

# PHONETIC CODE EMERGENCE IN A SOCIETY OF SPEECH ROBOTS: EXPLAINING VOWEL SYSTEMS AND THE MUAF PRINCIPLE

Ahmed-Reda Berrah & Rafael Laboissière  
Institut de la Communication Parlée – INPG  
46, Av. Félix Viallet, 38031, Grenoble, France  
E-mail: {berrah,rafael}@icp.grenet.fr

## ABSTRACT

The purpose of this paper is to explain how it is possible to simulate the emergence of a common phonetic code in a society of speech robots using evolutionary techniques. Simulations of the prediction of vowel systems and the explanation of the Maximum Use of Available distinctive Features (MUAF) principle are discussed. These experimental results show how simple local rules of interaction between robots may explain some of the universals characteristics of the phonological structure of world’s languages. On going work aiming to answer more complex questions, such as the use of supplementary features in large vowel systems, is presented.

## 1. INTRODUCTION

The origins and evolution of language is still clouded in mystery. Typological studies [7] have shown that the sound systems of world’s languages exhibit systematic structural characteristics. The most common hypothesis is that language is based on human innate ability and on the refinement of innate language (universal grammar) by a parameter setting process [4]. According to this theory, all humans are born with the same set of distinctive features. Then, they select the set of distinctive features used by their mother tongue. Still, there are a number of fundamental problems with this theory such that most proposed feature set are not able to account for all the sounds found in world’s languages. We espouse however a radically different point of view, which posits that sound system universals emerge from the basic characteristics of the speech production and perception mechanisms as well as from the speaker/listener interaction (cf. [6]). The main notion related to this approach is that of **functional efficiency** 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 allow the best accomplishment of the communication task taking into account the constraints of the speech production and perception systems).

The present paper reports some advances in solving the problem of emergence of a common phonetic code in a

group of distributed communicating agents. First, we will describe the general emergent model on which our simulations are based. Then, we will discuss the simulation results of two different studies: prediction of vocalic systems and integration of the MUAF principle.

## 2. METHODS

Speech communication can be seen as a sequence of perceptual objectives negotiated between a speaker and a listener. The term “negotiation” in this context means that (1) speakers and listeners have to agree on a common code, and that (2) speakers can modulate the physical realisation of the utterances as a function of environmental conditions, like noise, and extra-linguistic information, as long as the listener is able to decode the message (cf. Lindblom’s hypo-hyper scale [6]).

An important question immediately raises: how are the perceptual objectives negotiated? The answer to this question lies in a deep comprehension of the mechanisms involved in speaker-listener interactions. Of course, this is a overwhelming task. In the simulations presented here, we implemented a simplified version of local interactions between speech agents, who can communicate using vowel-like, static sounds.

The main plan is to breed a “society” of speech robots [1] that can adapt their productions through local interactions, called hereafter *transactions*. For each simulation, the number of distinct phonetic symbols (i.e. the size of the lexicon) is fixed in advance. This is in contrast with other similar recent studies, in which a society of communicating agents can change the number of symbols during the course of a simulation [3].

### 2.1. Speech Robot

Speech robots in our model are simple agents which can produce sustained vowels, represented at a perceptual space. This space is composed of subspaces, each one representing a particular vocalic feature. Although some of this subspaces are unidimensional (adequately describing vowel features like duration or nasalization), we will

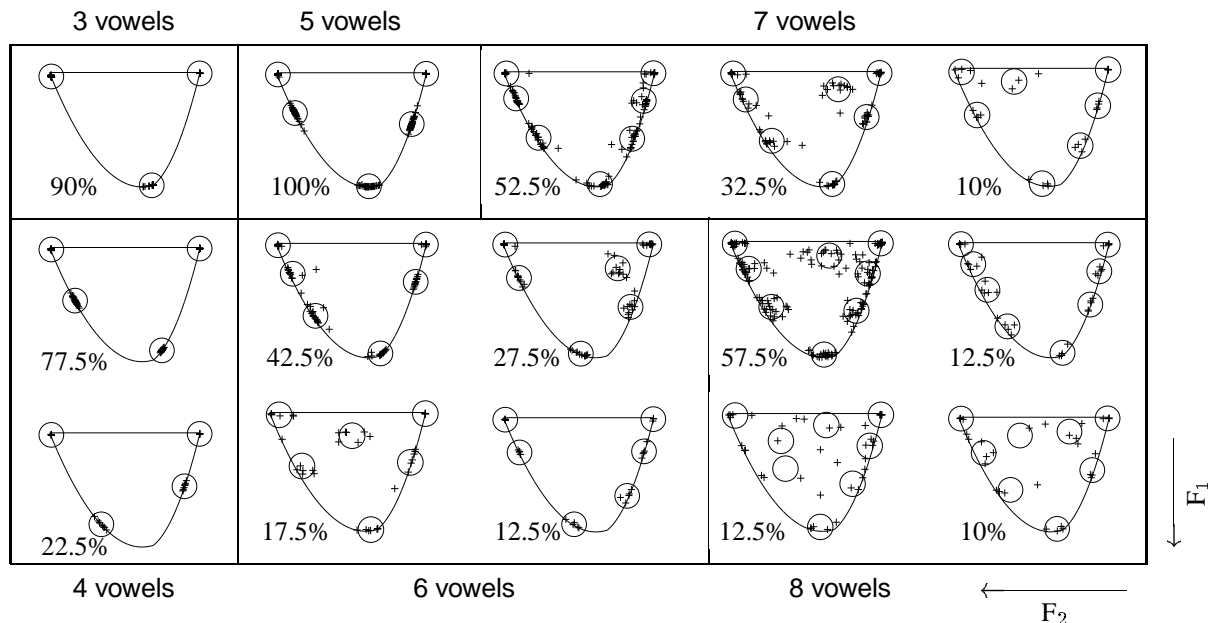


Figure 1: Distribution of vowels, used by the agents when they are constrained to use a lexicon of 3, 4, 5, 6, 7 and 8 vowels, in the vocalic triangle (projection of the maximal vocalic space [2] on the  $F_1$ - $F_2$  plan).

always consider one of the subspaces as the bidimensional formant space  $F_1$ - $F_2$ .

In a former version of our model [5], the speech production level was also included, thanks to the inclusion of an articulatory model. We decided to not include it in the present work in order to thoroughly understand both the mechanisms and the implications of the perceptual level alone.

## 2.2. Speaker-listener transaction

Each robot of the society has a lexicon composed of a fixed number of items randomly initialised. Pairs of robots are randomly chosen to communicate using an item of their lexicons. We define an interaction between two robots as a transaction, in which one of the robots takes the place of the speaker and the other that of the listener. The speaker chooses randomly one of its items and produces it. The listener robot relates the features of the signal sent by the speaker to its own lexicon at the perceptual level. Finally, it adjusts its items in order to adapt its lexicon to this new sensed item.

The listener adaptation to a new sensed item constitutes the critical point of the simulations. Indeed, lexicon convergence, key of our problem, depends in a great part on it. We suppose that a person who learns a new word (or sound) will try to repeat this word (or sound) at first. So, a speech robot, which perceives an unknown symbol for it, will try to repeat it. This is achieved by moving in the for-

mant space the nearest item closer to the perceived item, and moving the others away in order to avoid confusion risks and to maximise the perception distinction. Due to the limitation of space, the details of the simulations will be published elsewhere.

## 3. RESULTS AND DISCUSSION

Despite the striking simplicity of our model, the simulations replicate well-known phenomena observed in phonetic inventories of world's languages. We will briefly discuss the simulation results of three different studies.

### 3.1. Vowel Systems

The first one concerns prediction of vowel systems with small ( $\leq 8$ ) number of items. In these simulations, the feature space is composed only of the formant space. We aim to obtain, after an important number of transactions, a common lexicon for all speech robots, which will confronted with the universals of world's languages. In this case, the theoretical possibility that language units are the fruit of exchanges and cooperations based on perceptual distinctivity, would be plausible and coherent.

Vowels are generally represented in a vocalic triangle. This triangle is the projection of the maximal vocalic space on the  $F_1$ - $F_2$  plan. The principal parameters involved in the simulations are population size (typically 10 agents), lexicon size (3-8 vowels), and maximal number of transactions (5000). We ran 40 simulations for each

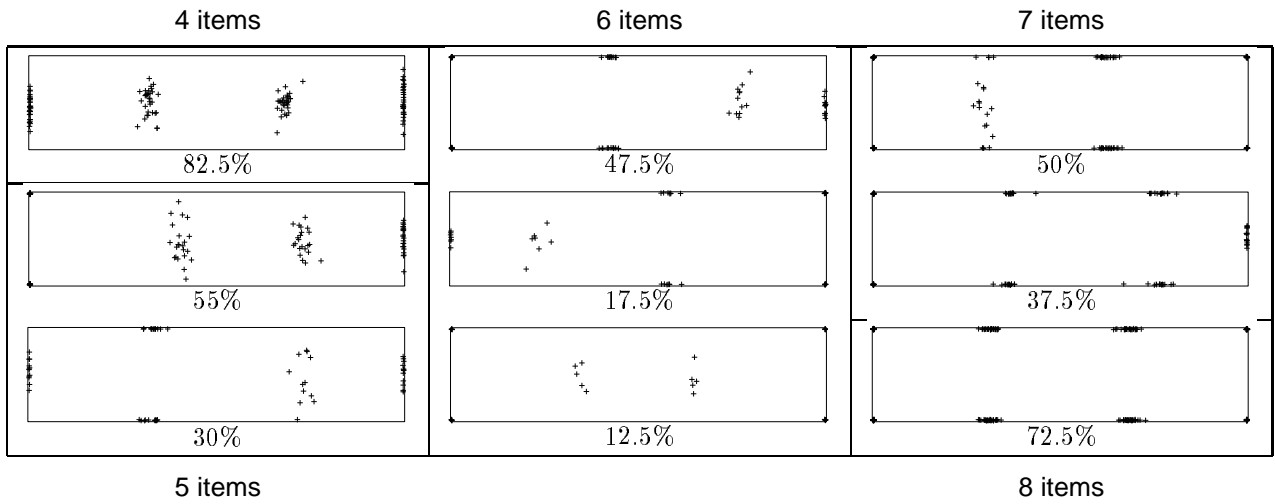


Figure 2: Progressive exploitation of the second dimension in systems using 4-, 5-, 6-, 7- and 8-items.

lexicon size and we computed the mean lexicon reached by the population. After that, we grouped similar systems together and plotted each group (what is called a “system type”) as a panel in Fig. 1. The percentage value appearing aside each vocalic triangle represents the proportion of each system type.

As shown in Fig. 1, for 3- and 5-sized lexicons the society settles systematically to /i,a,u/ and /i,e,a,o,u/ respectively. Two 4-vowel systems are possible: /i,e,a,u/-like and /i,a,o,u/-like. These predictions are in accordance to the statistics of world’s languages [2]. For 6-, 7- and 8-sized lexicons, several different systems, observed also in world languages, are obtained.

It is important to notice that our simulations produce vowel systems in which the items are mainly placed along the /i/-/e/-/a/ and the /u/-/o/-/a/ continua. Both central vowels (close to /ə/) and high-front vowels (like /y/) are nearly inexistent, which is in contradiction with the statistics of world’s languages. Introduction of a more elaborated formant space, including  $F_3$ , or the use of the Focalization Theory [2] could remedy this problem.

### 3.2. MUAF principle

In the second set of simulations, we explore what J.J. Ohala called “Maximum Use of Available distinctive Features (MUAF)” in structuring phonetic inventories. The MUAF principle is invoked to explain why languages with crowded vowel inventories tend to use additional features to preserve distinctiveness, like nasalization and duration. Lindblom [6] suggests the introduction of developmental constraints to explain why phonetic systems obey the MUAF principle: acquisition of new forms will be easier when their associated motor commands overlap with the

previously stored motor patterns. Everything else (articulatory cost and perceptual salience) being equal, learning a new form that overlaps with old patterns should involve storing less information in memory than acquiring one with nothing in common with old items, since part of the new motor score is already in storage.

We propose a modified version of our emergent model to explain the MUAF principle in vowel system. The goal is to replicate the fact that once a feature is used, it is explored systematically before going to the next. The rules of adaptation were also changed, such that repulsion of different items in the listener lexicon is not done using an Euclidean distance: only one of the two features will be used at a time (namely that for which the gain in distinctiveness is greater).

In order to test this idea, a synthetic exemple was implemented, in which the perceptual space is two-dimensional and limited by a rectangle. The features are here the horizontal and vertical axes. As in the previous case, we ran 40 different simulations for each lexicon size (4–8 items), stopping at 5000 iterations. The results (grouped mean lexica) are summarized in Fig. 2. The figure shows that (1) only one feature is explored with lexicons of small size (under 4-items systems), and (2) progressively, a new feature is used for large systems (8-items, for instance, systems exploit the two dimensions, being equivalent to a doubled 4-items system). Fig. 3 shows the result obtained using an Euclidean metric, in which the both dimensions are explored (a) and that obtained with principle described here (b).

Beyond the simplicity of this simulation, the results reveal how a particular metric associated to our general model could replicate an interesting phenomenon: how systems

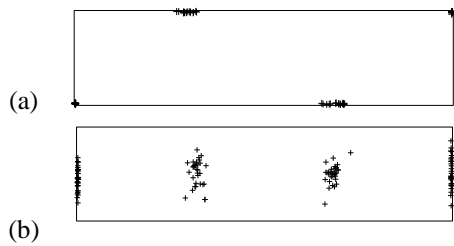


Figure 3: Results for two different distance functions used computing repulsion between items: (a) Euclidean and (b) MUAF principle

are contented with only one dimension while it is possible to explore it, otherwise, they exploit progressively a new dimension.

### 3.3. Additional Feature for Vowels

We were tempted to use this concept in a speech communication experiment and to realize a similar experiment with vowels. We suggest the introduction of a feature, the perceptual space becoming three-dimensional (two for the formants and one for the additional feature). We combined the two precedent simulation principles so that repulsion of different items in listener is carried out in only one dimension ( $F_1$ - $F_2$  or along the new feature) at a time depending on whether the gain in distinctiveness is greater to the duration difference or not. Fig. 4 shows two preliminary results. left panels represent projections of 3-D perceptual space into the  $F_1$ - $F_2$  plan. (a.1) and (a.2) show a 5-vowel systems which is a /i,e,a,o,u/-like as in the first simulation (see Fig. 1). We can see that new feature is not explored in (a.2) contrarily to (b.2) which represents a 8-vowel systems. One could say that this last system is also a variant of /i,e,a,o,u/ but the more interesting is that, like shown in (b.1), vowels tend to use the additional feature and that is quite clear for /i/, /a/ and /u/.

## 4. CONCLUSION

The above preliminary results on vowels universals strengthen the hypothesis that language is a cultural phenomenon which is created actively by speakers and evolves under functional pressures. Indeed, such an approach seems to be very promising in the investigation of systematic phonetic properties.

On going simulation should allow the introduction of geographical distribution for the agents. This concept is materialized by a matrix containing the probabilities of communication among the agents. The population occupies a regular 2D grid.

Moreover, such experiments, inspired by evolutionary and learning processes, can answer fundamental questions

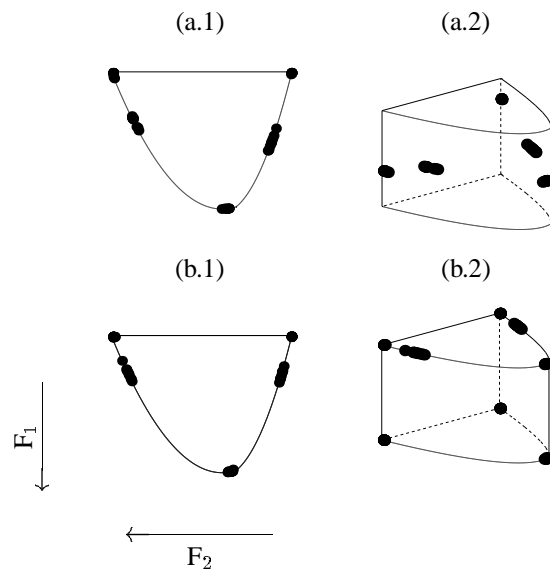


Figure 4: Final lexica for simulations in an extended perceptual space, composed of the formant space and an additional feature. (a) 5 vowels, (b) 8 vowels.

which are essential for building a true theory of language: How has language originated in the first place? How can language be learned? Why do language keep evolving? And why are there so many different languages?

## 5. REFERENCES

- Berrah, A.R., Glotin, H., Laboissière, R., Bessière, P. & Boë, L.J. (1996) From form to formation of phonetic structures: An evolutionary computing perspective. *Proceedings of the 13th International Conference on Machine Learning Workshop on Evolutionary Computing and Machine Learning*, In Fogarty, T. & Venturini, G. (Eds), 23-29.
- Boë, L.-J., Schwartz, J.-L. & Vallée, N. (1994). The prediction of vowels systems: Perceptual contrast and stability. *Fundamentals of Speech Synthesis, and Speech Recognition*, In Keller E. (Ed.), John Wiley & Sons, 185-212.
- de Boer, B. (1997). Emergent vowel systems in a population of agents. *Proceedings of the 1997 European Conference on Artificial Life*, Brighton, England.
- Chomsky, N. (1957). *Syntactic Structures*, Mouton & Co., La Haye.
- Glotin, H. & Laboissière, R. (1996) La vie artificielle d'une société de robots parlants : Émergence et changements du code phonétique, In *Du Collectif au Social — Actes de Journées de Rochebrune*, ENST, Paris, 113-125.
- Lindblom, B. (1996) Systemic constraints and adaptive change in the formation of sound structure. *Evolution of Human Language*, In Hurford, J. (Ed), Edinburgh Univ. Press., Edinburgh.
- Maddieson, I. (1986) *Patterns of Sounds*. 2nd edition, Cambridge University Press, Cambridge (1st edition: 1984).
- Steels, L. (1996) Synthesising the Origins of Language and Meaning using Co-evolution and Self-organisation. *Evolution of Human Language*, In Hurford, J. (Ed), Edinburgh Univ. Press., Edinburgh.