TESTING THE DISPERSION-FOCALIZATION THEORY: PHASE SPACES FOR VOWEL SYSTEMS

J.L. Schwartz, L.J. Boë, N. Vallée
Institut de la Communication Parlée, Grenoble, France

ABSTRACT

The Dispersion-Focalization Theory (DFT) attempts to predict vowel systems thanks to a competition between two perceptual costs, namely global dispersion vs local focalization. The competition is controlled by two parameters, α and λ. We describe a new methodology for testing the DFT predictions: for a given number of vowels, phase spaces allow to determine the DFT winner in the (α, λ) space. We derive an (α, λ) region for which the theory predictions fit quite well with the phonological inventories.

1. INTRODUCTION

Substance-based theories of linguistic systems propose a deductive approach, which looks at the primary language-specific facts from an external point of view by considering non-linguistic constraints on possible speech sounds. Stevens [1] and Liljencrants and Lindblom [2] introduced the basic categories of arguments about the nature of the listener-speaker interaction and its role in shaping phonological systems, namely Lindblom's Dispersion Theory (DT) and Stevens' Quantal Theory (QT). The principle of the Dispersion-Focalization Theory (DFT) is to set a competition between a structural dispersion cost based on inter-vowel perceptual distances and a local focalization cost based on intra-vowel perceptual salience [3].

2. IMPLEMENTING THE DFT

2.1. Cost of a vowel system

The DFT assumes that for a given number of vowels, namely n, the preferred system (i.e. the most frequent in phonological bases) is obtained by minimizing a global cost summing two components, namely a structural Dispersion cost and a local Focalization cost, both applied on acoustic parameters characterizing each vowel. Vowels are described in our work by four formants (F1, F2, F3, F4), with F4 fixed at 3560 Hz, and all values expressed in bark, as computed by the formula proposed in [4]: bark = 7 ArgSh (Hz / 650)

The energy function of a given system with n vowels V^i, $i \in \{1, .., n\}$, is given by:

$$E_{DF} = E_D + E_F$$

with E_D a dispersion cost and E_F a focalization cost. E_D is defined, as in the DT, by :

$$E_D = \sum_{\substack{i = 1 \,..\, (n-1) \\ j = (i+1) \,..\, n}} (1/d_{ij})^2$$

with d_{ij} the perceptual distance between vowels V^i and V^j. To compute this distance, we use an Euclidian distance in the (F_1, F'_2) space, where the "second perceptual formant" F'_2 is computed from F_2, F_3 and F_4 on the basis of a model we have developed in the 80s [5]. In order to deal with the excessive number of high non-peripheral vowels in the DT predictions, we introduce Lindblom's proposal of a "stretching" of the acoustic space in the F_1 dimension [6] by using an (F_1, F'_2) weighted Euclidian distance, namely:

$$d_{ij} = [(F1^j - F1^i)^2 + (\lambda F'2^j - \lambda F'2^i)^2]^{1/2}$$

where λ can be chosen at any value lower than 1, assuming that higher formants play a lesser part in vowel phonetic quality than do lower ones.

The DFT discards from the DT by the introduction of a second energy term, called focalization cost, diminishing the energy of configurations with vowels with close F_1 and F_2, F_2 and F_3 or F_3 and F_4 ("focal vowels", [7]) and hence making such configurations more stable. This cost is defined by:

$$E_F = \alpha (E_{12} + E_{23} + E_{34})$$

with

$$E_{12} = -\sum_i 1/(F_2^i - F_1^i)^2$$

$$E_{23} = -\sum_i 1/(F_3^i - F_2^i)^2$$

$$E_{34} = -\sum_i 1/(F_4^i - F_3^i)^2$$

where α is a second free parameter. We thus obtain an energy function depending respectively on the parameter λ, which sets the weighting between F_1 and F'_2, and the parameter α, which determines the weighting of the additional focalization cost.

2.2. Selection criterion

Various criteria have been proposed in the literature for selecting vowel configurations. Whatever the criterion, a crucial methodological point concerns the way one deals with the well-known impossibility to analytically derive the solution of a non-linear minimization process (namely with non-quadratic energy landscapes, resulting in local minima). The original solution we have adopted here is based on what we call the "phase space" in reference to a classical procedure in Chemistry. In this method, we a-priori define a number N of "prototypes" covering all the vowel space, and for each number n we attempt to determine, according to the values of the free parameters λ and α, which is the system, made of n vowels selected among the N prototypes, which minimizes the total energy E_{DF}. Hence the problem becomes tractable: it consists in choosing one between a finite number (in theory, C_N^n) of systems, thanks to an associated variable E_{DF}. We use 33 prototypes with positions as "regular" as possible, in terms of distances in the (F_1, F'_2) space. The methodology may be summarized in the following way: For each value of n, determine the "phase space ", namely the regions in the (λ, α) space in which a given system of n vowels chosen among our 33 prototypes "wins", in the sense that it has the minimal E_{DF} cost in respect to all his concurrents. These "phase spaces" are then compared with phonological inventories, derived from the UPSID Database [8].

3. PHASES SPACES FOR SYSTEMS FROM 3 TO 7 VOWELS

3.1. Simulation results

The methodology described in Section 2 allowed us systematically determine the "phase spaces" for all values of n between 3 and 9. We shall concentrate the discussion on values of n from 3 to 7, which provide the most significant trends in the UPSID basis. The results are given in Fig. 1 to 5, which display, respectively for n = 3, 4, 5, 6 and 7, the best system for a given value of the pair (λ, α) in the square region [0, 1] x [0, 1]. These results provide the following trends.

Decreasing λ favours peripheral systems, which is exactly what we expected, since it results in vertically shrinking the vowel space. A more unexpected consequence is that a too small value of λ leads to either reduced (Fig. 1, 2) or asymmetrical (Fig. 3, 5) peripheral configurations, since the interactions between the front and the back side of the peripheral systems become important.

Increasing α favours focal vowels, namely first [i] and [y], then front unrounded vowels with the highest focalization benefit for the highest vowels, and finally back rounded vowels, which have all the same E_F cost. This mainly results in switching from a central high vowel (be it [ɨ] or [ɯ]) to an [y] . It may also produce a switch from a peripheral vowel to an [y].

Increasing n increases the dispersion cost of peripheral systems, hence it decreases the λ boundary necessary for making these systems optimal. Conversely, large n values favour systems with one or even two non-peripheral high vowels (for n=6 or 7).

3.2. Comparison with UPSID data

UPSID inventories provide the following trends. Symmetrical peripheral systems are the great winners for an odd number of vowels, namely [i, a, u] for 3-vowels systems, [i, 'e', a, 'o', u] for 5-vowels systems and [i, e, ɛ, a, ɔ, o, u] for 7-vowels systems. For 4-vowels systems the dominant acoustic structure is

[i,'e',a,u], while for 6-vowels systems, if we discard systems including schwa, which is according to us a "special" vowel [9], the situation is balanced between [i, e, ɛ, a, 'o', u] and [i, 'e', a, 'o', u, i].

If one considers all these constraints together, the correct (λ, α) region is roughly defined by:

$0.2 \leq \lambda \leq 0.3$ and $0 \leq \alpha \leq 0.4$

Additional constraints based on the stability for systems which contain an [y] as the single non-peripheral high vowel, namely an unbalanced [i, y, u] structure for high vowels, provide a floor value for α higher than 0 (see [3, 10, 11]), namely:

$0.2 \leq \lambda \leq 0.3$ and $0.3 \leq \alpha \leq 0.4$

4. CONCLUSION

This study shows that we are able to define a region for the two DFT parameters for which theoretical predictions are quite in line with experimental data coming from the UPSID database.

The DF Theory provides some kind of generalization of the D Theory. Indeed, the first simulations by Liljencrants and Lindblom (1972) should correspond more or less to the results displayed in Fig. 1 to 5 for a value λ equal to 1 (same weighting for F1 and F'2, or M1 and M'2 in their terms, with formants expressed in mels) and a value α equal to 0 (no focalization). However, our simulations clearly show that λ must be much lower than 1 in order to solve the problem of peripheral vowels and α higher than 0 in order to solve the problem of front rounded vowels. This second point is crucial. It confirms that the focalization term *is* necessary for understanding the [i, y, u] structure for high vowels, which is not negligeable in the UPSID base, since it represents 4.5 % of the whole base, and more than 25 % of the structures with three high vowels (namely two peripheral and one non-peripheral). Therefore the DFT provides a good basis for understanding vowel systems in detail.

REFERENCES

[1] Stevens, K.N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In (E.E.Davis Jr. & P.B.Denes, eds.) *Human Communication: A unified view*, 51-66. New-York: Mc Graw-Hill.

[2] Liljencrants, J., & Lindblom, B. (1972). Numerical simulations of vowel quality systems: The role of perceptual contrast. *Language*, 48, 839-862.

[3] Boë, L.J., Schwartz, J.L., & Vallée, N., (1994). The prediction of vowel systems: perceptual contrast and stability. In E. Keller (Ed.) *Fundamentals of Speech Synthesis and Speech Recognition* (pp. 185-213). John Wiley.

[4] Schroeder, M.R., Atal, B.S., & Hall, J.L. (1979). Objective measure of certain speech signal degradations based on masking properties of human auditory perception. In (B. Lindblom, S. Ohman, eds.) *Frontiers of Speech Communication Research*, 217-229. London: Academic Press.

[5] Mantakas, M., Schwartz, J.L., & Escudier, P. (1986). Modèle de prédiction du 'deuxième formant effectif' F'2 – application à l'étude de la labialité des voyelles avant du français. *15th JEP, Société Française d'Acoustique*, 157-161

[6] Lindblom, B. (1986). Phonetic universals in vowel systems. In (J.J. Ohala, ed.) *Experimental phonology*, 13-44. New-York: Academic Press.

[7] Badin, P., Perrier, P., Boë, L.J., & Abry, C. (1991). Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences. *Journal of the Acoustical Society of America*, 87, 1290-1300.

[8] Maddieson, I. (1984). *Patterns of sounds*. Cambridge studies in speech science and communication, Cambridge: Cambridge University Press.

[9] Schwartz, J.L., Boë, L.J., Vallée, N., & Abry, C. (1995). The Dispersion – Focalization Theory of vowel systems. Submitted to the *J. of Phonetics*.

[10] Vallée, N. (1994). *Systèmes vocaliques: de la typologie aux prédictions*. Thèse de Doctorat en Sciences du Langage, Université Stendhal, Grenoble.

[11] Jomaa, M., & Abry, C. (1992). *La base de systèmes vocaliques RHÔNSON*. Rapport PPSH 78B, ICP.

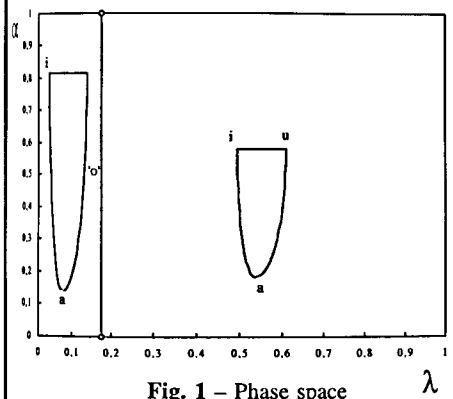

Fig. 1 – Phase space for 3-vowels systems

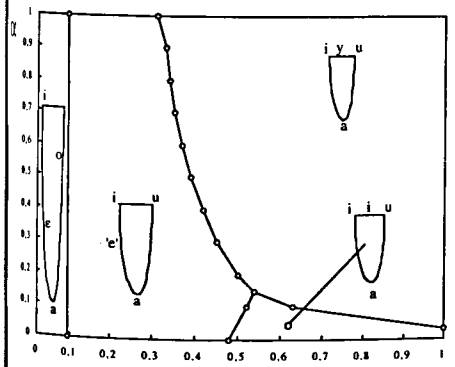

Fig. 2 – Phase space for 4-vowels systems

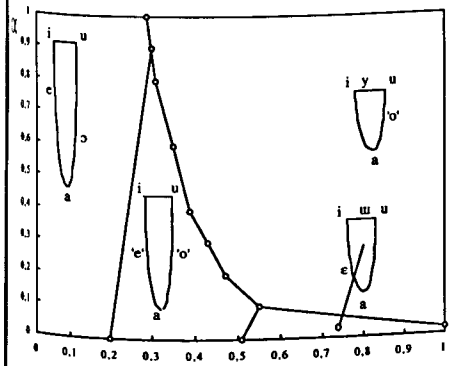

Fig. 3 – Phase space for 5-vowels systems

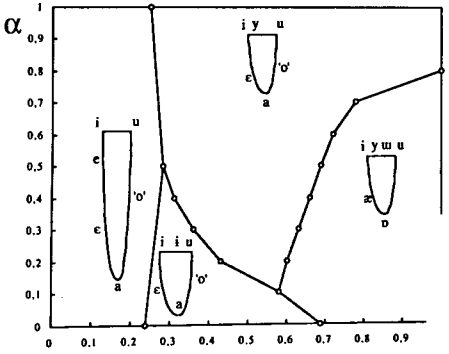

Fig. 4 – Phase space for 6-vowels systems

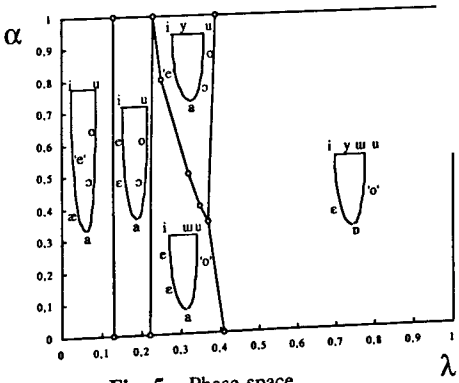

Fig. 5 – Phase space for 7-vowels systems