

Master's thesis proposal 2021

Multi-sensor target localization and tracking for sound enhancement

Traditional hearing-aids only have a few microphones as sensors to produce improved sound for the hearing impaired. With the use of more microphones, other complementing sensors, and tailored algorithms a significant performance increase of target speech amplification and noise suppression is expected. Current trends in audio-visual speech enhancement such as [1] and [2] are very promising and inspiring.

In this master's project the student(s) will integrate several data sources to a common sensor platform consisting of a pair of Tobii Pro 2 eye tracking glasses (ETG) including a scene facing camera, an inertial measurement unit (IMU), and a single microphone. From earlier work in [3], and a student project, several software modules were implemented allowing for real-time tracking and mapping of faces, head pose estimation, global gaze filtering, a simulation environment, and more.

The student(s) will derive models for all sensor measurements and their relations. Several estimators can then be derived using the sensor models. Depending on the student's interest and the progress of the project several studies can be done.

Suggested investigations can include:

- Face pose tracking from video
- Facial features and voice activity tracking
- Sound source mapping and tracking
- Joint ego-localization and mapping (SLAM) from video/audio and IMU
- Source separation and clean speech reconstruction from video and single microphone [1-2]
- Eye gaze behavior analysis in various scenarios

The Qualisys motion tracker in Visionen Arena will be used for evaluation of algorithms and to support estimators.

Supervisors: Martin Skoglund (Oticon/LiU)

Location: LiU, Reglerteknik.

Start: January 2020.

Duration: 20 weeks, 30 HP.

References:

[1] Looking to listen at the cocktail party: a speaker-independent audio-visual model for speech separation. Ariel Ephrat, Inbar Mosseri, Oran Lang, Tali Dekel, Kevin Wilson, Avinatan Hassidim, William T. Freeman, Michael Rubinstein. *In ACM Trans. Graph.* 2018.

[2] Effects of Lombard Reflex on the Performance of Deep-Learning-Based Audio-Visual Speech Enhancement Systems. Daniel Michelsanti, Zheng-Hua Tan, Sigurdur Sigurdsson, Jesper Jensen. *In Proc. Int. Conf. Acoust., Speech, Signal Processing*, 2019.

[3] Mapping an Auditory Scene Using Eye Tracking Glasses. Alfred Fredriksson, Joakim Wallin. *Master's Thesis, Linköping University*. 2020.