

## The ATCO2 project: Creating a commonly available data base for training ATC ASR models

*Petr Motlicek, Idiap Research Institute*

The talk will introduce the ongoing project ATCO2 (Automatic collection and processing of voice data from air-traffic communications, 2019 - 2022), funded by CleanSky EU programme (<https://www.cleansky.eu>). Goal of the project is to develop a platform allowing to collect, organize and pre-process air-traffic control (voice communication) data from air space users. Several hundreds of hours of communication data have already been collected by a community of people and are either manually or automatically transcribed. Transcriptions are enriched by the corresponding ADS-B data. The data is already partially used in subsequent research as well as starts to be offered through the Open Sky Network (<https://opensky-network.org>) to potential developers of ASR applications for ATM.