EMERGENT SYLLABLE USING ARTICULATORY AND ACOUSTIC PRINCIPLES

A.R. Berrah, L.J. Boë and J.L. Schwartz
Institut de la Communication Parlée, Grenoble, France

ABSTRACT

Speech sequences can be regarded as a succession of syllable gestures composed of consonants and vowels. The choice of these basic units is not arbitrary: languages have a tendency to optimize their sound structures. This paper deals with the prediction of CV syllables. In order to bring about these gestures, we have suggested syllabic prototypes defined acoustically and articulatorily. Then, we have defined new optimization criteria.

1. INTRODUCTION

"Where do phonological universals such as segments and features come from?" In 1984, Lindblom et al. [1] addressed this ambitious question by trying to simulate the emergence of a self-organized model of phonological structure. They have restricted their attention to articulation involving transitions from a closed (stoplike) to an open (vowellike) state, in other words, CV syllables: the most frequent in the languages of the world [2].

With a 7D articulatory space (lip height and protrusion; jaw; body, dorsum and apex of the tongue; height of the larynx), and a 4D acoustic-perceptive space (F1-F4 in barks) our research has consisted in predicting the rank of "efficiency" of CV syllables among all the 20 possible combinations stoplike [b d g ɢ] (with [ɢ] an hypothetic but not observed pharyngeal stop) and with vowellike [i 'e' a 'o' u]. The CV prototypes have been designed with Maeda's articulatory model [3]. For designing CV prototypic transitions, we have taken into account X-Ray data for coarticulation [4], and Sussman's locus values for formant transitions [5]. Each syllable is characterized by a global efficiency: ratio of acoustic efficiency and articulatory cost. We define a criterion of maximization of intersyllabic distances by using F'_2 (in Barks) evaluation in the perception of transitions. For a given system of syllables, we propose a global criterion taking into account intersyllabic distances and efficiency of each syllable. The emergence of a syllable in a set of syllables is then simulated and its rank is discussed.

As a by-product a 3D space is proposed for the 20 syllables; it has been derived from a Kruskal analysis [6] on intersyllabic acoustic distances of our prototypes.

2. PROBLEM SOLVING

A great number of questions still remain to try to explain how linguistic sounds are made up. Topological studies have shown the occurrence in the systems of a group of phonetic properties which are found in most languages of the world.

The first questions asked in the face of the universal tendencies in the systems are notably:

- Which factors entail these restrictions?
- What are the causes which provoke the tendency to use only one little set of sound qualities for communication?

Assumptions are made by examining universals from a phonetic point of view and point towards the notion of ***functional efficiency***. The systemic nature of sounds is shown: a sound is examined as the constituent of a system.

Two very important fields of research will stem from these hypotheses in the seventies: The *Quantal Theory* of Stevens [7] and the *Dispersion Theory* of Liljencrants & Lindblom [8].

The initial hypothesis is based on two principal ideas: articulatory simplicity and perceptive distinction.

3. SYLLABIC PROTOTYPES

We have restricted our research to the twenty syllables obtained by combining the following 4 consonants: [b d g ɢ] and the 5 vowels [i 'e' a 'o' u]. The choice of these vowels is not arbitrary: Typology reveals that most languages of the world with 5 vowels have a vowel system with [i 'e' a 'o' u]. We want to obtain an articulatory and an acoustical

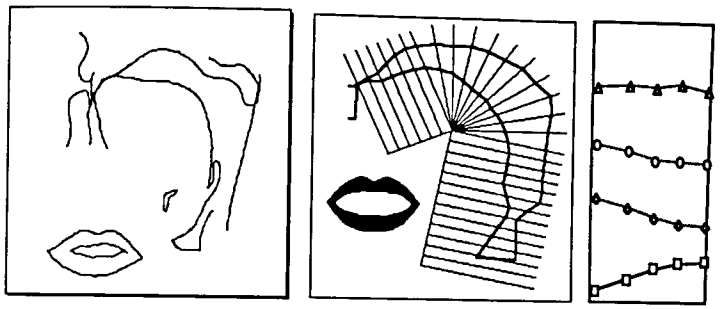

Figure 1. From left to right: X-Ray photography, vocal tract shape and the formant transitions from [d] to [a].

target for every syllable to determine which configurations are typical for syllabic systems.

In fact, to select syllabic standards in the acoustic and articulatory space, we must characterize a typical form of the vocal tract for a certain number of acoustic parameters (the formants).

Implementation

The articulatory and the acoustic targets of every syllable are obtained from the prototype of the vowel and the coarticulated occlusion of the consonant.

(i) We have used the Vallée et al.'s vowel prototypes [9].

(ii) Thanks to Maeda's model, we have generated a syllabic prototype from the vowel prototype by adjusting the parameters of some articulators [10] until occlusion is reached.

Prototypic syllables

Thanks to the X-Ray data of Bothorel et al., we have proposed [10] a typical form of the vocal tract for the fifteen most frequent syllables CV [b d g] x [i 'e a 'o' u] and five hypothetical CV syllables [ɢ] x [i 'e' a 'o' u] [11].

Locus notion

To validate the syllabic prototypes, we have computed the locus of consonants [b], [d], [g] and [ɢ], [g] presents two loci associated to front and back vowels. Figure 2 presents the locus for our prototypic syllables in a 3D space in agreement with Sussman's data.

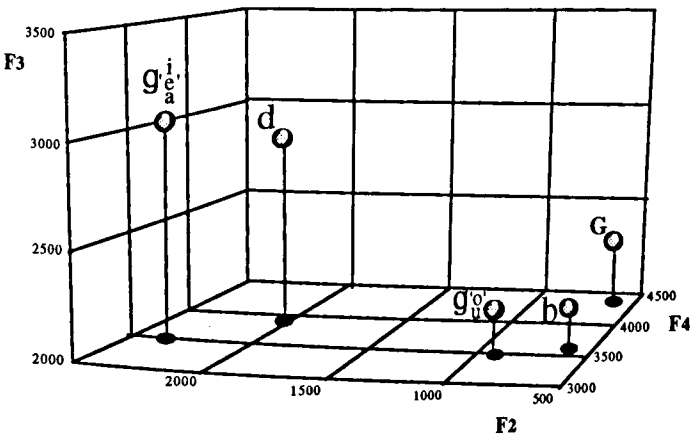

Figure 2. Consonantal locus of our CV syllables in a 3D space F2-F4.

4. THE PREDICTION MODEL

Our syllable model consists in predicting the most frequent syllables used in languages. We hypothesize that they are the most efficient for communication. The general postulate of this study is that phonetic signals evolve in such a way that their production and perception is more easy for training, production and perception.

Acoustic efficiency

The acoustic space considered is 4D which represents the first 4 resonances of the vocal tract. Our predictive model is inspired of Liljencrants' & Lindblom's model [8] and the DFT model [12]. This model is based on a bidimensional vowel space defined by the first (F_1) and the second effective formant (F'_2). In order to take into account physiological and perceptual constraints [13], The contribution of F_1 is increased.

Intrasyllabic efficiency

Intrasyllabic efficiency (or perceptual salience) is a characteristic of individual transitions. It is defined in terms of the extent of syllable trajectory, i.e., the distance between the initial and the final auditory spectra. A syllable is more salient if the resonance frequencies of the consonant and the vowel are different. The acoustic efficiency, Acoust-Eff_{CV}, is given by:

$$\sqrt{(F_{1_c} - F_{1_v})^2 + \lambda^2 * (F'_{2_c} - F'_{2_v})^2}$$

λ^2: ponderation between F_1 et F'_2.

Intersyllabic efficiency

Intersyllabic efficiency (or perceptual distance) of two arbitrary CV transitions is a dimension used to rank all possible pairs of CV events in order to minimize the risk of confusion between the items of the lexicon.

The intersyllabic distance ds_1s_2 (S_1=C_1V_1, S_2=C_2V_2) is described by:

$$\sqrt{\begin{aligned}&(F_{1_{c1}} - F_{1_{c2}})^2 + \lambda^2 * (F'_{2_{c1}} - F'_{2_{c2}})^2 \\ &+(F_{1_{v1}} - F_{1_{v2}})^2 + \lambda^2 * (F'_{2_{v1}} - F'_{2_{v2}})^2\end{aligned}}$$

Articulatory cost

The preference for less extreme articulations introduces a ranking of both static configurations and movements. The articulators can have not the same weight [12] in the evaluation of the articulatory cost.

The expression of the articulatory cost of a CV syllable produced by m articulatory parameters P is given by:

$$Art\_Cost_{CV} = \sqrt{\sum_{k=1}^{m} w_p^2 * (P_{kc} - P_{kv})^2}$$

System and energy

The contrast properties are determined by the relations between the syllables inside a system, and not the own acoustic and articulatory characteristics of each one. The prediction principle of syllabic systems consists to accomplish a research of all optimal systems.

Global efficiency

Each syllable is identified by its acoustic and articulatory characteristics. The emergence of a syllable depends, in part, on the ratio of the acoustic efficiency and the articulatory cost. This ratio constitutes the *global efficiency* described by:

$$Glob\_Eff_{cv} = \frac{Acoust\_Eff_{cv}}{Art\_Cost_{cv}}$$

Maddieson & Precoda [2] have done the observation according to that there is as many occurrences of [di] as of [du]. The computed global efficiency of [di] and [du] are roughly the same. Thus, our model has been adjusted to verify this strong observation basis.

Energy

The stability of a system is appreciated in the acoustic level by a criterion of minimal energy. The optimal systems minimize the following expression:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{d_{s_i s_j}^{\ 2}} + \beta \sum_{i=1}^{n} \left(\frac{1}{Glob\_Eff_{s_i}} \right)$$

n: number of system syllables.

The parameter β weights the intrasystemic values with regard to the intersystemic values.

5. RESULTS

Acoustic criterion

The emergence order of the twenty syllables according to the acoustic criterion is: [ba ga da be bo gi do ge ga de go ge bi du go bu gi di gu gu].

It is not strange that salient syllables as [ba] and [da] are in the top of the list.

Articulatory criterion

The emergence rank according to the articulatory criterion is: [bu bo gi gu de di ge bi be go du ga go ba da ga do gu ge gi]. We have penalized the pharyngeal syllables by an important weight for the backward displacement of the tongue body.

Global efficiency

Once, the perceptive and articulatory characteristics computed, we can evaluate the global efficiency. The list of the syllables according to their value is: [bo bu de ge be ba ga go da gi bi gu du di ga do go gi ge gu].

Syllable space

We have applied Kruskal's analysis [6] on intersyllabic acoustic distances of the prototypes to obtain a 3D space [10]. We have noted a predominance of the vowel qualities on the location of the syllables.

System energy

At this step, we want to select among the twenty syllables, a lexicon of nine syllables highly discriminable and easy to produce. For this purpose, we must generate the optimal syllable lexicon, i.e., the system of nine syllables presenting the minimum of energy. After different simulations, allowing to adjust the intrasyllabic weight (β), the system [bu ba ga da gi bi gu du di] has emerged.

6. CONCLUSION

Our *substance-based* approach can be used to predict the main tendancies of emergence of lexicons. This first step has allowed to predict the universally favoured syllables taking into account constraints of production and perception.

REFERENCES

[1] Lindblom, B., MacNeilage, P. & Studdert-Kennedy, M. (1984), *Self-Organizing Processes and the Explanation of Phonological Universals*, In *Explanation of Language Universals*. B. Butterworth et al. (Eds.), Mouton, The Hague, 181-203.

[2] Maddieson, I. & Precoda, K. (1992), Phonetic models and syllable structure, *Phonology*, vol. 9, 45-60.

[3] Maeda, S. (1989), *Compensatory Articulation during Speech: Evidence from the Analysis and Synthesis of Vocal-Tract Shapes using an Articulatory Model*, In *Speech Production and Modelling*, W.J. Hardcastle & A. Marchal (Eds.), Academic Publishers, Kluwer, 131-149.

[4] Bothorel, A., Simon, P., Wioland, F. & Zerling, J. (1986), *Cinéradiographies des voyelles et des consonnes du français*, Institut de Phonétique, Strasbourg, France.

[5] Sussman, H.M., Hoemeke, K.A. & Farhan, S.A. (1993), A Cross-linguistic Investigation of Locus Equations as a Phonetic Descriptor for Place articulation, *J. Acoust. Soc. Am.*, vol. 94, 1256-1268.

[6] Kruskal, J.B. (1977), *Multidimensional Scaling and other Methods for Discovering Structure*, In *Statistical Methods for Digital Computers. Mathematical Methods for Digital Computers*, K. Enslein et al. (Eds.), John Wiley, New-York, vol. 3, 296-339.

[7] Stevens, K.N. (1972), *The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data*, In *Human Communication: a Unifed View*, P.B. Denes & J.R. Davis (Eds), McGraw-Hill, New-York, 51-66.

[8] Liljencrants, J. & Lindblom, B. (1972), Numerical Simulation of Vowel Quality Systems: the Role of Perceptual Contrast, *Language*, vol. 48, 839-862.

[9] Vallée, N., Boë, L.J., Payan, Y. (1995),Vowel prototypes for UPSID'S 33 phonemes, In *these Proceedings*.

[10] Berrah, A.R. (1994), L'émergence des structures sonore, D.E.A. Sciences Cogntives, INP, Grenoble.

[11] Maddieson, I. (1986), *Patterns of Sounds*, Cambridge University Press.

[12] Schwartz, J.L., Boë, L.J. & Vallée, N. (1995), Experimenting the dispersion focalization theory: Phase spaces for vowel systems, In *these Proceedings*.

[13] Lindblom, B.E.F. & Lubker, J. (1985), *The Speech Homonculus and a Problem of Phonetic Linguistics*, In *Phonetic Linguistics*, V. Fromkin (Ed), Academic Press, Orlando, 169-192.