Deep Learning-based Audio-visual Speech Enhancement

Mission confiée

Context: Audio-visual speech enhancement (AVSE) refers to the task of improving the intelligibility and quality of a noisy speech utilizing the complementary information of visual modality (lips movements of the speaker) [1]. Visual modality can help distinguish target speech from background sounds especially in highly noisy environments. Recently, and due to the great success and progress of deep neural network (DNN) architectures, AVSE has been extensively revisited. Existing DNN-based AVSE methods are categorized into supervised and unsupervised approaches. In the former category, a DNN is trained to map noisy speech and the associated video frames of the speaker into a clean estimate of the target speech. The unsupervised methods [2] follow a traditional maximum likelihood-based approach combined with the expressive power of DNNs. Specifically, the prior distribution of clean speech is learned using deep generative models such as variational autoencoders (VAEs) and combined with a likelihood function based on, e.g., non-negative matrix factorization (NMF), to estimate the clean speech in a probabilistic way. As there is no training on noisy speech, this approach is unsupervised.

Supervised methods require deep networks, with millions of parameters, as well as a large audio-visual dataset with diverse enough noise instances to be robust against acoustic noise. There is also no systematic way to achieve robustness to visual noise, e.g., head movements, face occlusions, changing illumination conditions, etc. Unsupervised methods, on the other hand, show a better generalization performance and can achieve robustness to visual noise thanks to their probabilistic nature [3]. Nevertheless, their test phase involves a computationally demanding iterative process, hindering their practical use.

Principales activités

Project description: In this PhD project, we are going to bridge the gap between supervised and unsupervised approaches, benefiting from both worlds. The central task of this project is to design and implement a unified AVSE framework having the following features: 1) Robustness to visual noise, 2) Good generalization to unseen noise environments, and 3) Computational efficiency at test time. To achieve the first objective, various techniques will be investigated, including probabilistic switching (gating) mechanisms [3], face frontalization [4], and data augmentation [5]. The main idea is to adaptively lower bound the performance by that of audio-only speech enhancement when the visual modality is not reliable. To accomplish the second objective, we will explore techniques such as acoustic scene classification combined with noise modeling inspired by unsupervised AVSE, in order to adaptively switch to different noise models during speech enhancement. Finally, concerning the third objective, lightweight inference methods, as well as efficient generative models, will be developed. We will work with the AVSpeech [6] and TCD-TIMIT [7] audio-visual speech corpora.

Informations générales

- **Thème/Domaine**: Langue, parole et audio
- **Ville**: Villers-lès-Nancy
- **Centre Inria**: CRI Nancy - Grand Est
- **Date de prise de fonction souhaitée**: 2021-10-01
- **Durée de contrat**: 3 ans
- **Date limite pour postuler**: 2021-04-25

Contacts

- **Equipe Inria**: MULTISPEECH
- **Directeur de thèse**: Sadeghi Mostafa / mostafa.sadeghi@inria.fr

A propos d’Inria

Inria est l’institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3500 scientifiques pour relever les défis du numérique, souvent à l’interface d’autres disciplines. L’institut fait appel à de nombreux talents dans plus d’une quarantaine de métiers différents. 900 personnels d’appui à la recherche et à l’innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 180 start-up. L’institut s’efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l’économie.

L’essentiel pour réussir

Application deadline: April 25th, 2021 (Midnight Paris time)

How to apply: Upload your file on jobs.inria.fr in a single pdf or zip file, and send it as well by email to mostafa.sadeghi@inria.fr & romain.serizel@loria.fr. Your file should contain the following documents:

- Your CV.
- A cover/motivation letter describing your interest in this topic.
- A short (max one page) description of your project (thesis, internship, or equivalent) of the work in progress if not yet completed.
- Your degree certificates and transcripts for Bachelor and Master (or the last 5 years).
- Master thesis (or equivalent) if it is already completed and publications if any (it is not expected that you have any). Only the web links to these documents are preferable, if possible.

In addition, one recommendation letter from the person who supervised(d) your Master thesis (or research project or internship) should be sent directly by him/her to mostafa.sadeghi@inria.fr & romain.serizel@loria.fr.

Applications are to be sent as soon as possible.

Consignes pour postuler

Sécurité défense:

Ce poste est susceptible d’être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L’autorisation d’accès à une zone est délivrée par le chef d’établissement, après avis ministériel favorable, tel que défini dans l’arrêté du 03 juillet 2012, relatif à la PPST Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l’annulation du recrutement.
**Avantages**

- 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

**Rémunération**


Monthly salary after taxes: around 1596,05€ for 1st and 2nd year. 1678,99€ for 3rd year. (medical insurance included).

---

**Politique de recrutement**

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.

**Attention**

Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.