VOWEL PROTOTYPES FOR UPSID'S 33 PHONEMES

N. Vallée, L.J. Boë and Y. Payan
ICP URA CNRS n° 368 INPG/ENSERG – Université Stendhal
BP 25 38 040 Grenoble Cedex 9 France

ABSTRACT

UPSID is a phonological database made up of 33 primary vowel symbols. In this paper we propose 33 vowel prototypes at articulatory and acoustic levels. We can generate them with an anthropomorphic articulatory model. A wide-ranging bibliographic study has enabled us to (i) establish a classification of values for articulatory input parameters, geometrical values of the midsagittal section and crucial values of the area functions; and (ii) specify F1-F2-F3 formant values. These prototypes have been used for a prediction model of the sound systems of world languages.

1. INTRODUCTION

The research of vowel prototypes has been done in the frame of a substance-oriented phonology. Tradionally, the functional efficiency hypothesis, which is looked for at the substance level, is widely used to explain the contents of phonological systems. In this hypothesis, distinctiveness plays a fundamental part: each component is defined in relation to the other components of the system. In advocating a *"substance-based"* analysis, Liljencrants & Lindblom [1] have proposed to reverse the trend by stressing the importance of the "lower levels" in the emergence of systems. The selection of sound units would rest on articulatory and perceptive physiological constraints, which would allow us to explain and predict system structures. The theory of maximal contrast and therefore the design of predictive models both rely on the principle of sound discrimination. To improve modelization, Schwartz & al. [2] have proposed the DFT model *(Dispersion- Focalization Theory)*: They have added auditory pregnancy criteria to distinctiveness [3] (see Schwartz, Boë, Abry & Vallée in these Proceedings). To simulate typical vowel configurations in an acoustic space with a model, we need prototypes, specified both by articulatory and acoustic characteristics [4]. Consequently, we will switch from *form* to *substance* by establishing a relationship between some linguistic units of a representative sample of the phonological inventory of world languages (UPSID [5]), and their physical shape (articulatory and acoustic parameters). In a typological study [4], we have listed 33 vowel qualities which permit to describe the set of symbolic constituant elements of UPSID's 317 systems. From a normalization of the acoustic vowel space [6], we have surveyed and estimated the corresponding values for articulatory input parameters and vocal tract geometry (e.g. location and dimension of the constriction, upper point of the tongue body, furthest back point of the tongue root in the pharynx). For this task, we have used macro-variations [7], – interface functions between articulatory input and acoustic output –, which allow to provide for (i) the acoustic consequences of gestures, (relationship between formants and articulators); (ii) the influence of crucial geometrical parameters [8] (location and dimension of the constriction, and lips area) on the acoustic output. The task has been executed with SMIP, software developed at the ICP within the framework of a European project (ESPRIT/BR N°6975) whose central core is made up of Maeda's articulatory model [9].

2. PROBLEMS TO SOLVE

The traditional description [10], which provides a position for each vowel in terms of height and advancing tongue arching in the buccal cavity, is inadequate. The highest point of the tongue is an operational descriptive parameter whereas the location of the constriction can be directly linked to the acoustic output. Recently, Boë & al. [11] have attempted to unify the traditional description *"lips, tongue arching"* and the acoustic oriented description *"troat-tongue-lips"*.

2.1. INVERSION:

It consists in deriving the vocal tract shape from the acoustic output. Several static articulatory configurations of the vocal tract constitute what is called "a fiber" of the articulatory space, i.e. they make up a set of configurations which supply the same acoustic output [12] [13]. We then need to select a single configuration of the vocal tract and get rid of the rest of the fiber by imposing articulatory and acoustic constraints on these prototypes with the help of experimental and theoretical publications available.

2.2. ARTICULATORY-ACOUTIC RELATIONSHIP:

Secondly, we have to deal with the non-linear and discontinuous relationship between articulation and acoustics. Thanks to the study of macro-variations we are able to foresee the relationship between formants and articulators (Boë, Badin & Perrier, in these Proceedings).

2.3. VARIABILITY:

Different strategies of tongue and jaw allow to produce acoustically identical vowels. Experiments such as "bite-block" [14] show that the vocal tract is capable of using articulatory compensations to produce the same vowel under different conditions.

However, research on invariance [8] [13] [15] has shown an important regularity of the location of the constriction in vowel articulation, whatever the language.

These 3 fundamental issues have led us to collect results of acoustic surveys as well as data on articulatory descriptions [4].

3. METHOD

3.1. VOWEL SPACE:

There are two fundamental constraints: (i) any prototype must be included into maximal vowel space [6]; (ii) the configuration proposed in that space must not fall into the "gap" observed in natural language systems around 300 Hz for F1 and 1,000 Hz for F2, which corresponds to formant area linked to the nasal-pharyngeal tract [16][17].

3.2. PROTOTYPES FOR FRENCH:

All prototypes have been elaborated by calculating dispersion ellipsoids, at the acoustic level, of the 10 oral vowels of French [i e ɛ a ɔ o u y ø œ] for which numerous data were available [4]. Thanks to 60,000 sagittal views of the vocal tract, generated by Maeda's articulatory model [9], we have looked for the ones which fitted our dispersion ellipsoids. With sagittal views, the model supplies us with the values of 7 control parameters determining the position of articulators: lips (retracted and protruded), tongue (dorsum, body and tip), jaw and larynx.

Thanks to a sagittal section, we can calculate with SMIP: (i) the area functions whose crucial zones are: Xc the position of the narrowing, its aperture Ac, and Al the lips area; (ii) the transfer function of the vocal tract and formants.

In the same way as Majid & al. [7], and thanks to SMIP, we can infer articulatory data from acoustic targets.

3.2. OTHER PROTOTYPES:

Acoustic targets of the remaining 23 vowels have been positioned in vowel space thanks to surveys of formant data from work done on modelling (synthetic vowels), and various acoustic studies on over 30 languages [4]. A database has thus been constituted. It contains vowel systems of various sizes.

For choosing the value of articulatory parameters, we have also worked with macro-variations of French oral vowels and a wide bibliographic survey. Comparing the various data has enabled us to find a coherence between articulatory control parameters, crucial values of the area function and position in the space formant, even though the variability of sources has sometimes forced us to make compromises in adjusting parameters (Figures 1 & 2).

The 33 acoustic prototypes retained can be synthetized, allowing an auditory control.

4. PUTTING PROTOTYPES TO GOOD USE

More than a stage between form and substance to evaluate predictions, prototypes are the raw material for a whole field of research:

- First, vowel prototypes remind us of the first definitions of "standard vowel quality" of phoneticians [10] or the Jones's cardinal vowels [18], whose primary objectives were to be a reference for the IPA user.

- The hierarchical classification of articulators for all prototypes allow to address again the issue of traditional articulatory description of vowel and its relationship with acoustic production [11].

- Prototypes are used as a preliminary phase in any attempt at predicting vowel systems. It is now common knowledge that psycho-acoustic parameters are not sufficient for all types of prediction and we must look into articulatory production process for criteria that could improve simulations. Results of this type of research look promising in order to associate articulatory dimension to acoustic and perceptive criteria of distinctiveness – e.g. a description of the articulatory distance (Berrah & al., in these proceedings).

5. REFERENCES

[1] LILJENCRANTS, J. & LINDBLOM, B. (1972). "Numerical Simulation of Vowel Quality Systems: the Role of Perceptual Contrast". *Language 48*, 839-862.

[2] SCHWARTZ, J.L., BOË, L.J., PERRIER, P., GUÉRIN, B. & ESCUDIER, P. (1989). "Perceptual Contrast and Stability in Vowel Systems: A 3-D Simulation Study". *Eurospeech 89*, Paris, Vol. 1/2, 63-66.

[3] BOË, L.J., SCHWARTZ, J.L. & VALLÉE, N. (1994). "The prediction of Vowel Systems: Perceptual Contrast and Stability". *Fundamentals of Speech Synthesis and Speech Regognition*, Ed. by Keller E., Wiley & Sons Ltd, London, England.

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

[5] MADDIESON, I. (1986). *"Patterns of Sounds"*. 2nd edition, Cambridge University Press, Cambridge (1st edition: 1984).

[6] BOË, L.J., PERRIER, P., GUÉRIN, B. & SCHWARTZ, J.L. (1989). "Maximal Vowel Space". *Eurospeech 89*, Paris, Vol. 2/2, 281-284.

[7] MAJID, R., BOË, L.J. & PERRIER, P. (1986). "Fonctions de sensibilité, modèle articulatoire et voyelles du français". *15e Journées d'Étude du GALF-G.C.P.*, Aix-en-Provence, 59-63.

[8] BOË, L.J., PERRIER, P. & BAILLY, G. (1992). "The Geometric Vocal Tract Variables Controlled for Vowel Production: Proposals for Constraining Acoustic - to - Articulatory Inversion". *Journal of Phonetics 20*, 27-38.

[9] MAEDA, S. (1989). "Compensatory Articulation During Speech: Evidence from the Analysis and Synthesis of Vocal-Tract Shapes using an Articulatory Model". *Speech Production and Speech Modelling*, Ed. by Hardcastle W.J. & Marchal A., Academic Publishers, Kluwer, Netherlands, 131-149.

[10] BELL, A. (1867). *"Visible Speech"*. Ed. by Simpkin & Marshall, London.

[11] BOË, L.J., GABIOUD, B., PERRIER, P., SCHWARTZ, J.L. & VALLÉE, N. (1994). "Vers une unification des espaces vocaliques". *Levels in Speech Communication: Relations and Interactions*, Ed. by Beekmans R., Jospa P., Schoegen J., & Serniclaes W., Elsevier Science Publishers B.V., Amsterdam, Hollande.

[12] BOË, L.J., & PERRIER, P. (1988). "C.F. Hellwag 200 ans après ou les éléments d'une fibre conductrice". *17e Journées d'Étude sur la Parole*, S.F.A., G.C.P., 200-205.

[13] STEVENS, K.N. & HOUSE, A.S. (1955). "Development of a Quantitative Description of Vowel Articulation". *J. Acous. Soc. Am. 27*, Vol. 5, 484-493.

[14] GAY, T., LINDBLOM, B. & LUBKER, J. (1981). "Production of Bite-Block Vowels Acoustic Equivalence by Selective Compensation". *J. Acous. Soc. Am. 69*, 802-810.

[15] WOOD, S.A.J. (1982). "X-Ray and Model Studies of Vowel Articulation". *Working Papers 23*, Lund University, Department of linguistic, Lund.

[16] MAEDA, S. (1984). "Une paire de pics comme corrélat acoustique de la nasalisation des voyelles". *13e Journées d'Étude du GALF-G.C.P.*, Bruxelles, 223-224.

[17] FENG, G. (1986). *"Modélisation acoustique et traitement du signal de parole : le cas des voyelles nasales"*. Thèse de Docteur Ingénieur, INP Grenoble.

[18] JONES, D. (1918). *"An Outline of English Phonetics"*. 1st edition (9th edition: 1960), Heffer W. & Sons L.T.D., Cambridge.

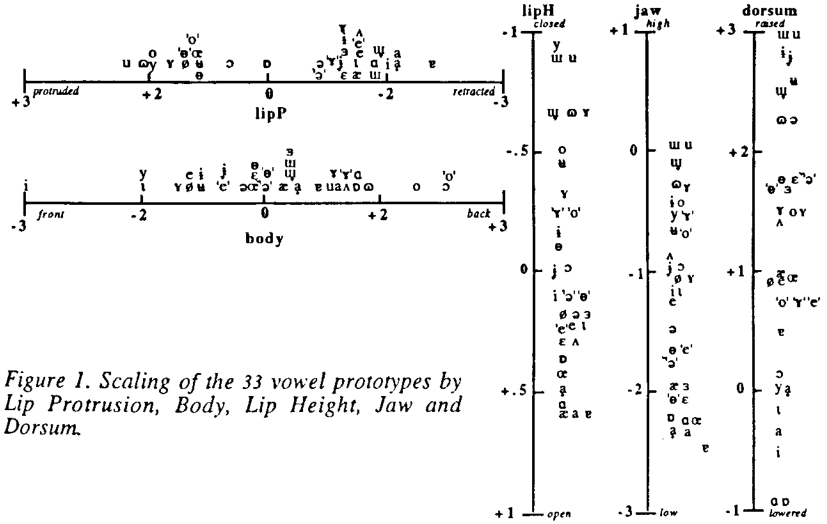

Figure 1. Scaling of the 33 vowel prototypes by Lip Protrusion, Body, Lip Height, Jaw and Dorsum.

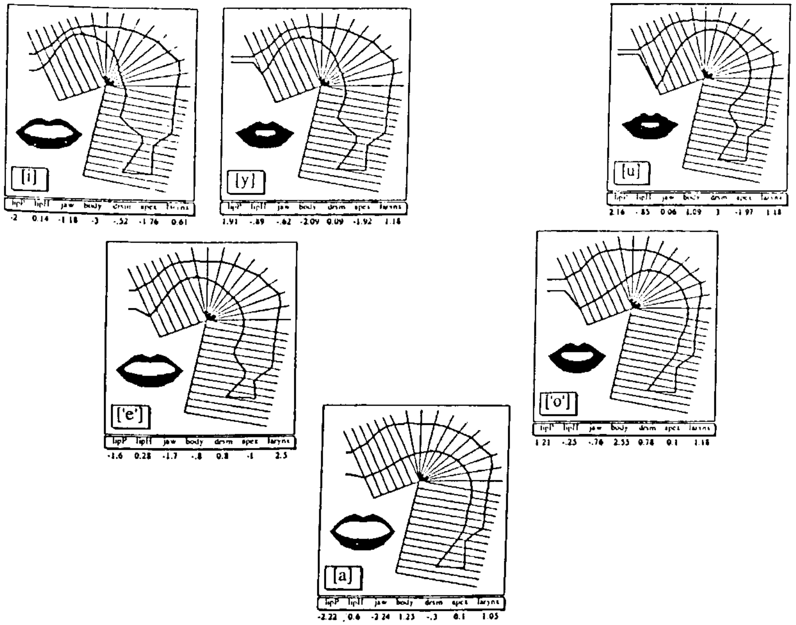

Figure 2. Sagittal views of vocal tract for the five most frequent vowels of world languages /i 'e' a 'o' u/. And the French vowel /y/.