COSMO, a Bayesian computational model of speech communication: Assessing the role of sensory vs. motor knowledge in speech perception
Marie-Lou Barnaud, Julien Diard, Pierre Bessiere, Jean-Luc Schwartz

To cite this version:
Marie-Lou Barnaud, Julien Diard, Pierre Bessiere, Jean-Luc Schwartz. COSMO, a Bayesian computational model of speech communication: Assessing the role of sensory vs. motor knowledge in speech perception. 5th International Conference on Development and Learning and on Epigenetic Robotics (ICDL-Epirob), 2015, Providence, RI, United States. <hal-02004350>
Modeling the concurrent development of speech perception and production in a Bayesian framework

Marie-Lou Barnaud1, Julien Diard2, Pierre Bassire3, Jean-Luc Schwartz1

1 GIPSA-lab, UMR 5221 - CNRS & Université Grenoble Alpes, France
2 LPNC, UMR 5105 - CNRS & Université Grenoble Alpes, France
3 ISIR, UPMC, UMR 7222 - CNRS & Université Paris Sorbonne, France

It is widely accepted that both auditory and motor representations intervene in the perceptual processing of speech units. However, the question of the functional role of each of these systems remains seldom addressed and poorly understood. This is where computational models can play a crucial role. We developed COSMO, a Bayesian model comparing sensory and motor processes in the form of probability distributions which enable both theoretical developments and quantitative simulations.

Introduction

The three trends about the invariants of speech perception

Auditory theories: the invariants are acoustic properties.
Motor theories: the invariants are the intended phonetic gestures of the speaker.
Perceptuo-motor theories: the invariants are perceptuo-motor units characterized by both their articulatory coherence and their perceptual value.

Context and Bayesian model

Issues

Question of interest: Why perceptuo-motor units?
Hypothesis: The auditory and the motor systems would be complementary.

COSMO: from a model of communication...

COSMO: ... to a model of communicating agent (the internalization hypothesis)

Model learning and simulation of perception processes

Model learning

Sensory learning

Motor learning

Simulation of perception processes

Comparison of three perception processes:
- Sensory system P(O | S)
- Motor system P(O | S)
- Perceptuo-motor system P(O | S) (C=1)

In 1D simulations:
- 2 objects O1 and O2
- Spaces S and M in one dimension
- P(O | S) and P(O | S) Gaussians

Results

Learning pace: Comparison of the entropy of the sensory and motor learning

The entropy of the sensory model converges quickly to a level close to the entropy of the stimuli produced by the master (in less than 200 learning steps), while the entropy of the motor model converges more slowly (in more than 20,000 learning steps).

Evaluation of perception: Comparison of recognition rates of the three perception processes in noisy conditions

The sensory system provides good recognition scores without noise, with a quick degradation of performance when noise is added. The motor system is better than the sensory model in noise, though still being poorer without noise. The perceptuo-motor system performs better than both isolated systems in all conditions.

Discussion

The sensory model is of lower variance than the motor model and yields a less uncertain categorization probability distribution than the motor categorization process. By contrast, the motor model is of larger variance than the sensory model and generalizes better. Consequently, in nominal condition, both systems are able to categorize the stimulus but the sensory system is better than the motor system. In adverse condition, the sensory system performs a random categorization whereas the motor system succeeds to correctly categorize.

Conclusion

We have compared and illustrated in detail the behavior and the performance of the learned sensory and motor models. We have shown that the sensory model directly exploits the associations between objects and stimuli to learn the sensory classifier in a quick and efficient way. In contrast, the motor model needs to build both motor repositories and an internal model of the sensory-motor mapping. In order to do so the Learning Agent explores its motor space. As a result, the motor model is less efficient for the processing of stimuli typical of the learning set ("nominal conditions") but more robust to degradations and adverse conditions. The motor model has some generalization capacities thanks to its explorative phase, when the sensory model, in some sense, overlearned. This is what we summarize as the "sensory narrow-band vs. motor wide-band" property.

Perspectives

- We have already extended COSMO to more complex configurations in multi-dimensional spaces involving synthetic plosive-vowel sequences.
- We are currently exploring further the learning algorithm and its ability to produce "idiomsynchronies" which are variations in learned motor and sensory strategies in the learning agent.

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(s)”, PI Jean-Luc Schwartz).