
From *Form* to *Formation* of Phonetic Structures : An evolutionary computing perspective

| | | | | |
|---|---|---|---|--|
| Ahmed-Reda Berrah Leibniz-Imag 46, Av. Félix Viallet Grenoble, France berrah@imag.fr | Hervé Glotin ICP 46, Av. Félix Viallet Grenoble, France glotin@icp.grenet.fr | Rafael Laboissière ICP 46, Av. Félix Viallet Grenoble, France rafael@icp.grenet.fr | Pierre Bessière Leibniz-Imag 46, Av. Félix Viallet Grenoble, France bessiere@imag.fr | Louis-Jean Boë ICP 46, Av. Félix Viallet Grenoble, France boe@icp.grenet.fr |
|---|---|---|---|--|

Abstract

The purpose of this paper is to explain how evolutionary computing and machine learning open new perspectives in Phonetics and Speech Science. Using these techniques, it is possible to simulate the emergence and the evolution of a common language in a society of speech robots. Experimental results show how simple local rules of interaction between robots may explain some of the universal characteristics of the phonological structure of world’s languages. On going work aiming to answer more complex questions, such as language evolution or dialect apparition, is presented.

1 INTRODUCTION

Languages have very specific forms, their phonetic structures are not completely arbitrary. Typological studies have shown that the sound systems of world’s languages exhibit systematic structural characteristics. For example, the vowel /i/ ¹ is present in 87% of world’s languages, /a/ in 87% and /u/ in 82% [13, 18]. As regards the consonants, some regularities are also found, for example, the frequency of /p,b/ is 99%, of /t,d/ is 100%, and of /k,g/ is 99% [19].

These tendencies raise an important question: Why do the sound systems exhibit universal regularities?

¹Throughout this paper, we will adopt the usual notation in Linguistics, namely phonetic characters between slashes (/./) mean the intended, phonological representation of an utterance, and characters between square brackets ([.]) stand for the actual, phonetic realisation of that utterance.

One of the ways of coping with these regularities would be to adopt the post structuralist point of view of *Linguistics* [3] following which the universals are axiomatically imposed. This consists essentially in finding ad-hoc rules that allow the structure of sound systems replicating those found in world’s languages.

We espouse however a radically different point of view, which is traditional in *Phonology* [15]. We posit that sound system universals emerge from the basic characteristics of the speech production and perception mechanisms. In other words, we take a substance-based approach, reminiscent of the study of *Phonetics*. The main notion related to this approach is that of *functional efficiency* [9], also found in evolutionary studies, and means that “nature” will favour organisms (in our case: sound systems) that adapt, or fit, the best to the environment (in our case: that allows the best accomplishment of the communication task taking into account the constraints of the speech production and perception systems). This approach shows two principal ideas: articulatory simplicity and perception distinction. The preferred sound structures in languages seem to be those which have sufficient perceptual benefits at acceptable articulatory costs.

These substance-based hypotheses have fed since the seventies several theories that aimed at explaining not only the origin of sound systems, but also their evolution and the existence of universals in languages. Two of the most important of these theories are the Stevens’ “*Quantal Theory*” [17] and Liljencrants’ and Lindblom’s “*Dispersion Theory*” [9]. The former theory establishes criteria for the choice of the individual elements (or phonemes) composing the sound systems, namely that these elements should be placed at regions where the stability between production and perception is maximum. The latter theory is more systemic in nature, in the sense that elements are selected in or-

der to maximize the perceptual contrast among them. More recently, Schwartz et al. [16] tried to unify those approaches in their “*Dispersion Focalization Theory*” yielding a comprehensive theory of vocalic systems.

One can point out two weaknesses of the theories discussed above: on the one hand, they concentrate on the perceptual distinctiveness leaving almost without development the aspects of economy of production. On the other hand, these theories try to explain the sound systems from the global point of view, i.e. paying almost no attention to the basis process of communication, namely the oral exchange between speakers and listeners.

The aim of the present paper is then twofold. First, we will show how to improve sound system predictions by explicitly taking into account the contradictory demands of economy of production and sufficient contrast in perception, applied to the case of syllable (consonant/vowel sequences) emergence. We will show how this approach refines the previous studies in explaining better the *form* of sound systems. Second, we will discuss an approach that aims at modelling the *formation* of the phonetic inventory of languages. Indeed, it will be shown how the introduction of evolutionary, as well as learning and developmental aspects can be beneficial for the understanding of the origin and change of languages.

2 THE FORM OF SOUND SYSTEMS: A MACROSCOPIC APPROACH

In the first study we are going to report in this paper, we focused on the macroscopic analysis of phonological systems, our work being a direct extension of those of Liljencrants & Lindblom [9] and Schwartz et al. [16]. Before discussing the details of our simulations and results, it is worth to note that it is possible to draw an analogy with the study of physics of gases. Namely, our macroscopic approach is similar to *Thermodynamics*. Indeed, studying a system from the thermodynamical point of view is equivalent to relate some measurable macroscopic variables [like pressure (P), volume (V) and temperature (T)] by the means of global equations (like $PV = NRT$). The underlying mechanisms from which those relationships arise are not relevant for the theory, the main goal being the correct explanation of the macroscopic level.

In the explanation of the structure of sound systems, such like in Thermodynamics, we are interesting in

finding the global laws that will enable us to say that some phonological systems are preferable in respect to others. For instance, why almost all languages in the world have the three vowels /i/, /a/ and /u/?² Is it possible to find an evaluation function “**f**” taking as argument a phonological system S (such as {/i/, /a/, /u/}) knowing that f(S) gives a measure of “goodness” of the system S? Furthermore, if it is possible to find such a function, it would be possible to classify phonological systems S_1, S_2, S_3, \dots in preference order, such that $f(S_1) > f(S_2) > f(S_3) > \dots$

Several studies have been carried out following these principles and addressed both vocalic and syllabic systems emergence [9, 10, 2, 1, 16]. The simulations presented here extend these previous studies by including an articulatory model [14] which generates vocal tract shapes in the midsagittal plane, taking eight parameters as inputs: two for the lips, one for the jaw, one for the larynx and four for the tongue. Examples of shapes can be seen in Figure 5. The midsagittal contour is converted into an area function with which the transfer function of the vocal tract is computed. The first four maxima of the transfer function are called formants. Their frequency position and bandwidths are taken as the output of the model.

Among syllables of lexicals of 31 languages [13], Consonant/Vowel (CV) syllables are the most frequent. A syllable is an opening gesture in the vocal tract. It can in this way be identified by both the changes in the articulatory parameters from the consonant to the vowel, and the changes in the short-term spectrum of the sound, stylised by the trajectory in the formant space. As the plosive /p b t d k g/ being the most occurred consonants, and /i a u/ being the most frequent vowels followed by /e' o'/, we can say, without taking any number of risks, that most occurred syllables are /ba da ga bi di gi bu du gu/ followed by /b'e' d'e' g'e' b'o' d'o' g'o'/ (Cf. *Figure 1*). Our research consisted in predicting the rank of efficiency of Consonant/Vowel (CV) syllables among all the 20 possible combinations of stop-like /b d g ?/ (/??/ is a little observed pharyngeal stop) and vowel-like /i 'e' a 'o' u/.

The selection of a syllable depends, in part, on the ratio of the acoustic efficiency (*Acoust Eff*) and the articulatory cost (*Art Cost*). This ratio constitutes the *Global Efficiency* described by:

$$Glob\ Eff_{CV} = \frac{Acoust\ Eff_{CV}}{Art\ Cost_{CV}} \quad (1)$$

²/i/ like in “feed”, /u/ like in “food” and /a/ like in “father”.