Offer #2024-08369

Internship - Diffusion-based Unsupervised Joint Speech Enhancement, Dereverberation, and Separation

Contract type : Internship agreement

Level of qualifications required : Master's or equivalent

Fonction : Internship Research

Context

This master internship is part of the**REAVISE** project: "Robust and Efficient Deep Learning based Audiovisual Speech Enhancement" (2023-2026) funded by the French National Research Agency (ANR). The general objective of REAVISE is to develop a unified audio-visual speech enhancement (AVSE) framework that leverages recent methodological breakthroughs in statistical signal processing, machine learning, and deep neural networks in order to design a robust and efficient AVSE framework.

The intern will be supervised by Mostafa Sadeghi (researcher, Inria), Romain Serizel (associate professor, University of Lorraine), as members of the MULTISPEECH team, and Xavier Alameda-Pineda (Inria Grenoble), member of the RobotLearn team. The intern will benefit from the research environment, expertise, and powerful computational resources (GPUs & CPUs) of the team.

Assignment

Speech restoration regroups several downstream tasks that share a common goal of recovering a ground-truth speech signal that has been affected by one or many deformations. These deformations can be for example due to: a) a noise or a concurrent speech that adds up to the original speech signal, b) reflection of the speech signal by the walls in a room, c) limited dynamic range of a recording system that clips the speech waveform's amplitudes exceeding a certain threshold, d) packet loss occurring in transmission in telecommunication systems. Each of these degradations, has been most of the time studied separately in the literature leading to respective techniques like a) speech enhancement or speech separation b) dereverbation, c) declipping, and d) inpainting. Recently, the interest increased in learning universal models able to tackle simultaneously two or more tasks of speech restoration [1]. This is motivated by the fact that in real-life applications, a speech signal is likely tainted by several degradations at once. Various approaches have been proposed. They can be generative based [2, 3] or not [4], but they are mostly implemented in a supervised way leading to the requirement of pairs of training data, where each pair is composed of a degraded speech and the corresponding clean speech target. Better generalization for such a model is achievable at the cost of important training data. Particularly, speech denoising (or enhancement) is known to be vulnerable to mismatches since its standard approaches heavily rely on paired clean/noisy speech data to achieve strong performance.

Main activities

Recent advances in generative modeling have seen the emergence of diffusion models as strong data distribution learners. It consists of gradually turning clean data into noise and learning a deep neural network to reverse this process so that one can generate samples of the clean data distribution starting from a pure Gaussian noise for example. This generative modeling has been successfully applied in an unsupervised way individually for the task of speech enhancement [10], and speech dereverberation [5, 6]. That is, the training no longer requires paired data as opposed to the supervised; only the clean speech data is required in training, and the enhancement or dereverberation task is performed in inference with some statistical modeling. Promising results for generalization have been found for these approaches. Utilizing diffusion models [7] also attempts to solve individually in an unsupervised way, various speech restoration but with little success particularly for the speech separation task while in a supervised way and still with diffusion model [8, 9] achieve competitive performance.

The goal of this internship will be threefold:

- Address joint speech enhancement and dereverberation tasks rather than separately with diffusion models in an unsupervised way,
- Address speech separation, with a diffusion model learned in an unsupervised way (i.e. learned only on
  clean speech),
- Address the three tasks (enhancement, dereverberation, separation) with a single unsupervised framework.

Skills

Preferred qualifications for candidates include a strong foundation in statistical (speech) signal processing, and computer vision, as well as expertise in machine learning and proficiency with deep learning frameworks, particularly PyTorch.

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Remuneration

€ 4.35/hour

General Information

- **Theme/Domain :** Language, Speech and Audio
  Scientific computing (BAP E)
- **Town/city :** Villers lès Nancy
- **Inria Center :** Centre Inria de l'Université de Lorraine
- **Starting date :** 2025-04-01
- **Duration of contract :** 6 months
- **Deadline to apply :** 2025-01-15

Contacts

- **Inria Team :** MULTISPEECH
- **Recruiter :**
  Sadeghi Mostafa / mostafa.sadeghi@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

The keys to success

Prospective applicants are invited to submit their academic transcripts, a detailed curriculum vitae (CV), and, if they choose, a cover letter. The cover letter should highlight the reasons for their enthusiasm and interest in this specific project.

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Defence Security :
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.