al.*, 1989; Ladefoged *et al.*, 1988).

We briefly return to the open question of phonological enrichment of phonetic vowel models. An unsolved, and perhaps unsolvable, drawback inherent to phonetic models is that they cannot easily account for the linguistic demand for vowel contrast, although linguistic oppositions are ultimately based upon phonetic contrast. Is there a relation between the need of intervowel contrast and the 'lexical load' of the opposition? The relation between phonetic contrast and phonological contrast seems not to be derivable directly from the statistics on lexemes in a language. In Dutch, /ɑ/ and /ɔ/ have the largest (most often frequented) minimal set in common, despite they are a very close pair in the Dutch vowel system.

3. EXTERNAL STRUCTURE

We mean by external structure of vowel systems the description of the vowel space boundary in articulatory terms. It opposes the internal structure, with which we mean the positional organization of the vowels themselves. External structure is related to the notion of 'possible speech sound' (Lindblom, 1990). From a phonological point of view, the boundary of the vowel space is globally anchored between the combinations [low], [back, round] and [front, unround], representing /a/, /u/ and /i/, respectively. From a phonetic point of view, the set of possible speech sounds is a subset of the total sound-producing potential of the vocal tract. The relation articulation-to-acoustics and the inverse problem, the computation of the vocal tract shape from the acoustic output, plays here a central role.

The problem, how to relate vocal tract shape and acoustic output can be tackled in different ways: (1) in terms of electric LC-circuits. Historically, this has been the usual paradigm; (2) in terms of the n-tube representation of the tract (Fant, 1960; Atal *et al.*, 1978; Bonder, 1983; Ten Bosch *et al.*, 1987; Stevens, 1972, 1989). (3) in terms of articulatory-based tract models (by Lindblom, Sundberg, Ladefoged, Mermelstein, Maeda). (4) in terms of eigenfunctions of the Webster horn equation (Karal, 1953; Mrayati *et al.*, 1988).

Apart from their starting points, these four approaches are in fact mathematically equivalent.

Perrier *et al.* (1985), using Maeda's statistical analyses of articulatory positions has shown that the boundary of the vowel triangle can adequately be simulated by putting specific lower and upper bounds on the tube segment areas. Bonder (1983) and Ten Bosch (1991) studied this phenomenon by using the n-tube as articulatory model.

Since Atal *et al.* (1978), it is well known that the inverse problem has no unique solution (fibre). In order to specify one unique exemplar from the fibre, additional constraints have to be defined. This provides us the possibility to define an effort value to each formant position. The acoustic output being given, let ϕ denote the correponding fiber of all positions x in the articulatory space. Furthermore, we have some articulatory effort function e defined on the articulatory space. Then

$$\min\{e(x) \mid x \text{ on } \phi\}$$

denotes the minimal effort value on the fiber. This value depends on the fiber, i.e.

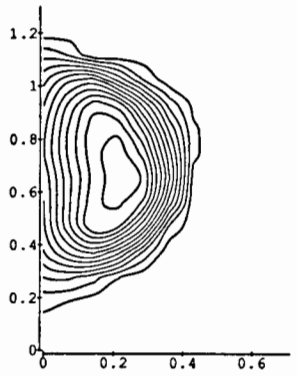

Fig. 1. Contour lines in the $(x, y) = (F_1, F_2)$ plane of an effort function defined on the articulatory space. Scaling: 1 ≡ 2000 Hz.

the acoustic output. Accordingly, the minimum effort value defines a 'effort landscape' on the acoustic space. It is shown in Ten Bosch (1991) that a relatively simple effort function e can be found such that the boundary of the vowel space, as found in languages, resembles closely one of the contour lines of that landscape (fig. 1).

Acknowledgements

This study has been supported by the Dutch Organization for the Advancement of Pure Scientific Research NWO (project 300-161-030).

4. REFERENCES

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. Am. 63. p. 1535 - 1555.

BADIN, P., Perrier, P., Boë, L.-J., and Abry, C. (1990). Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences. J. Acoust. Soc. Am. 87. p. 1290-1300.

BLADON, R.A.W. & Lindblom, B. (1982). Modeling the judgment of vowel quality differences. J. Acoust. Soc. Am. 69. pp. 1414-1422.

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

BONDER, L.J. (1986). A prediction method for modal n-vowel systems. Procs. Inst. Phon. Scs. Amsterdam, vol. 10. p. 73-90.

TEN BOSCH, L.F.M., Bonder, L.J., and Pols, L.C.W. (1987). Static and dynamic structure of vowel systems. Procs. 11th Intern. Congress Phon. Scs., vol. 1. p. 35-238.

TEN BOSCH, L.F.M. and Pols, L.C.W. 1989). On the necessity of quantal assumptions. Questions to the Quantal Theory. Journal of Phonetics 17. p. 63 - 70.

TEN BOSCH, L.F.M. (1991). On the structure of vowel systems. An extended dispersion model. PhD-thesis (in preparation). University of Amsterdam, The Netherlands.

DISNER, S.F. (1983). Vowel quality. The relation between universals and language-specific factors. UCLA WPP 58. Un. of California, LA.

FANT, G. (1960). Acoustic Theory of Speech Production. Mouton & Co., 's-Gravenhage.

KARAL, F.C. (1953). The analogous acoustical impedance for discontinuities and constrictions of circular cross section. J. Acoust. Soc. Am. 25. p. 327 - 334.

KEWLEY-PORT, D., and Atal, B. (1989). Perceptual differences between vowels located in a limited phonetic space. J. Acoust. Soc. Am. 85. p. 1726-1740.

LADEFOGED, P. and Lindau, M. (1988). Modeling articulatory-acoustic relations. UCLA Working Papers 70. p. 32 - 40.

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

LINDBLOM, B. (1986). Phonetic universals in vowel systems. In: Experimental Phonology (J. Ohala and J. Jaeger, eds.). Academic Press, Orlando, Florida. p. 13 - 44.

LINDBLOM, B. (1990). On the notion of "possible speech sound". Journal of Phonetics 18. p. 135 - 152.

MRAYATI, M., Carré, R., and Guérin, B. (1988). Distinctive regions and modes: A new theory of speech production. Speech Communication 7. p. 257 - 286.

PERRIER, P., Boë, L.J., Majid, R., and Guérin, B. (1985). Modélisation articulatoire du conduit vocal: exploration et exploitation. In: Proceedings of the 14emes Journées d'Etudes sur la Parole (Groupement des Acousticiens de Langue Française, Paris). p. 55 - 58.

SCHWARTZ, J.L., Boë, L.J., Perrier, P., Guérin, and Escudier, P. (1989). Perceptual contrast and stability in vowel systems: a 3-D simulation study. Proceedings of the Eurospeech Conference, Paris. p. 63-66.

STEVENS, K.N. (1972). The quantal nature of speech: evidence from articulatory-acoustic data. In: Human communication: A unified view (E. David and P. Denes, eds.). McGraw-Hill, New York. p. 51 - 66.

STEVENS, K.N. (1989). On the quantal nature of speech. Journal of Phonetics 17. p. 3 - 45.