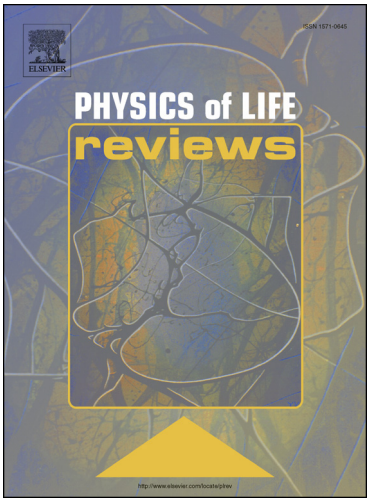
Phonology in the mirror

Jean-Luc Schwartz, Marie-Lou Barnaud, Pierre Bessière, Julien Diard, Clément Moulin-Frier

Accepted Manuscript



Phonology in the mirror

Jean-Luc Schwartz, Marie-Lou Barnaud, Pierre Bessière, Julien Diard, Clément Moulin-Frier

PII: S1571-0645(16)00008-7
DOI: http://dx.doi.org/10.1016/j.plrev.2016.01.007
Reference: PLREV 701

To appear in: *Physics of Life Reviews*

Received date: 15 January 2016
Accepted date: 21 January 2016

Please cite this article in press as: Schwartz J-L, et al. Phonology in the mirror. *Phys Life Rev* (2016), http://dx.doi.org/10.1016/j.plrev.2016.01.007

Phonology in the mirror

Comment on "Towards a Computational Comparative Neuroprimatology: Framing the language-ready brain" by Michael A. Arbib

Jean-Luc Schwartz (1), Marie-Lou Barnaud (1), Pierre Bessière (2), Julien Diard (3), Clément Moulin-Frier (4)

(1) GIPSA-Lab
CNRS,
Univ. Grenoble Alpes
Grenoble, France

(2) ISIR
CNRS, Sorbonne Univ.,
Univ. Pierre et Marie Curie
Paris, France

(3) LPNC
CNRS,
Univ. Grenoble Alpes
Grenoble, France

(4) SPECS
Universitat Pompeu Fabra
Barcelone, Spain

The contribution by M.A. Arbib over the years and as it appears summarized and conceptualized in this paper (Arbib, 2016) is admirable, extremely impressive, and very convincing in many aspects. A key value of this work is that it systematically attempts to introduce formal conceptualization and modeling in the reasoning about facts and interpretations.

We would like to focus on a component of language – actually minor in Arbib's chapter – that is phonology, in light of a model of speech communication developed by our group.

If the mirror system paved the way for a "language-ready brain", this should include a "phonology-ready brain". Among the seven properties of the "language-readiness" listed in Section 1.6, two of them should provide the basis for crucial aspects of a phonological system. Firstly, property (1), "Complex action recognition", involves both action analysis, necessary for learning the components of a complex vocal action and decomposing it into phonological segments, and action chunking, enabling to utter phonological sequences likely to convey meaning. Secondly, property (2), "parity", ensures that the communicative value of a phonological unit plays the same role for the speaker and the listener.

Let us begin by parity. In Section 4.1, Arbib insists that the mirror system does *not* correspond to a motor theory of speech perception. He proposes that vocal actions (speech utterances) can be recognized and understood by "general mechanisms which need not involve the mirror system strongly", and that the mirror system would just complement such general mechanisms – which is actually the basis of the model elaborated by Moulin-Frier & Arbib (2013) and presented in the paper. This is where our computation studies could shed some more light on when general mechanisms for recognizing phonemes could suffice, and when the mirror system could be useful.

Indeed, we developed COSMO, a sensori-motor model for speech communication, associating general auditory processing and specific mirror mechanisms by which the listener may exploit motor knowledge associated to the way he/she would produce a sound similar to the acoustic stimulus to be recognized (Moulin-Frier et al., 2012). COSMO (for "Communicating Objects using Sensori-Motor Operations") is based on the assumption that the whole communicating process by which an "object" (in a very general sense) is communicated by a speaker to a listener by motor actions generating sensory stimuli, is fully internalized inside an individual's brain (see Fig. 1).

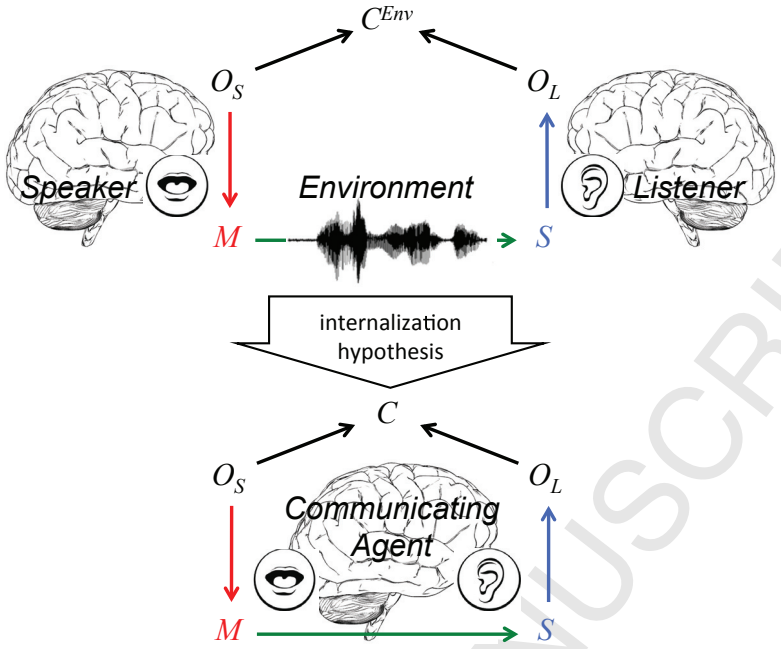


Figure 1. The COSMO model: the communication situation, which involves a speaker and a listener interacting within an environment, is internalized in communicating agents.

A COSMO agent is then fully described by a probability distribution associating the "object" to be produced (O_S, S for speaker), the object to be perceived (O_L, L for listener), the motor action M, the sensory stimulus S and a variable C ensuring the success of communication, by a probability distribution $P(C, O_S, M, S, O_L)$. If considered objects O_L and O_S are restricted to phonemes, COSMO enables to define what would be a general mechanism for auditory recognition of a phoneme, characterized by the distribution $P(O_L \mid S)$, and a motor recognition mechanism based on motor inference equivalent to the mirror system hypothesis, by the distribution $P(O_S \mid S)$. Bayesian inference enables to show that the motor recognition mechanism $P(O_S \mid S)$ actually involves sensori-to-motor inversion and articulatory decoding, according to the formula:

$$P(O_S \mid S) \propto \sum_M P(M \mid O_S)\, P(S \mid M)$$

From the COSMO model and the phoneme recognition models it yields, three major results emerge, that precise and complete the notes by Arbib on the relative roles of general processes and mirror processes in speech decoding (Moulin-Frier et al., 2012, Laurent et al., submitted).

(1) Under conditions of perfect learning of the whole communication process, auditory recognition and motor recognition actually provide exactly the same performance – this is an "indistinguishability theorem" which sheds very interesting light on the longstanding difficulty to provide clear cut experimental arguments for proving or disproving auditory or motor theories of speech perception;

(2) When learning is imperfect (as is unavoidably the case in human communication), then the auditory recognition performs better for learned stimuli, and the motor recognition is better in noise or adverse conditions;

(3) In all situations, combining auditory and motor mechanism thanks to sensori-motor fusion enhances accuracy and robustness.

Additional experiments by Moulin-Frier et al. (2015) on interaction games between COSMO agents display realistic simulations of vowel, plosive or syllable systems in line with phonological inventories available in large databases of human sound systems.

What are the lessons of these simulations? Firstly, that there are underlying principles in the way auditory and motor theories should be compared and combined (e.g. in the framework of the "Perception-for-Action-Control Theory, PACT, Schwartz et al., 2002, 2007 2012), and that these principles could guide future experiments on phonological representations in the human brain, and the development of future computational models in the field.

Secondly, that the phoneme is likely a perceptuo-motor construction, represented both in the ventral and in the dorsal stream, which is actually compatible with the regions likely associated with phonological coding and representation in the brain – posterior superior temporal sulcus, angular gyrus, supramarginal gyrus, all situated in the dorsal pathway but close to and tightly connected to auditory areas.

This is where we return to the second set of questions about the learnability and compositionality of phonological units and rules. Property (1) about complex action recognition assumes that the brain is able to parse a complex vocal action into its subcomponents – or likely words into phonemes. Phoneme discovery is classically a difficult scientific problem, considering the intimate coupling of successive phonemes into words making their decomposition not transparent inside the acoustic stream because of coarticulation processes. In COSMO, we explored a small part of the general problem, by showing how an agent learning to perceive plosive-vowel syllables could learn similarities in vowels in the auditory space, and similarities in consonants in the articulatory space (Laurent et al., submitted).

This shows once again that phonemes would be perceptuo-motor units rather than pure auditory or pure motor units – and that in this sense, they are probably not only ingredients of the "vocal pantomime" in the dorsal pathway, but possible elements of convergence of sound and gesture closely associated to the extraction of meaning in the ventral pathway.

We must remember the origin of the motor theory of speech perception: it was based on the idea that phonetic invariance at the basis of phonological processes would be articulatory or motor rather than acoustic or auditory. With COSMO, we propose that invariance is in fact perceptuo-motor in nature. We suggest that the phonology-ready brain would probably put phonology in the mirror. We claim that, while much of the phonological descriptions are rather speaker-oriented and conceived as part of the speech production process, there is still a long road ahead for defining a "phonology for the speaker and for the listener" which would connect speech perception and speech production in an integral speech communication system.

Acknowledgements
The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007 – 2013, Grant Agreement no. 339152- "Speech Unit(e)s").

References

Arbib, M.A. (2016). Towards a Computational Comparative Neuroprimatology: Framing the language-ready brain. *Physics of Life Reviews*, this issue.

Laurent, R., Barnaud, M.L., Schwartz, J.L., Bessière, P., & Diard, J. (submitted). A computational assessment of the role of motor information in speech perception.

Moulin-Frier, C., & Arbib, M.A. (2013). Recognizing speech in a novel accent: the motor theory of speech perception reframed. *Biological Cybernetics*, 107, 421–47.

Moulin-Frier C., Laurent R., Bessière P., Schwartz J.-L., & Diard J. (2012). Adverse conditions improve distinguishability of auditory, motor and perceptuo-motor theories of speech perception: an exploratory Bayesian modeling study. *Language and Cognitive Processes*, 27, 1240–1263.

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,* 53, 5-41.

Schwartz, J.L., Abry, C., Boë, L.J., & Cathiard, M. (2002). Phonology in a theory of perception-for-action-control. In J. Durand, B. Laks (eds.) *Phonology: from Phonetics to Cognition* (pp. 255-280). Oxford: Oxford University Press.

Schwartz, J.L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception for Action Control Theory (PACT): a perceptuo-motor theory of speech perception. *Journal of NeuroLinguistics*, 25, 336-354.

Schwartz, J.L., Boë, L.J., & Abry, C. (2007). Linking the Dispersion-Focalization Theory (DFT) and the Maximum Utilization of the Available Distinctive Features (MUAF) principle in a Perception-for-Action-Control Theory (PACT). In M.J. Solé, P. Beddor & M. Ohala (eds.) *Experimental Approaches to Phonology* (pp. 104-124). Oxford University Press.