

On the cognitive nature of speech sound systems

Jean-Luc Schwartz, Clément Moulin-Frier, Pierre-Yves Oudeyer

▶ **To cite this version:**

Jean-Luc Schwartz, Clément Moulin-Frier, Pierre-Yves Oudeyer. On the cognitive nature of speech sound systems. Journal of Phonetics, 2015, Special Issue : "On the cognitive nature of speech sound systems", 53, pp.1-175. 10.1016/j.wocn.2015.09.008 . hal-01222752

HAL Id: hal-01222752
https://hal.science/hal-01222752

Submitted on 3 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Special Issue for the Journal of Phonetics

On the cognitive nature of speech sound systems

Jean-Luc Schwartz, Clément Moulin-Frier, Pierre-Yves Oudeyer

During the last 50 years, the question of the cognitive nature of phonological units has followed the rhythm of the persistent debate between auditory and motor theories of speech communication. Though recent advances in cognitive neuroscience and cognitive psychology have largely renewed this debate, a consensus is still out of reach, and the true nature of speech units in the human brain remains elusive.

A dimension of importance in this debate is a *systemic* one: speech units are not isolated, they are part of a phonological system, and they obey structural principles regarding well-investigated properties as distinctiveness, compositionality, contextual dependencies or systemic regularities. The phonological system itself is also part of a complex network of interaction with low-level biomechanical and sensory-motor systems, with higher-level brain structures regulating cognition, emotion and motivation, and finally with the social structures in which all these systems are embedded.

Connecting assumptions or theories about the nature of speech units with a structuralist view about the relationship between phonetic properties and phonological systems has given rise to a number of major breakthroughs in speech science, for instance Lindblom's bridges between the Variable Adaptive Theory (or its Hyper-Hypo variant) of speech communication (Lindblom, 1990) and the Dispersion Theory of vowel systems (Lindblom, 1986); or Stevens' Quantal Theory (Stevens, 1972, 1989) addressing both the invariance issue and the search for the origins of distinctiveness and phonetic features; or the tandem between the Motor Theory of Speech Perception (Liberman & Mattingly, 1985) and Articulatory Phonology (Browman & Goldstein, 1992) in the Haskins Labs.

This Special Issue is centered around a target paper by Moulin-Frier et al. that aims at relating the question of the auditory vs. motor vs. perceptuo-motor nature of speech units with simulations of vowel, plosive and syllable systems of human languages emerging from agent interactions, in a computational Bayesian framework. In this context, the papers in the special issue explore further the systemic perspective, studying how various dimensions of physical, cognitive, motivational and interactional systems can inform our understanding of the origins of speech forms.

Contributions were expected from researchers known for their interest in both speech units and sound systems, using indifferently the tools of cognitive neuroscience and cognitive psychology, experimental phonetics and computational modeling (various ingredients of the "laboratory phonology" program). The systemic component could be addressed by various means either directly (through contributions focused on the question of the content and

implementation of phonetic features and the link between phonetics and phonology) or through the various temporal scales at which diachrony operates, in relation with developmental, cultural or phylogenetic evolution.

This resulted in a set of 12 papers that we organized and present in the following way.

In the target paper, Moulin-Frier and coll. present COSMO (for "Communicating about Objects using Sensory-Motor Operations"), a model of the cognitive architecture supposed to be at hand in speech communication. COSMO is based on the hypothesis that the communication process between two agents is internalized in each agent's brain, which results in an architecture combining phonological units as well as sensory and motor variables into a single probabilistic distribution. This is then implemented in the framework of Bayesian reasoning. This allows the implementation and comparison of auditory, motor and perceptuo-motor theories of both speech production and speech perception, in a single probabilistic framework. Moulin-Frier and colleagues develop simulations of interaction paradigms between pairs of agents, based on what they call "deictic games". They present simulations of vowel, plosive and plosive-vowel systems and compare them with preferred systems in phonological databases.

The participants of the special issue were left free to integrate or not the content of the target paper in their own contribution. Six papers are actually directed towards one or another ingredient of the simulations presented in the target paper.

Fowler applauds the sensory—motor model put forward in COSMO and the use of a probabilistic framework to capture the nature of speech communication. She underlines the potential of this approach to infer premises of both Lindblom's dispersion theory and Stevens' quantal theory, but she claims that the motor theory implemented in the COSMO framework is actually different from the theory put forward by Liberman and coll. She bases her claims on her analysis of the Bayesian equations derived in the target paper, and particularly on the lack of an explicit term linking sensory variables to motor variables (the "inversion"). She concludes by the stance, similar in the motor theory and in the direct realist theory (Fowler, 1986), according to which produced and perceived language forms are "the same thing, publicly available phonetic gestures".

Nguyen & Delvaux focus their contribution on the concepts of imitation and inter-speaker convergence. They describe in detail the various forms taken by convergence processes according to which two interacting speakers tune their actions in respect to their interlocutor, which results in decreasing the distance between their productions, at multiple levels. This produces alignment processes, which may concern the whole communication chain, from the choice of linguistic units and structures to modifications in sounds and physical postures. They claim that these imitation and convergence processes play an important role in learning mechanisms at hand both in first language learning in children and second language learning in adults, and that it should be incorporated in a model such as

COSMO. They discuss how such kinds of interaction principles could be added to COSMO and what could be their role, particularly in the emergence and stabilization of diversity and innovation in linguistic systems.

Oudeyer acknowledges the interest and potential of the unified Bayesian framework at hand in COSMO to compare predictions of different theoretical frameworks in speech communication. However, he introduces additional mechanisms in computational scenarios trying to let speech forms emerge from non-speech. Firstly he insists on the role of dynamic structures, both at the level of the individual (related with the coupled dynamic properties of sensory—motor systems) and the collectivity (associated with the dynamics of self-organization and the stabilization of dynamic behavior). Then he puts forward the role of motivation and curiosity in the processes of (vocal) exploration involved in the dynamics of developing and socially interacting individuals. He discusses how such mechanisms could contribute to bridge speech and non-speech systems. Finally he stresses the interest of an evo-devo approach associating in a single framework individual and collective pressures in the emergence of human languages.

Warlaumont focuses her comment on syllables and the specific role attributed to the Frame-Content Theory (FCT) put forward by MacNeilage and colleagues (MacNeilage, 1998; MacNeilage et al., 2000) in this process. She recalls a number of possible caveats about the FCT, but most importantly she claims that the syllabic process cannot be conceived as a pure consequence of an independent and pre-existing biomechanical oscillation. Quite on the contrary, she proposes that syllables emerge from an exploration-and-learn sensory—motor process, and she presents two possible mechanisms by which infants could actually learn and control to produce syllabic structures, one based on curiosity and the other one on the dynamics of a neural spiking model. She concludes on the possibility to associate a biomechanical bootstrap such as FCT with adequate tuning/learning processes towards a complete theory of the emergence of syllables – that remains to be elaborated.

Plummer and Beckman discuss three major fundamental limitations of the target paper in their view. Firstly they raise doubts on the evaluation process at hand in the COSMO simulations, related to the difficulty to link phonological inventory databases with simulations. Then they argue that simulations should focus on diversity rather than similarity of human languages. Finally they focus on ontogeny rather than phylogeny and discuss in detail why they think that the assumptions in the target paper could not be applied to speech and language development. Indeed, they notice that referential communication develops later than emotional communication in infants, and stress the intrinsic asymmetry between the infant and the caregiver in the learning and communication process at hand in development. They suggest that a key question is compositionality and its relation with prosody, and conclude with the role and importance of dyadic interactions and imitative processes in the emergence of human languages.

De Boer notices the interest of the Bayesian approach in COSMO to explicit and assess hypotheses in a Bayesian computational framework – but he considers that this addresses only a small part of the whole question. He insists on the cultural aspect of the evolutionary mechanisms, which may largely hide the visibility of the cognitive processes at hand in communication. He illustrates this by the fact that cultural optimality is different from individual optimality, and that "decisions" at the level of the collectivity may be different from what they are at the individual level. He then introduces a number of proposals about the interaction between biological, cognitive and cultural processes in language emergence. This involves various "experimental" approaches to language emergence, exploiting either behavioral paradigms with simulated language games, or simulations focusing on specific questions at the level of individual behavior.

The other contributors provide at distance a number of echoes to the questions raised in the target paper though in a completely independent way.

Kröger and Cao present their model of phonetic-phonological emergence in a perceptuo-motor framework anchored in the concept of topologically-organized maps (Self-Organizing Maps). They describe the architecture of their model based on both sensory-motor and semantic inputs, and discuss in the context of specific simulations how phonological units might emerge in development. For this aim, they consider a developmental sequence beginning by sensory—motor babbling then tuning the system on the language input through imitation driven by both sensory-motor and semantic information. This results in progressively elaborating both a "Phonetic Map" and a "Semantic Map" related through learning co-occurrence stimuli. They discuss how phonetic representations and units might emerge, with some features (such as syllabic structure, consonant voicing or vowel height) directly driven by sensory—motor processes while other features such as consonant manner of place of articulation would require more precise attunements driven by the link with semantic maps at a further stage.

Ma and colleagues come back on the nature and content of syllabic units around the question of the possible differences in the coherence of intrasyllabic units from one language to the other. Considering the possibility that a language such as Mandarin Chinese could rely less on phonemes and more on syllables in speech production than Indo-European languages, they propose a quantitative test based on articulatory data. They analyze a corpus of data obtained by an electromagnetic magnetometer, and involving VCV sequences uttered by three native speakers of Mandarin Chinese and three native speakers of French. Their data reveal in their view different patterns of coarticulation in the two languages, with across-syllable anticipatory coarticulation in French but not in Mandarin – while coarticulation is strong within syllables in both languages. They interpret these results in relation with their original hypothesis, though they describe in detail possible limitations of their study in terms of statistical power associated with "case-studies" typical of speech production corpora.

Messum and Howard focus their contribution on one key problem in the study of both speech development and the cognitive representations of phonological units: the resolution of the correspondence problem, in which a child finds the way to associate her own productions with those of her caregiver in reference to a given phonological target. They notice that the problem is generally solved by a call to imitation processes, which involves a judgment of similarity between the target and the trial. This “Similarity Based Equivalence” process should operate in spite of the potential “opacity” of the sensory information provided by the target stimuli: between other difficulties, in the case of a difference in morphology, the child must capture a normalized information which is perhaps out of her abilities. Therefore the authors introduce another mechanism that they call “Mirrorred Equivalence”, in which the caregiver interprets the stimulus provided by the child and provides a reinterpretation enabling the infant to solve the correspondence problem. They discuss in detail why the Mirrorred Equivalence is compatible with behavioral data. Indeed they show that it provides a framework for the relationship between speech perception and production, discuss a number of consequences in terms of learning and communication, and conclude on the perceptuo-motor nature of the speech code, compatible with the Mirrorred Equivalence process.

Redford presents a theoretical paper dealing with speech production, in which she attempts to relate the literature from psycholinguistics – associating productions with meaning, in relation with compositionality and sequencing, and dealing with constraints of the working memory – with the literature on motor control – on plans and programs, learning and automation. This integration is attempted in a developmental framework compatible with a usage-based approach of phonology. Schemas are put at the center of the reasoning. Hence the paper discusses what these schemas could be and how they are linked with both the general linguistic project of the speaker that is to produce meaning and communication, and the requirements of motor control routines. Redford discusses how the automatization of schemas in development might shed light on the adequate units in the phonological process at hand in speech production.

In the last paper of this special issue, Oh and colleagues come back to the inventories from a functional point of view. Indeed, they propose a complete reanalysis of phonological inventories in relation with lexical inventories, thanks to the central concept of functional load. This concept enables to quantify what does a given phonological contrast explain in the formation of lexical units. The authors firstly describe their linguistic data, based on lexical corpora with 9 languages. They introduce the mathematical framework for defining functional load, which basically consists in computing what is the cost for a lexicon to loose a given phonological contrast. This enables them to describe the functional load carried by vowels, consonants, stress and tones, and the distribution of load within individual segments. The authors come back on the hypothesis of the “consonantal bias” (according to which there would be more functional weight for consonants in the lexicon and for vowels in morphology), and discuss all these findings in relation with the psycholinguistic literature.

As a conclusion, let us come back on the choice of this topic for the Journal of Phonetics, and for phonetics in general. It is striking to see how strongly cognitive sciences have entered in the field more and more over the years. In some sense, the “Lindblom’s program” to derive language from non-language (Lindblom, 1984) has been really fruitful, allowing cognitive and developmental psychology to shed light on phonology. The neuroscience (r)evolution of the last twenty years provided an abundance of new tools, facts and proposals fostering the perceptuo-motor link. Also “Steels’ program” inviting us to play language games with robots and computational agents (Steels, 1994, 1997) impelled a series of advances in which interactions between agents in development and in society has enabled us to propose mechanisms for the emergence of phonology in ontogeny, cultural evolution and phylogeny. The current special issue presents a view on this exciting and active matter of studies and thoughts, and possibly helps us to know a little more about segments, syllables, prosody, normalization, development, perception, production and phonology in general. We hope that this contribution will induce more and more works in this domain in the coming years.

Acknowledgements

This work benefitted from funding from the European Research Council under the European Community's Seventh Framework Programme (ERC Advanced Grant 339152 and ERC Starting Grant 240007).

References

Browman, C. P., & Goldstein, L. M. (1992). Articulatory Phonology: An Overview. Phonetica, 49, 155–180. http://dx.doi.org/10.1159/000261913

De Boer, B. (2015). Biology, culture, evolution and the cognitive nature of sound systems. Journal of Phonetics, this issue.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics, 14, 3–28.

Fowler, C.A. (2015). COSMO's "motor theory" is not the motor theory of Liberman, Cooper, and Mattingly. Journal of Phonetics, this issue.

Kröger, B.J., Cao, M. (2015). The Emergence of Phonetic-Phonological Features in a Biologically Inspired Model of Speech Processing. Journal of Phonetics, this issue.

Ma, L., Perrier, P., & Dang, J. (2015). Strength of syllabic influences on articulation in Mandarin Chinese and French: Insights from a motor control approach. Journal of Phonetics, this issue.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21, 1–36.

Lindblom, B. (1984). Can the models of evolutionary biology be applied to phonetic problems?. In Proceedings of the 10th international congress of phonetic sciences (pp. 67–81), Utrecht.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala, & J. J. Jaeger (Eds.), Experimental phonology (pp. 13–44). New York: Academic Press.

Lindblom, B. (1990). On the notion of possible speech sound. Journal of Phonetics, 18, 135–152.

MacNeilage, P. (1998). The frame/content theory of evolution of speech production. Behavioral and Brain Sciences 21, 499-546.

MacNeilage, P. F., Davis, B. L., Kinney, A., Matyear, C. L. (2000). The motor core of speech: a comparison of serial organization patterns in infants and languages. Child Development 71, 151-163.

Messum, P., & Howard, Y.S. (2015). Creating the cognitive form of phonological units: the correspondence problem could be solved by mirrored vocal interactions in infancy rather than by the imitation of speech sounds. Journal of Phonetics, this issue.

Moulin-Frier, C., Diard, J., Schwartz, J.L., & Bessière, P. (2015). COSMO ("Communicating about Objects using Sensory-Motor Operations"): a Bayesian modeling framework for studying speech communication and the emergence of phonological systems. Journal of Phonetics, this issue.

Nguyen, N., & Delvaux, V. (2015). Role of imitation in the emergence of phonological systems. Journal of Phonetics, this issue.

Oh, Y.M., Coupé, C., Marsico, E., & Pellegrino, F. (2015). Bridging Phonological System and Lexicon: Insights from a Corpus Study of Functional Load. Journal of Phonetics, this issue.

Oudeyer, P.Y. (2015). Open challenges in understanding development and evolution of speech forms: the roles of embodied self-organization, motivation and active exploration. Journal of Phonetics, this issue.

Plummer, A.R., Beckman, M.E. (2015). Framing a socio-indexical basis for the emergence and cultural transmission of phonological systems. Journal of Phonetics, this issue.

Redford, M.A. (2015). Unifying Speech and Language in a Developmentally Sensitive Model of Production. Journal of Phonetics, this issue.

Steels, L. (1994). The artificial life roots of artificial intelligence. Artificial Life Journal , 1, 89–125.

Steels, L. (1997). The synthetic modeling of language origins. Evolution of Communication, 1, 1–34.

Stevens, K. N. (1972). The quantal nature of speech: evidence from articulatory-acoustic data. In E. E. Davis, Jr., & P. B. Denes (Eds.), Human communication: A unified view (pp. 51–66). New-York: Mc Graw-Hill.

Stevens, K. N. (1989). On the quantal nature of speech. Journal of Phonetics, 17, 3–45.

Warlaumont, A.S. (2015). Modeling the Emergence of Syllabic Structure. Journal of Phonetics, this issue.