Journal of Phonetics 53 (2015) 55–60

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Research Article

Open challenges in understanding development and evolution of speech forms: The roles of embodied self-organization, motivation and active exploration

Pierre-Yves Oudeyer [a,b,*]

[a] Inria, France
[b] Ensta ParisTech, France

ARTICLE INFO

Article history:
Received 5 January 2015
Received in revised form
30 June 2015
Accepted 9 September 2015
Available online 21 October 2015

Keywords:
Origins of speech forms
Self-organization
Curiosity
Active exploration
Development
Evolution
Evo-devo

ABSTRACT

This article discusses open scientific challenges for understanding development and evolution of speech forms, as a commentary to Moulin-Frier, Diard, Schwartz and Bessière, 2015. Based on the analysis of mathematical models of the origins of speech forms, with a focus on their assumptions, we study the fundamental question of how speech can be formed out of non-speech, at both developmental and evolutionary scales. In particular, we emphasize the importance of embodied self-organization, as well as the role of mechanisms of motivation and active curiosity-driven exploration in speech formation. Finally, we discuss an evolutionary-developmental perspective of the origins of speech.

1. Comparing theories of speech formation in a unified Bayesian framework

Studying the forms and formation of speech has long been a topic of tremendous interest for cognitive science in general. It has been repeatedly used in the last century as the cradle in which alternative theories of language as well as sensorimotor control have been expressed and debated. Jakobson (1941) used it as a strong ground for the early elaboration of structuralist theories of cognition. Later on, it has been the pivot of theories of perception, and their potential links to action (Galantucci, Fowler, & Turvey, 2006), as well as theories of language development in the child (Oller et al., 2013). It has also gathered efforts in the quest for understanding the origins of language, where a mystery is how linguistic forms can arise, be shared and evolve in a population of individuals (Kirby, Griffiths, & Smith, 2014; Moulin-Frier, Diard, Schwartz & Bessière, 2015; Oudeyer, 2006; Steels, 2011).

Across these scientific enterprises, mathematical and computational modeling has been prominently used in the latest decades, grounded in the physics of the speech system and in the dynamics of neural and learning architectures. Such models constitute a formal language allowing us to formulate and analyze precisely hypotheses about complex mechanisms. Yet, an obstacle to scientific progress has been that alternative theories have often been expressed through different formal languages, making it challenging to articulate and compare them in a single framework. This challenge applies both to models of speech evolution (e.g. Berrah, Glotin, Laboissière, Bessière, & Boë, 1996; Browman & Goldstein, 2000; de Boer, 2000; Liljencrants & Lindblom, 1972; Oudeyer, 2005a; Pierrehumbert, 2006; Wedel, 2011) and models of speech acquisition (e.g. Guenther, 1994; Howard & Messum, 2014; Moulin-Frier, Nguyen & Oudeyer, 2014; Warlaumont, Oller, Buder, & Westermann, 2013). In this perspective, the COSMO Bayesian framework (Moulin-Frier et al., 2015) is a significant step forward as it leverages Bayesian modeling to develop a formal framework that allows us

* Correspondence address: Inria, France. Tel./fax: +33 1 524574030.
E-mail addresses: pierre-yves.oudeyer@inria.fr,
URL: http://www.pyoudeyer.com

0095-4470/$ - see front matter © 2015 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2015.09.001

to formulate in a unified manner many key theories of speech (Moulin-Frier, Laurent, Bessière, Schwartz, & Diard, 2012), as well as theories of how speech forms can arise in populations of individuals.

A particularly useful feature of such a Bayesian framework is that it constrains model designers to be as explicit as possible on the assumptions of their models. Furthermore, Bayesian modeling allows compact expression of relationships between subparts of a model, abstracting details of implementation to highlight the global functional dynamics. In their article, Moulin-Frier et al. show in detail how various theories of speech perception and production compare in the context of communication loops among individuals. They also show how such a framework can encode the dynamics of so-called "language games" to account for the formation of shared speech codes in a population of individuals, as well as to explain why certain vowel and consonant structures are more frequent than others in world languages.

2. Speech from non-speech: open questions

As such formalism provides a compact and general view on a large family of models of the formation and learning of speech structures, it also affords identifying open scientific questions. This can be done in particular through analyzing the assumptions of the COSMO framework. As Moulin-Frier et al. very clearly state, COSMO attempts to explain some properties of phonological systems out of speech communication principle, i.e. out of (Bayesian) optimization processes that lead individuals to learn and negotiate a communication system that is efficient under physiological constraints. Furthermore, thanks to the method of Bayesian modeling, Moulin-Frier et al. also make it explicit that their model assumes mechanisms that solve requirements of "adequacy" (availability of a system of forms easy to produce and perceive), "parity" (capability to play symmetric roles in speech interaction) and "reference" (capability to use a device like pointing to ensure shared attention on a referent).

Thus, the model relies on a pre-existing set of linguistic abilities, as well as abstracts away from many non-linguistic processes, such as sensorimotor development outside speech, non-linguistic activities such as sensorimotor coordination in joint tasks, or properties of the body outside the speech system.

This in itself is not a weakness of the framework, especially because this is made explicit. But it points to a very important question, formulated already long ago by Lindblom (1984): "[Which are the processes that allow to] derive language from non language"?. This question applies at several scales: development in individuals, cultural evolution and phylogenetic evolution in populations.

At the developmental scale, one needs to explain how young infants come to discover and master speech sounds, and how they understand that these sounds can be used to produce effects on their social peers and coordinate with them. Infants indeed are not born with a refined understanding of what "speech communication" is, and optimization processes driving the formation of (speech) codes efficient for communication may hardly be at work at the beginning. Indeed, notions of "code" and "communication" shall themselves be formed though cognitive and social development, leveraging in particular the capability to assign new functions to behaviors previously mastered (which has been called "functional flexibility", Oller et al., 2013).

At the evolutionary scales, cultural and phylogenetic, an analogous mystery is still far to be solved: how did the capacity to linguistically communicate through speech or gestures appear out of non-language? Language games models such those presented in Moulin-Frier et al. have mostly focused on the question of how a shared linguistic convention can form and change at the population level, but assuming the capacity to handle the syntax and meaning of these language games, i.e. assuming a capacity for language. But how did language form in communities of individuals who did not already have such tools to build and negotiate a linguistic system?

Speech and language, in general, are embedded in a network of diverse non-linguistic activities, as well as influenced by constraints due to biological implementation of the body and the brain: what consequences can this have on the formation of speech forms at developmental and evolutionary scales?

Such non-speech mechanisms from which speech communication shall emerge, or be influenced by, are bound to have consequences both at the individual/developmental level and at the population/evolutionary level, acting as structure providers and filters constraining and guiding the formation of speech forms.

We will now discuss three families of non-speech mechanisms that may be useful starting points to further understand how speech can be formed out of non-speech: self-organization and spontaneous pattern formation in physical systems, the role of intrinsic and extrinsic motivational systems, and finally some commonalities between speech development and the development of other sensorimotor skills through the prism of active exploration.

2.1. Embodied self-organization of speech forms

Nature is full of complex organized patterns, in particular in the inorganic world: spiral galaxies, sand dunes, deltas of rivers, polyhedrons in water foam, ice crystals are all macro-patterns that spontaneously form out of the physical interaction of their micro-components. Such self-organization of structures, due to the laws of physics in complex systems, is also at play in the living world. For example, it has been identified in the formation of spots and stripes on the skin of animals, of hexagonal honeycombs, or for organizing group behavior in insects or birds (Ball, 1999). At the developmental scale, these spontaneous patterns can be at play to generate organized behavior without any process of explicit optimization. At the evolutionary scale, such self-organized developmental processes can act as constraints, or might have been recruited and shaped to serve optimally a functional purpose.