The Epigenesis of Syllable Systems : an Operational Model.

Pierre-yves Oudeyer
Sony Computer Science Lab, Paris, France
e-mail : py@csl.sony.fr

Abstract
A computational model of the origins of syllables systems is presented : a society of robotic agents endowed with realistic motor, perceptual and cognitive apparati is shown to build from scratch shared syllable systems in a decentralized manner. Furthermo re, these systems share many structural properties with those of human languages.

Résumé
Un modèle opérationel de l'origine des systèmes syllabiques est présenté : une société d'agents robotiques munis d'appareils moteurs, perceptuels et cognitifs réalis tes construit un système de syllabes partagées entre eux de manière décentralisée. De plus, ces systèmes partagent de nombreuses proriétés structurelles avec ceux des langues du monde.

Introduction

The phonological repertoires of human languages have very well defined structural tendencies (Vennemann,1988). For instance, the vowels /i/ and /a/ are present in 87 percent of languages, and /p/ and /b/ in 99percent. More strikingly, phonemes are composed in very precise ways : while all languages have CV syllables, few allow CCVC ; also, phonemes tend to be ordered in order of increasing sonority until the peaks of syllables, and then in order of decreasing sonority.

Why is this so ? Several answers have been proposed. On the one hand, the Chomskian post -structuralist point of view explains these structures by a number of formal rules, which can be generative rules as in the Principle and Parameters Theory (Chomsky and Halle 1968) or a linguistically specific formal constraint system like in Optimality Th eory (Archangeli and Langendoen 1997). An example of rule in OT is the *COMPLEX rule constraint which states that syllables can have at most one consonant at an edge. These rules are supposed to be known innately by children (they are coded in the genome) , which would explain why they learn so easily the idiosyncracies of languages. This point of view is more and more challenged by biological evidence (no language acquisition device has ever been found) and by the more fundamental question of the origins o f speech : where do these rules come from ? How could they have been introduced in the genome ?

In the past 20 years, a number of researchers have argued that this structure could be explained by the self - organizing interaction between generic motor, per ceptual, cognitive, social and task constraints (Lindblom 1992, Steels 1998). A number of experiments and models have already been devised for phoneme and syllable repertoires. Lindblom (1992) and Schwartz et al. (1995) showed that the optimization of a n umber of analytically defined perceptual and motor constraints allowed to predict which phonemes systems occur more frequently in world's languages (they are the ones that have the lowest costs). Lindblom, (1992), and then Schwartz extended this result ove r CV syllables, and then Redford (1998) for syllables of any type.

While these models are clearly steps forward a credible alternative to innatist theories, there needs a lower level model that explains not only the ``why'', but also the “how” : how this optimization process could have been implemented in reality by human societies (in particular can it be implemented completely culturally ?), and how humans could have negotiated in a decentralized manner so as to settle on a particular solution that will constitute their sound system (indeed, there are many good solutions, as found by Redford) ?

Indeed, it is difficult to imagine that a genius once made a complicated calculation in order to pick up an optimal system and then imposed it to all the populatio n.

The present work presents such a model. It is based on previous work by de Boer (1999) about vowel systems, and is part of a current of research on the origins of language consisting in using the multi -agent paradigm (Steels 1998). de Boer (1999) developed a society of agents, each endowed with non-linguistically specific motor, perceptual and cognitive constraints, who tryed to build a shared vowel system from scratch and in a decentralized cultural self -organized manner. He showed that the produced v owel systems are very similar in type and frequency to human languages. Here we use the same modeling framework for syllable systems. The model was made much more realistic, and is also more complex because it involves more sofisticated control, perceptual and cognitive apparati.

Next section gives an overview of this setup. Then section 3 shows that the systems allows to answer the two ``how'' questions. Section 4 concludes.

The experimental setup

Agents are endowed with a motor, a perceptual and a cognitive apparatus. They use these to interact in the framework of the imitation game (de Boer, 1999). A game consists in picking up two agent, one is the speaker and the other the hearer. The speaker ut ters one syllable of its repertoire, and the hearer tries to imitate it given its knowledge/skills about the sound world at this point. Then the speaker gives feedback about whether he judged the imitation was good or not. This is the case when he categori zes the imitation as he categorizes its initial utterance.

The production system of agents consists on the one hand of a physical model of the vocal tract based on (Cook, 1989) to which we added stiffness and intertia features, and on the other hand of a control system that steers it when executing articulatory programs. Articulatory programs are the specifications of the items in agent's repertoire, corresponding to the syllables of human languages. Here we approach the concept of syllable from the viewpo int of the frame/content theory of MacNeilage (1998) : it is one oscillation of the jaw within which are inserted articulatory targets. Note that in the model the targets are not pre - programmed to be inserted in a particular manner, so agents may have at s ome point items whose targets imply such a deformation of the pure frame that they actually perform 2 oscillations. Each articulatory execution is given a cost which depends on the length of the displacement of articulators and how well targets were reach ed. The co -articulation strategy used is very similar to the one described in (Massaro 1998), and models well both anticipative and perseverative influences among targets.

The perceptual apparatus of agents is composed on the one hand of a model of the c ochlea based on (Lyon, 1997), and on the other hand on a similarity measure between two perceptual/cochlear trajectories implemented with dynamic time warping, which is a widely used technique in speech recognition.

The cognitive architecture of agents c onsists in 2 memories of examplars. Examplars are associations between an articulatory program and the associated perceptual trajectory. The first memory consists in prototype examplars that represent the syllables of the agent's repertoire. It is initiall y empty, and grows either by invention (modeling a pressure for more forms coming for example from the need to develop a larger lexicon), or by learning new syllables from others. New syllables are learnt when an imitation fails (but this is not systematic, because an agent may think an imitation failed because he had already a prototype for what he heard, though a bad one). Agents constantly try to improve their prototypes thanks to their second memory. The later is composed of many examplars and is gene rated initially through babbling : agents try random articulatory programs and perceive the produced sounds. They remember these examplars with a probability inverse to their articulatory cost, which models the principle of least effort (Linblom 1992). This memory has an imposed limited size to model cognitive limited resources, and evolves along with time. When used to improve the imitation of a given sound, the closest item is found and a number of small variations are tryied and added to the memory. To keep with the size limit, some other random examplars are pruned. As a consequence, at the beginning, the exemplars are spread across all the space, and progressively concentrate in the regions appropriate to the syllable system currently built.

Predictions of the model

20 runs of the experiment were made (doing more was difficult, since one run lasts 1 day). Each experiment consisted in 20000 interactions with 20 agents ; the number of pregiven phonemes was 10 and they were allowed to have up to 4 ta rgets within a syllable (which means the syllable space was approximately of size 12000). Phonemes consisted in 4 vowels, 3 liquids and 3 plosives. The inverse -mapping memory of agents had an upper size limit of 100, which means they could never know more than 0.8 percent of the syllable space at a given point in time.

The first thing one wants to know is simply whether populations of agents manage to develop a sound system of reasonable size and that allows them to communicate (imitations are successful). Both mean success in imitation over the last 50 interactions (each run consisted in 20000 interactions) and mean size of syllable repertoires of agents (there were 20 agents) were monitored. The result was that the mean success in the last 1000 interact ions over the 20 runs was 96.9 percent while the mean number of syllables in their repertoires was 79.1.

Now that we have seen that a communication system was effectively built, one has to look whether the structural properties of the produced repertoire s of syllables resemble human syllable systems. A first study about the syllable types of the produced systems was achieved. Statistics about the set of all the syllables produced by the 20 runs were computed (for each run, measured were done after 20000 g ames). The following order of most frequently used types of syllables was found : CV > CVC > CC > CCV > CVVC/CCVC/CVC. This order is significantly different than the order found in randomly generate syllable systems, and identical to the one found in huma n languages, except for the presence of CC in the 3rd position.

A second important tendency of human languages is the ``sonority hierarchy principle". In the 20 runs, 70.9 percent of used syllables at the end of simulations respected this principle, w hile chance is only 5.3 percent. In order to investigate the role of the jaw oscillation constraint, it was removed and statistics were made on the resulting systems : 21.5 percent of produced syllables respected the principle. The conclusion is that the jaw constraint is the main responsible but not the only one. The interaction between the fact that syllables always have to start from a rest closed jaw and the fact that syllables must be different enough in order to be well distinguished can account for part of it.

Conclusion

The model proposed a plausible operational solution to the two previously unanswered ``how" que stions expressed above : 1) our society of realistic agents manage to negotiate in a decentralized manner and to settle on a particular syllable system that will constitute their sound system ; 2) the syllable systems they choose have structural properties typical of human languages, and this is due to only local interaction among generic non-linguistically specific constraints.

References

Archangeli D., Langendoen T. (1997) Optimality theory, an overview, Blackwell Pulishers.

de Boer, B. (1999) Inve stigating the Emergence of Speech Sounds. In: Dean, T. (ed.) Proceedings of IJCAI 99. Morgan Kauffman, San Francisco. pp. 364 -369.

Chomsky, N. and M. Halle (1968) The Sound Pattern of English. Harper Row, New york.

P. R. Cook, "Synthesis of the Singing Voice Using a Physically Parameterized Model of the Human Vocal Tract," Proc. of the International Computer Music Conference, pp. 69 -72, Columbus, OH, 1989.

Lindblom, B. (1992) Phonological Units as Adaptive Emergents of Lexical Development, in Ferguson, Menn, Stoel -Gammon (eds.) Phonological Development: Models, Research, Implications, York Press, Timonnium, MD, pp. 565-604.

Lyon, R. (1997), All pole models of auditory filtering, in Lewis et al. (eds.)
Diversity in auditory mechanics, World Scientific Publishing, Singapore.

Massaro, D. (1998) Perceiving talking faces, MIT Press.

MacNeilage, P.F. (1998) The Frame/Content theory of evolution of speech production. { \it Behavioral and Brain Sciences}, 21, 499-548.

Pinker, S., Bloom P., (1990), Natural Language and Natural Selection, The Brain and Behavioral Sciences, 13, pp. 707-784.

Redford, M.A., C. Chen, and R. Miikkulainen (1998) Modeling the Emergence of Syllable Systems. In: Proceedings of the Twentieth Annual Conference of the Cognitive Science Society. Erlabum Ass. Hillsdale.

Schwartz, J.L., Boe, L.J., Vallée, N. (1995) Testing the Dispersion -Focalization Theory: Phase Spaces for Vowel Systems, XIIIth Int. Congr. of Phonetics, 1, 412 -415.

Steels, (1998), Synthesizing the origins of language and meaning using coevolution, self -organization and level formation, in Hurford, Studdert Kennedy, Knight (eds.), Cambridge University Press, pp. 384 -404.

Steels L., Oudeyer P -y. (2000) The cultural ev olution of syntactic constraints in phonology, in Bedau, McCaskill, Packard and Rasmussen (eds.), Proceedings of the 7th International Conference on Artificial Life, pp. 382-391, MIT Press.

Vennemann, T. (1988), Preference Laws for Syllable Structure, B erlin: Mouton de Gruyter.