MLSP 2021

IEEE International Workshop on Machine Learning for Signal Processing

October 25-28, 2021 – Gold Coast, Queensland, Australia

L3DAS21: Machine Learning for 3D Audio Signal Processing

IEEE MLSP Data Challenge 2021

Introduction

3D audio is gaining increasing interest in the machine learning community in recent years. The field of application is incredibly wide and ranges from virtual and real conferencing to game development, music production, autonomous driving, surveillance and many more. In this context, Ambisonics prevails among other 3D audio formats for its simplicity, effectiveness and flexibility. Ambisonic recordings permit to obtain an impressive performance in many machine learning-based tasks, usually bringing out a significant improvement over the mono and stereo formats. Tasks like Sound Source Localization, Speech and Emotion Recognition, Sound Source Separation Separation, Speech Enhancement and Denoising, Acoustic Echo Cancellation, among others, benefit from tridimensional representations of sound field, thus leading to higher accuracy and perceived quality.

Tasks

The L3DAS21 Challenge aims at encouraging and fostering research on machine learning for 3D audio signal processing. In multi-speaker scenarios, it is very important to properly understand the nature of a sound event and its position within the environment, what is the content of the sound signal and how to leverage it at best for a specific application (e.g., teleconferencing rather than assistive listening or entertainment, among others). To this end, L3DAS21 Challenge presents two tasks: 3D Speech Enhancement and 3D Sound Event Localization and Detection, both relying on first-order Ambisonics recordings in reverberant office environment.

Each task involves 2 separate tracks: 1-mic and 2-mic recordings, respectively containing sounds acquired by one Ambisonics microphone and by an array of two Ambisonics microphones. The use of two first-order Ambisonics microphones definitely represents one of the main novelties of the L3DAS21 Challenge. We expect higher accuracy/reconstruction quality when taking advantage of the dual spatial perspective of the two microphones.

Both tasks rely on the same audio recordings, but with completely different targets, as described below.

- Task 1: 3D Speech Enhancement

  The objective of this task is the enhancement of speech signals immersed in the spatial sound field of a reverberant office environment. Here the models are expected to extract the monophonic voice signal from the 3D mixture containing various background noises. The evaluation metric for this task is the sum of the short-time objective intelligibility (STOI) and word error rate (WER). The first estimates the intelligibility of the output speech signal. The latter, instead, is aimed at assessing the effects of the enhancement for speech recognition purposes.

- Task 2: 3D Sound Event Localization and Detection in Office Environment

  The aim of this task is to detect the temporal activities of a known set of sound event classes and, in particular, to further locate them in the space. Here the models must predict a list of the active sound events and their respective location at regular intervals of 100 milliseconds. We consider up to 3 simultaneously active sounds, which may also belong to the same class. The models' performance on this task are evaluated according to the location-sensitive detection error, which joins the localization and detection errors.

Dataset

The LEDAS21 datasets contain multiple-source and multiple-perspective B-format Ambisonics audio recordings. We sampled the acoustic field of a large office room, placing two first-order Ambisonics microphones in the center of the room and moving a speaker reproducing the analytic signal in 252 fixed spatial positions. Relying on the collected Ambisonics impulse responses (IRs), we augmented existing clean monophonic datasets to obtain synthetic tridimensional sound sources by convolving the original sounds with our IRs. We extracted speech signals from the Librispeech dataset and office-like background noises from the FSD50K dataset. We aimed at creating plausible and variegate 3D scenarios to reflect possible real-life situations in which sound and disparate types of background noises coexist in the same 3D reverberant environment. Further information on the dataset can be found on the L3DAS21 Challenge website.

Challenge Organization and Timeline

We first provide the training and development sets, alongside with a supporting Python-based API to facilitate the data download and pre-processing. We also supply baseline results for both tasks, obtained by using state-of-the art deep learning architectures. In a second step, we will release the test sets without truth labels. Participants must submit the results obtained for the latter. In the end, the final ranking of the challenge will be presented at the IEEE Workshop on MLSP 2021 and also released on the challenge webpage. Here is a timeline of the challenge.

- 27 Mar 2021 – Release of the training and development sets
- 08 Apr 2021 – Release of supporting code, baseline methods and documentation
- 10 May 2021 – Release of the evaluation test set
- 20 May 2021 – Deadline for submitting results for both tasks
- 27 May 2021 – Notification of the results of participants
- 31 May 2021 – Deadline for 6-page paper submission
- 31 Jul 2021 – Notification of paper acceptance
- 02 Aug 2021 – Notification of challenge winners
- 31 Aug 2021 – Deadline for camera-ready papers
- 25 Oct 2021 – Opening of the IEEE Workshop of MLSP 2021

Submissions

Results and papers can be submitted via the IEEE MLSP 2021 submission site.

Besides submitting papers related to L3DAS21 Challenge, authors are encouraged to submit to this special session also papers related to the topic of machine learning for 3D audio signal processing.

Challenge Website and Contacts

L3DAS21 Challenge Website: https://www.l3das.com/mlsp2021
Dataset Download: https://doi.org/10.5281/zenodo.4642005
Email contact: l3das@uniroma1.it

Organizers

Danilo Comminiello, Associate Professor, Sapienza University of Rome, Italy
Eric Guizzo, Research Fellow, Sapienza University of Rome, Italy