2018-00750 - Audio-visual conversational sorting with autonomous systems

Level of qualifications required : Graduate degree or equivalent
Function : PhD Position
Level of experience : Recently graduated

About the research centre or Inria department
Grenoble Rhône-Alpes Research Center groups together a few less than 800 people in 35 research teams and 9 research support departments.

Staff is localized on 5 campuses in Grenoble and Lyon, in close collaboration with labs, research and higher education institutions in Grenoble and Lyon, but also with the economic players in these areas.

Present in the fields of software, high-performance computing, Internet of things, image and data, but also simulation in oceanography and biology, it participates at the best level of international scientific achievements and collaborations in both Europe and the rest of the world.

Context
The Ph.D. candidate will be hosted in the Perception Team, at Inria Grenoble Rhône-Alpes. Our team has extense technical expertise on audio-visual data processing and learning (specifically with egocentric centers, e.g. cameras and microphones mounted on a robotic platform or embedded in a smart device). Similarly, the team has the necessary technology to acquire, process and do learning on large audio-visual datasets.

This Ph.D. will be co-supervised by Dr. Radu Horaud (head of the Perception Team) and Dr. Xavier Alameda-Pineda.

Assignment
Scene understanding with egocentric (as opposed to distributed) sensors is an active field of research. Both, the computer vision and audio processing communities proposed method to address important tasks such as people detection and tracking, gaze following, object detection, speaker diarization, speech enhancement and separation, etc. All these tasks are key for conversational sorting, that is to understand which part of the speech was uttered by whom. To do that, the aforementioned tasks must be jointly solved. We propose to exploit the information contained in the auditory and visual modalities as well as the complementarity between the two.

Indeed, audio-visual processing and learning offers the possibility of overcoming the limitations associated to the exploitation of a single modality. However, it comes at the price of designing methods and algorithms that properly link and fuse auditory and visual data. In this Ph.D., we aim to develop machine (probabilistic and deep) learning methods to sort the conversational situation and assign different speech pieces to their utterer. The challenge is further increased by the usage of autonomous systems (i.e. companion robot or smart device).

Main activities
The main activities are: scanning the literature to properly understand where is the bottleneck, propose ideas how to overcome the current limitations, run experimental protocols to validate/reject the research hypothesis and finally to write research reports/scientific articles and publish them.

Skills
Technical skills and level required : strong mathematical and programing background, basic notions of probabilistic and deep learning, computer vision and signal processing.

Languages : Outstanding oral and writing skills in English.

Benefits package
- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Sports facilities

Remuneration

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

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.