Popular science summary

Sara Enander & Louise Karsten

Intelligent hearing aids hope to amplify only the sound a person is attending to. This thesis uses machine learning to see if there is a connection between brain signals, eye gaze and the attended source.

Imagine being in a crowded space, surrounded by lots of talking people. Standing in the same location, people with normal hearing can easily shift focus from one talking person to another. Somehow their brains manage to sort out the sound they want to listen to. The ability of the brain to perform this selection has been investigated but it remains unclear how it succeeds to do it. If a person has impaired hearing the ability of sorting out unwanted sounds is affected. This leads to hearing problems in noisy situations even though the person is using a hearing aid. Research for developing intelligent hearing aids has not yet come up with a solution for solving this in a good way and more research is needed.

One step towards intelligent hearing aids is to find a way to understand what a person is attending to. This thesis uses data from brain signals (EEG) and eye gaze to see if a machine learning model can predict whether a person is attending to a monologue or a dialogue. It is investigated what model is most successful with this prediction. The data comes from experiments where the test subjects are exposed to a monologue and a dialogue at the same time but are told to only focus on one of them. For the application of integrating this into a hearing aid device it is also interesting to investigate how few electrodes of the EEG that can be used, and also how short the trials that are being presented to the model can be. Another thing being investigated is what parts of the brain that are most important when predicting what a person is attending to. Three different models are investigated for this task: support vector machine, multilayer perceptron and convolutional neural network.

A convolutional neural network performs best overall and reaches an average prediction score of 87% for all subjects when using inputs from all electrodes at the same time. When using one electrode at the time, it performs worse than the support vector machine and multilayer perceptron that reaches average scores above 70%. There is however a clear pattern in what regions of electrodes that succeed best with the classification task. These are the electrodes at the temporal lobe as well as the sides of the front of the frontal lobe. It varies how short the trials can be before the accuracy decreases for each model when EEG data is used. The support vector machine and the multilayer perceptron performs best for longer trials while the convolutional neural network performs best for shorter trials. For the eye gaze data, the accuracy is not affected remarkably by decreasing the length of trials.