> *Open fully-funded PhD position:* **"Hybrid Bayesian and deep neural modeling for weakly supervised learning of sensory-motor speech representations"**

The Deep-COSMO project, part of the new AI institute in Grenoble, is welcoming applications for a 3-year, fully funded PhD scholarship starting October, 1st, 2019 at GIPSA-lab (Grenoble, France)

TOPIC: Representation learning, speech production and perception, Bayesian cognitive models, generative neural networks
RESEARCH FIELD: Computer Science, Cognitive Science, Machine Learning, Artificial Intelligence, Speech Processing
SUPERVISION: J. Diard (LPNC); T. Hueber, L. Girin, J.-L. Schwartz (GIPSA-Lab)
IDEX PROJECT TITLE: Multidisciplinary Institute for Artificial Intelligence – Speech chair (P. Perrier)
SCIENTIFIC DEPARTMENT (LABORATORY'S NAME): GIPSA-lab
DOCTORAL SCHOOL: MSTII (maths and computer science) or EEATS (signal processing) or EDISCE (cognitive science), depending on the candidate's profile and career plan
TYPE of CONTRACT: 3-year doctoral contract
JOB STATUS: Full time
HOURS PER WEEK: 35
SALARY: between 1770 € and 2100 € gross per month (depending on complementary activity or not)
OFFER STARTING DATE: October, 1st, 2019

SUBJECT DESCRIPTION:

General objective

How can a child learn to speak from hearing sounds, without any motor instruction provided by his/her environment? The general objective of this PhD project is to develop a **computational agent**, able to learn speech representations from raw speech data in a weakly supervised configuration. This agent will involve an **articulatory model of the human vocal tract**, an **articulatory-to-acoustic synthesis system**, and a **learning architecture** combining **deep learning algorithms and developmental principles** inspired from cognitive sciences. ***This PhD will be part of the "Speech communication" chair in the Multidisciplinary Institute for Artificial Intelligence in Grenoble (MIAI).***

Method

This work will capitalize on two bricks of research **recently developed in Grenoble**. First, **a Bayesian computational model of speech communication** called **COSMO (Communicating about Objects using SensoriMotor Operations)** (Moulin-Frier et al., 2012, 2015; Laurent et al., 2017; Barnaud et al., 2019) was jointly developed by GIPSA and LPNC. This model associates speech production and speech perception models in a single architecture. The random variables in COSMO represent the signals and the sensori-motor processes involved in the speech production/perception loop. COSMO learns their probability distributions from speech examples provided by the environment, and is then able to perceive and produce speech sounds associated to speech categories. So far, COSMO was mostly tested on synthetic data. One of the main challenges is now to confront COSMO to real-world data.

Second, we will also capitalize on **a set of computational models for automatic processing and learning of sensory-motor distributions in speech developed at GIPSA**. This comprises **a set of transfer-learning algorithms** (Hueber et al., 2015, Girin et al. 2017) aiming at adapting acoustic-articulatory knowledge on one speaker, towards another speaker, using a limited amount of data, possibly incomplete and noisy; together with **a set of deep neural networks able to process raw articulatory data** (Hueber et al., 2016; Tatulli & Hueber, 2017).

The first step will consist in designing, implementing and testing **a "deep" version of COSMO**, in which some of the probability distributions are implemented by **generative neural models** (e.g. VAE, GAN). This choice is motivated by the ability of such techniques to deal with raw, noisy and complex data, as well as their flexibility in terms of transfer learning. The second stage will consist in **reformulating entirely the speech communication agent in an end-to-end neural architecture**.

Outputs

The system will be tested in terms of both efficiency of the learning process – hence ability to generate realistic speech sequences after convergence – and coherence of the motor strategies discovered by the computational agent, in spite of the fact that no motor data will be provided for learning. The outputs are both (1) theoretical – for **better**

understanding the cognitive processes at hand in speech development and speech communication; (2) technical – **for integrating knowledge about speech production and cognitive processes in a machine learning architecture**; and (3) technological – for proposing a **new generation of autonomous speech technologies**.

ELIGIBILITY CRITERIA

Applicants must have:

- A Master's degree (or be about to earn one) or have a university degree equivalent to a European Master's (5-year duration), in Computer Science, Cognitive Science, Signal Processing or Applied Mathematics.
- Solid skills in Machine Learning or probabilistic modeling + General knowledge in natural language processing and/or speech processing (an affinity for cognitive sciences and speech sciences is welcome).
- Very good programming skills (mostly in Python).
- Good oral and written communication in English.
- Ability to work autonomously and in collaboration with supervisors and other team members.

SELECTION PROCESS

Applicants will have to send their CV + an application letter in English + copy of their last diploma to:
Jean-Luc.Schwartz@gipsa-lab.fr, Thomas.Hueber@gipsa-lab.fr;
Letters of recommendation are welcome. Contact before preparing a complete application are welcome too.
Applications will be evaluated as they are received: the position is open until it is filled, with deadline on July 10th, 2019.