# Master's thesis proposal 2020

## 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 and an accelerometer triad; and a multi-microphone platform with a 16 channel soundcard and processor, Spresense, developed by Sony.

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:

- Person tracking from video, moving camera/targets
- Ego-localization from video
- Sound source mapping and tracking
- Beamforming and beam selection using microphones and eye-gaze from (ETG)
- Joint ego-localization and mapping (SLAM) from video/audio and accelerometer
- Source separation from video and multiple microphones
- Distance from audio arrays. Spectral aspects (reverberation)

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

Supervisors: Martin Skoglund (Oticon/LiU), Clas Veibäck (LiU), Martha Shiell (Oticon), Sergi Rotger Griful (Oticon)

Location: LiU, Reglerteknik. At least one trip to Denmark.

Start: January 2020.

Duration: 20 weeks, 30 HP.

Salary: Yes.

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.*

Contact: Martin Skoglund, martin.skoglund@liu.se, Reglerteknik ingång 27, Rum 2A:488