

Internship PROPOSAL

Human-in-the-loop speech enhancement with generative models

Location: ADASP Group - Télécom Paris¹

Supervisors:

- Mathieu Fontaine (Télécom Paris)
- Marc Delcroix (NTT, France)

Contact: mathieu.fontaine@telecom-paris.fr

Date: April 1st (for 6 months)

Description

Speech enhancement consists of reducing acoustic interferences such as noise, reverberation, interfering speakers to make the speech captured by a microphone sound “cleaner”. This technology is a key to make smooth video conferencing or for audio post-processing to improve the sound quality in e.g. videos.

Human-in-the-loop speech enhancement, where the user provides feedback about the speech enhancement success/failure, could allow users to correct processing errors, allowing generating high-quality enhanced speech for audio post-processing applications [1].

The goal of this project is to develop such a speech enhancement system with user feedback, using a correction neural network model accepting the enhanced speech with a signal indicating the regions of the enhanced speech to correct.

We will follow the following plan:

- a. Reproduce prior works of [1] to get experience with human-in-the-loop speech enhancement.
- b. Extend prior work, by considering more powerful correction models, such as generative models
- c. Proposing fine-grained user feedback by differentiating the types of errors (speaker confusion, processing distortions, remaining noise, etc.), or other feedback such as manual transcriptions.

We aim to publish the results of the project to an international conference (ICASSP) or Journal (SPL).

Required profile and additional comment

- The intern candidate should have strong knowledge of deep neural networks, signal processing and Python

[1] Malek Itani et al. Neural Speech Extraction with Human Feedback. *INTERSPEECH*, 2025

¹ <https://adasp.telecom-paris.fr/>