Machine Learning for Audio Augmented Reality

Master Internship Offer

Dates: 5-6 months between February 1 and September 1, 2020

Internship advisors: Antoine Deleforge and Emmanuel Vincent (Inria Researchers)

Contact: antoine.deleforge@inria.fr; emmanuel.vincent@inria.fr

Location: Inria Nancy - Grand Est. 615 Rue du Jardin-Botanique, 54600 Villers-lès-Nancy, France. MULTISPEECH team: https://team.inria.fr/multispeech/.

Requirements: Bachelor degree in signal processing or machine learning. Good programming skills in Python and/or Matlab. Previous experience in audio or acoustics is a plus

Topic description:

Audio Augmented Reality (AAR) is receiving increasing interest in various application domains, such as artistic creation and cultural mediation, communication and information, and the entertainment industry. Embedding computer-generated or pre-recorded auditory content into a user's real acoustic environment creates an engaging and interactive experience that, for example, can be applied to video games, museum guides or radio plays. A key challenge is to spatialize sounds in 3D in the user's headset in a way that is consistent with the current environment of the user. For instance, a sound source placed 2 meters away in an office environment should not be rendered the same way as a source placed 10 meters away in a large museum hall. To achieve this, acoustical information about the user's environment is needed. A key task of the collaborative research project *HAIKUS* is to retrieve such information solely using microphones embedded in the user's headset.

In this context, this master internship will focus on developing machine learning methods that estimate the volume, reverberation time and/or absorption properties of a room based on multiple binaural recordings of a sound source from different user's positions. The problem of estimating the distance of the sound source will also be considered. The intern will perform the following tasks:

- Review of the existing recent literature on related topics, e.g. [1, 2]
- Creation of a simulated audio dataset corresponding to the considered audio augmented reality scenario, using available acoustic simulators such as [3]
- Development, training and testing of a deep neural network for the estimation of acoustical properties from these data, following previous works such as [4] and [5].

The intern will benefit from the strong expertise of the team MULTISPEECH in deep learning, acoustics and audio signal processing methodologies.

Bibliography:

[1] Genovese, A. F., Gamper, H., Pulkki, V., Raghuvanshi, N., & Tashev, I. J. (2019). Blind Room Volume Estimation from Single-channel Noisy Speech. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2019)*.

[2] Etienne Thuillier, Hannes Gamper and Ivan J. Tashev. "Spatial audio feature discovery with convolutional neural networks." In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2018)*.

[3] Steven M. Schimmel, Martin F. Muller, and Norbert Dillier. "A fast and accurate "shoebox" room acoustics simulator." In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009)*.

[4] Diego Di Carlo, Antoine Deleforge and Nancy Bertin. "MIRAGE: 2D sound source localization using microphone pair augmentation with echoes". In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2019)*.

[5] Saurabh Kataria, Clément Gaultier, and Antoine Deleforge. "Hearing in a shoe-box: binaural source position and wall absorption estimation using virtually supervised learning. In "*IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017)*.