SECTOR: Higher Education Institution

INSTITUTION: Univ. Grenoble Alpes, University of Innovation

One of the major research-intensive French universities, Univ. Grenoble Alpes\(^1\) enjoys an international reputation in many scientific fields, as confirmed by international rankings. It benefits from the implementation of major European instruments (ESRF, ILL, EMBL, IRAM, EMFL\(^*\)). The dynamic ecosystem, grounded on a close interaction between research, education and companies, has earned Grenoble to be ranked as the 5th most innovative city in the world. Surrounded by mountains, the campus benefits from a natural environment and a high quality of life and work environment. With 7000 foreign students and the annual visit of more than 8000 researchers from all over the world, Univ. Grenoble Alps is an internationally engaged university.

A personalized Welcome Center for international students, PhDs and researchers facilitates your arrival and installation.

In 2016, Univ. Grenoble Alpes was labeled «Initiative of Excellence ». This label aims at the emergence of around ten French world class research universities. By joining Univ. Grenoble Alpes, you have the opportunity to conduct world-class research, and to contribute to the social and economic challenges of the 21st century ("sustainable planet and society", "health, well-being and technology", "understanding and supporting innovation: culture, technology, organizations" "Digital technology").

\(^*\) ESRF (European Synchrotron Radiation Facility), ILL (Institut Laue-Langevin), IRAM (International Institute for Radio Astronomy), EMBL (European Molecular Biology Laboratory), EMFL (European Magnetic Field Laboratory)

Key figures:

- + 50,000 students including 7,000 international students
- 3,700 PhD students, 45% international
- 5,500 faculty members
- 180 different nationalities
- 1st city in France where it feels good to study and 5th city where it feels good to work
- ISSO: International Students & Scholars Office affiliated to EURAXESS

LOCATION: France, Grenoble

---

\(^1\) https://edu.univ-grenoble-alpes.fr/en/
RESEARCH FIELD: **multimodal machine perception**

RESEARCHER PROFILE:

- Recognized researcher (PhD holder not yet fully independent)
- Established researcher (Researchers who have developed a level of independence)

JOB PROFILE (Description): Deep generative models for speech enhancement with audio and audio-visual data

Speech signals recorded in a natural environment are generally largely corrupted by noise. Speech enhancement consists of cleaning the speech signal, i.e. removing the noise to obtain an estimation of the underlying clean speech signal. This is a crucial step for exploiting speech information in, e.g., automatic speech recognition and dialog systems. The overall goal of the proposed post-doc position is to develop cutting-edge machine learning methodologies for speech enhancement. The developed algorithms will eventually be implemented onto a companion robot for Human-robot social interaction.

This post-doc work aims to explore the use of deep generative models for speech enhancement. Deep generative machine learning models combine the principle of deep models, i.e. they basically consist of deep neural networks, and the principle of generative models, i.e. they model the way data are generated through explicit modeling of data distributions. One prominent example of such models are variational auto-encoders (VAEs) [1] which involve the projection of modeled data onto a so-called latent space of reduced dimension, and (re)generation of those data from the latent space. We already have exploited the VAE model for speech enhancement in [2, 3]. This model has proven very efficient for this task in an unsupervised configuration, where no clean and noisy datasets are necessary for model training, thanks to a separate model training phase (on clean speech signals) and a clean speech signal inference phase (from noisy data). This has led to very good generalization to unseen types of noise.

In the present post-doc work, we aim to further investigate and improve the use of this model for speech enhancement, with several different non-exclusive directions:

- At present, [2, 3] do not consider temporal dependencies across the different random variables, i.e. the VAE is exploited for independent frame-wise speech processing. Yet, speech signals have typical temporal dynamics that are routinely modeled and exploited in speech processing. Temporal models will thus be plugged into the VAE framework, e.g. in the line of the recurrent neural network (RNN) proposed in [4]. Different forms of temporal dependencies will be explored.

- The VAE-based speech enhancement model will be extended to a multimodal (audiovisual) configuration, where the video of the speaker’s face is an additional input that provides information about the lip shape and movements, which is useful for speech acoustic information recovery in noise [5].

- Explicit modeling/constraining of the latent space structure can also make the model more efficient. For example, a subset of the latent space code can be constrained to model some of the speech characteristics whereas another subset can be constrained to model some other characteristics. Such ‘disantangling’ of the latent information has been shown to allow more compact and more efficient data representation [6].

For each of these possible extensions, the model training algorithm and the clean speech signal inference algorithm will be derived, implemented and tested.


**Required languages:**

**TYPE of CONTRACT** temporary, 24 months  
**JOB STATUS** (Full time or part time): full time  
**HOURS PER WEEK:** 35  
**CONTRACT STARTING DATE:** 1 January 2020  
**APPLICATION DEADLINE:** 1 November 2019

**ELIGIBILITY CRITERIA**

Applicants must hold a PhD degree (or be about to earn one) or have a University degree equivalent to a European PhD (8-year duration) in a discipline connected to signal and information processing, computer vision and machine learning. A particular interest/experience in speech/audio processing, visual recognition, and/or multimodal fusion is a plus. Strong motivation for the research work. Ability to work both independently and to collaborate within a small team. Computer skills: MATLAB, Python, Deep Learning Toolkits (e.g. Keras, Pytorch).

Applicants will have to send an application letter in English and attach:
- Their last diploma  
- Their CV  
- Letters of recommendation are welcome.

Address to send their application to: [email address]