Offer #2023-06842

Fast and Efficient Algorithms for Diffusion-based Unsupervised Speech Enhancement

**Contract type:** Fixed-term contract  
**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 enhancement is a fundamental signal processing technique aiming to improving the quality and intelligibility of spoken language in noisy environments. It involves the removal of unwanted background noise and the enhancement of the target speech signal, making it clearer and more understandable for listeners. Speech enhancement plays a crucial role in various applications, including telecommunications, hearing aids, and automatic speech recognition systems, ensuring effective communication and speech comprehension even in challenging acoustic conditions. Numerous speech enhancement algorithms have been developed so far, thanks to the remarkable performance and flexibility of deep neural networks (DNNs) [1]. These algorithms typically operate in a supervised manner, where a DNN is trained to map noisy speech inputs to their corresponding clean estimates. While this approach yields impressive results, it does raise a significant challenge related to generalization. That is, ensuring that the algorithm could enhance speech signals in real-world scenarios beyond the training data, as the diversity and complexity of acoustic environments can vary widely.

**Main activities**

Unsupervised speech enhancement, utilizing deep generative models, offers an alternative approach that exhibits enhanced generalization capabilities [2]. Unlike entirely supervised techniques, this generative-based, unsupervised framework involves learning the statistical distribution of clean speech signals, using e.g., variational autoencoder (VAE) [3], and subsequently utilizing it as a prior distribution for the estimation of the target signal from its noisy observation. As a more efficient alternative to VAE, a diffusion-based generative model has been recently utilised in [4], demonstrating improved performance over the VAE-based approach. Nevertheless, as with other diffusion-based method, it suffers from high computational complexity.

The primary goal of this project is to develop fast and effective inference algorithms for unsupervised speech enhancement using the diffusion-based approach outlined in [4]. In pursuit of this objective, we will leverage recent advancements in diffusion-based image generation [5] and efficient diffusion-based techniques related to inverse problems [6]. Additionally, we will develop a latent diffusion model (LDM) [7] for speech data, which operates in a significantly lower-dimensional space compared to the original feature space. This approach will help us address the high computational complexity associated with the diffusion-based unsupervised speech enhancement problem.

**References**

Skills

Preferred qualifications for candidates include a strong foundation in statistical (speech) signal processing, 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

Smic level

General Information

- **Theme/Domain**: Language, Speech and Audio Statistics (Big data) (BAP E)
- **Town/city**: Villers lès Nancy
- **Inria Center**: Centre Inria de l'Université de Lorraine
- **Starting date**: 2024-04-01
- **Duration of contract**: 5 months
- **Deadline to apply**: 2024-01-08

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.