PHASE SPACES OF VOWEL SYSTEMS
A TYPOLOGY IN THE LIGHT OF THE DISPERSION-FOCALIZATION THEORY (DFT)

Nathalie Vallée, Jean-Luc Schwartz, Pierre Escudier
Institut de la Communication Parlée
INPG/Université Stendhal - BP 25X - 38040 Grenoble cedex - France
vallee, schwartz, escudier@icp.inpg.fr

ABSTRACT
The Dispersion-Focalization Theory (DFT) is a predictive model of vowel systems based on the minimization of two combined costs, one aiming at increasing auditory distances between vowel spectra ("dispersion"), the other at increasing the perceptual salience of each spectrum through formant proximities ("focalization"). This model is controlled by two parameters: \( \lambda \), the weight of \( F_1 \) with regards to the group of higher formants \( F_2, F_3, F_4 \) and \( \alpha \) the weight of focalization (encouraging "focal vowels") with regards to dispersion. For a given number of vowels we can predict in the \((\lambda, \alpha)\) space what should be the preferred system, together with other sub-optimal ones. This is achieved thanks to the so-called "phase space", a well-known procedure in thermodynamics used to predict the state of a given substance as a function of pressure and temperature. The present work consists in determining how many phonological vowel systems within the 451 UPSID languages agree with different possible structures predicted by the DFT. We present a complete typology of systems with 3 to 7 vowel qualities in the light of the DFT, we introduce a new variant inspired from physics - that is, polymorphism within a given phase, leading to define superstructures in vowel systems - and we show that the phase space methodology allows to predict a great part of the observed systems. This enables to better specify the \( \lambda \) value: a value around 0.2 seems more or less compatible with about 85% of the 3-to-7 vowels systems considered in the UPSID basis.

1. THE DISPERSION-FOCALIZATION THEORY

1.1. General principles
Since the beginning of the 70s several proposals have been done to predict the phonological structure of vowel systems with non-phonological principles, be they listener-oriented (perceptual contrast and stability) or speaker-oriented (articulatory contrast and economy). The so-called "sufficient perceptual contrast" theory [1] provides the best global fit with phonological data. However, to overcome its two main problems (that is, the excessive number of high non-peripheral vowels in the model predictions and the impossibility to predict the \([i\ y\ u]\) series within the high vowel set), we proposed at ICP the Dispersion-Focalization Theory (DFT) [2]. The DFT allows to predict vowel systems thanks to a competition between two perceptual costs: for a given number of vowels, the most frequent system in the world languages is supposed to be obtained by minimizing a global criterion combining a structural dispersion cost and a local focalization cost.

2. Implementation
Each vowel is characterized by the formants of its spectrum, that is \( F_1, F_2, F_3 \) and \( F_4 \) expressed in a perceptual Bark scale. The \((F_2, F_3, F_4)\) set allows to compute an integrated "effective perceptual formant" \( F'_2 \). The dispersion cost involves auditory distances between vowel spectra, computed through an Euclidean distance in the \((F_1, F'_2)\) space, and favours large inter-vowel distances. The focalization cost involves distances between consecutive formants within each vowel spectrum, and favours focal vowels, that is vowels with closed \( F_1 \) and \( F_2, F_3 \) and \( F_4 \) or \( F_3 \) and \( F_4 \). The model is controlled by two parameters, that is \( \lambda \) specifying the weight of \( F_3 \) in respect to \( F_1 \) in the dispersion cost, and \( \alpha \) specifying the respective weight of the focalization cost relative to the dispersion cost.

1.3. Phase spaces
For a given number of vowels, from 3 to 9 (beyond this limit, vocal systems introduce a new dimension, mainly nasality and less often quantity, [3]), we can predict, in the DFT frame, the different vowel systems in the \((\lambda, \alpha)\) space. This leads to the determination of what we call "the phase space", a well-known procedure in thermodynamics used to predict the states of a substance (for example the states of water: steam, liquid and ice), as a function of pressure and temperature. The general trend is that, for a given number of vowels in a system, decreasing \( \lambda \) favours peripheral systems while increasing \( \lambda \) favours systems with one and then two high non-peripheral vowels; and increasing \( \alpha \) favours focal vowels, and particularly stabilizes \([y]\) within an \([i\ y\ u]\) high series, while this series is unstable when \( \alpha \) is set to 0.

Previous work allowed us to verify that these predictions were more or less compatible with the observed preferred phonological vowel systems in the "UPSID 317" database [4]. Considering that peripheral systems are generally preferred from 3 to 7 vowels and that the \([i\ y\ u]\) series of high vowels exists in a significant amount of cases in the basis (about 5% of the cases in the whole basis, and 13% of the cases for systems with 7 vowels or more), we showed that setting the \( \lambda \) value around 0.2-0.3 and the \( \alpha \) value around 0.3-0.4 led to quite acceptable predictions [2]. In the present work, we try to go one step further: we shall attempt to determine where in the phase spaces one can find the different systems, preferred or not, existing in UPSID 451, and what kinds of "superstructures" can be derived from this analysis.
2. STRUCTURAL SYMMETRIES BETWEEN VOWEL SYSTEMS: A TYPOLOGICAL EQUIVALENCE CRITERION

2.1. Prototypical structures in phase spaces

Our previous simulations lead to determine "prototypical systems". These are winning n-vowels systems in the DFT framework, in the sense that they have the minimal global Dispersion-Focalization (DF) energy, according to the values of the two free parameters \( \lambda \) and \( \alpha \).

We have focused our study on values of \( n \) from 3 to 7 because they allow to capture the most significant phonological tendencies of the UPSID basis. The DFT simulation results are given in Figures 1-5, respectively for \( n = 3, 4, 5, 6 \) and 7. For each value of \( n \), the phase space determines regions in the \((\lambda, \alpha)\) space in which a given system wins (with its vowel qualities displayed as black points on a prototypical grid). We see that there are two prototypical systems for \( n = 3 \), which we called \( S_{31}T_{1} \) and \( S_{32}T_{2} \). There are four prototypical systems for \( n = 4, 5, 6 \), and five prototypical systems for \( n = 7 \): let us call them \( S_{71}T_{1} \) with \( n \) from 3 to 7, and i from 1 to 5. The global trend is that increasing \( n \) increases the dispersion cost of peripheral systems, hence it decreases the \( \lambda \) boundary necessary for making these systems optimal. Hence peripheral systems are favoured with small values of \( \lambda \). When \( \lambda \) is too small, the vowel space is completely vertically stretched (since higher formants play a minimal part in the determination of vowel phonetic quality); this favours asymmetrical peripheral configurations because of the interactions between front and back peripheral vowels in the systems. Non-peripheral configurations, that is systems with more than two high vowels, appear with large \( \lambda \) values, and when \( \alpha \) increases, focal vowels (especially \([i]\) with close \( F_{1} \) and \( F_{2} \), other front unrounded vowels together with \([y]\), all with close \( F_{1} \) and \( F_{2} \), and back rounded vowels, with close \( F_{1} \) and \( F_{3} \)) are favoured. Decreasing \( \alpha \) leads to replace the high rounded vowel \([y]\) by a high vowel acoustically more central (namely \([i]\) or \([u]\)).

2.2. Reverse prototypical structures

We set up the hypothesis that two structures having the same number of peripheral vowels but systematically replacing front unrounded vowels by back rounded ones with the same height, and vice-versa, are equivalent structures in the sense of DFT, that is to say that they have roughly the same DF energy for a given value of \( n \) and of the \((\lambda, \alpha)\) pair. This was systematically verified, by comparing the energy of the \( S_{71}T_{1} \) prototypical systems, with reverse systems that we called \( S_{71}T_{1}^{*} \): for example, for \( n = 4 \) we compared \( S_{41}T_{1} = [i u u a] \) with \( S_{41}T_{1}^{*} = [i e a] \), and \( S_{42}T_{2} = [i o u a] \) with \( S_{42}T_{2}^{*} = [i o a u] \). We have checked that \( n \)-vowels systems have a DF energy quite close to the \( S_{71}T_{1} \) ones whatever the region of the phase space, that is to say whatever the \( \lambda \) and \( \alpha \) values. Pushing the analogy with physics one step further, this reminds us of the "polymerization" of a number of solids (e.g. metals, or crystals). In this situation, while fusion produces an homogeneous liquid phase, solidification leads to mixtures of two or more variants of the solid phase, all stable and more or less with the same energy. This is exactly what happens here with the two variants within a given phase. Hence in the following we consider in the typologies of the phase spaces "superstructures" grouping prototypical structures and reverse ones (displayed with white points instead of black ones in Figures 1-5). The relevance of these superstructures for describing the UPSID basis will now be discussed in the next section.

3. COMPARING UPSID DATA WITH DFT SIMULATIONS

3.1. UPSID data reanalyzed

The UPSID basis gathers phonological systems of 451 languages in the world, sampling more or less uniformly all linguistic families [5]. The languages in UPSID have 3 to 28 vowels. In order to test our hypothesis, we have reanalyzed the UPSID basis of vowel systems, thanks to a two-steps methodology.

First, from raw data, that is to say without any typological equivalencies, we obtain 252 types of phonological structures from 3 to 17 vowel qualities. What we call vowel qualities corresponds to "basic segments" (vs. "elaborated" and "complex" segments) in the sense of [6]. We note that more than 96% of the languages have from 3 to 10 basic vowel qualities, and if we focus our study on systems with 3 to 7 qualities, we obtain 77% of the 451 languages (348 systems). This is due to the fact that there are in many cases more vowels than vowel qualities in a given system; for instance \([i e a o u]/\) is the phonological structure of four UPSID languages of which three have more than 5 vowels: Chipewyan with 14 vowels /i e a o u i: a: u/ \(\rightarrow i 'a' a u\)

3.2. Distribution of UPSID data within phase spaces

We now have at our disposal both a series of predictions, organized around the typology \( S_{71}T_{1}/S_{71}T_{1}^{*} \) defined in Section 2, and an inventory of 342 systems (the three-fourth of UPSID 451) with 3 to 7 vowel quantities, defined in Section 5.1. The final goal of this work was to try to associate to most of these 342 systems a region in the phase space where they would be optimal (i.e. viable in the sense of the DFT). This is what is displayed in Figures 1-5, where we have plotted within each region of the phase spaces the number of systems fitting with the corresponding structure. Let us now discuss the obtained results in more detail.
First, it appears that 303 of the 342 3-to-7 vowel systems (88.6%, or 67.2% of the whole UPSID 451 basis) fit with one of the $S_5 T_1$ or $S_5 T_1^*$ types. The 39 rejected systems (fitting with no prototypical or reverse type) correspond to one system in $S_5 5$ in $S_9 9$ in $S_{10} 10$ in $S_{14} 14$ in $S_7$; hence their number increases with $n$, which is logical since the complexity of the distribution of vowel qualities increases.

Second, the most widespread types in Fig.1-5 are those corresponding to $S_5 T_1$ (and sometimes to $S_5 T_1^*$). This provides a first confirmation on UPSID 451 of our results on UPSID 317 [1], that is that the $(\lambda, \alpha)$ region defined by $0.2 < \lambda < 0.3$ and $0 < \alpha < 0.4$ is compatible with preferred 3-to-7 vowel systems in UPSID.

But the data in Fig. 1-5 provide some new confirmation of this result. Indeed, it appears that systems corresponding to types associated to large $\lambda$ or $\alpha$ values are quite few. On the contrary, most systems on Fig. 1-5 are located at low $\lambda$ values. Indeed, apart from the "best" $S_5 T_1$ structures, other structures generally occupy quite close regions (mostly of types $S_5 T_1$ or $S_5 T_1^*$) and if we define a broad acceptable region such as $0.1 < \lambda < 0.3$ and $0 < \alpha < 0.4$, we obtain a total of 293 systems, that is 85.6% of the 342 systems of our inventory, which is quite important. Altogether, this confirms with a strong reliability the need of "stretching" the acoustic space along the $F_1$ dimension in auditory spectral distances, which indicates the dominant role played by the lower formant $F_1$ in the vowel phonetic quality.

The next observation deals with the symmetry between front and back peripheral vowels. Globally, the data in Fig. 1-5 confirm the well-known fact that vowel systems "prefer" both peripheral vowels and front-back symmetry. In the disymmetrical cases, when the numbers of front and back vowels are different, the (classical) trend is that there are more front than back ones: for example, $S_9 T_1 [i o e a]$ vs. $S_5 T_1^*$ [ueea] structures, $S_8 T_1 [i u e' a]$ vs. $S_6 T_1^*$ [iuoa] structures, $S_5 T_1 [i u o' a]$ vs. $S_5 T_1^*$ [iuoa'] structures. When the number of front and back vowels are the same, the (less classical) trend is that front vowels have often a more open degree than back ones: though this is not true for 3-vowels systems ($S_5 T_1 [i o a]$ vs. $S_7 T_1^* [o e a]$ structures), it is clearly the case for 5-vowels systems ($S_4 T_1 [i u o e a]$ vs. $S_5 T_1^*$ [iueoa] structures) and for 7-vowels systems ($S_5 T_1 [i u o e a]$ vs. $S_7 T_1^* [i u o e a]$ structures).

In what concerns focalization, its role is more important for stabilizing the vowel structures of larger systems, particularly those containing [y]; this vowel appears in only 2% of the UPSID languages from 3 to 7 vowel qualities, but almost 7% of all the UPSID languages; notice that more than half of them are Indo-European and Uralo-Altaic languages [8]. In our simulations, [y] is only present in the $S_5 T_1 [i u o e a]$ structure (3 examplars, that is only 1% of our restricted basis).

4. CONCLUSION

This study allowed us to go one step further in our program of experimental validation of the Dispersion-Focalization Theory. To begin with, we obtained a series of results on UPSID 451 which globally confirm previous simulations on UPSID 317, which is of course not unexpected, but still satisfying. But more importantly, we have proposed here an additional ingredient in our modelling phase-space framework, with the notion of polymorphism of a given phase, or superstructures in phonological organisations. Thanks to this new ingredient, we were able to obtain two new and important results. First, almost 90% of the 342 3-to-7 vowel systems kept in our restricted inventory could be located on the phase spaces in Fig. 1-5. Second, more than 85% of these systems correspond to $\lambda$ values around 0.2, which suggests that this value is indeed a crucial feature in the determination of vowel quality, and which provides overall a strong argument in favour of the DFT

A next step should consist in adding 8- and 9-vowels systems to the present analysis. Indeed, our analysis of raw UPSID 451 data demonstrate that more than 9 languages over 10 (91.5%) in the basis contain 3 to 9 vowel qualities. The difficulty however is that phase spaces become increasingly complex with such a large number of vowels.

But whatever the feasibility of this next step, we believe that the present study provides a new illustration of the convincing ability of substance-based theories, from Stevens’ quantal theory and Lindblom’s Dispersion Theory to ICP Dispersion-Focalization Theory, to produce realistic predictions and useful typologies, at least for vowel systems.

ACKNOWLEDGEMENTS

We would like to thank Nourredine Abbadeni, Élisabeth Pinto et Florence Mistrulli, for their participation in the implementation of UPSID 451, Cloé Bourdét, Catherine Clouzot, for having helped to clear up the job, Valérie Alcantara for the corrections made after the implementation of UPSID at the ICP. Finally, a special attention to Ian Maddieson who provided us with UPSID 451 data base and Louis-Jean Boë for the stack programmation. This research was partially financed by the Région Rhône-Alpes (ARASSH) and the GIS des Sciences de la Cognition.

REFERENCES

Figures 1-5. Phase spaces for 3-to-7 vowel structures.

$S_n T_i$ structures have their vowel qualities displayed as black dots on a prototypical grid. $S_n T_i^*$ vowels are displayed as white dots replacing the black ones. The number in the oval is the total number of UPSID languages with the $S_n T_i$ and $S_n T_i^*$. 