

# Decoding Auditory Attention with Automatic Speech Recognition

Using Whisper to improve the performance of smart hearing aids

*Advanced machine learning applied to the decoding of brain signals will bring us closer to intelligent, neuro-steered devices for hearing loss patients.*

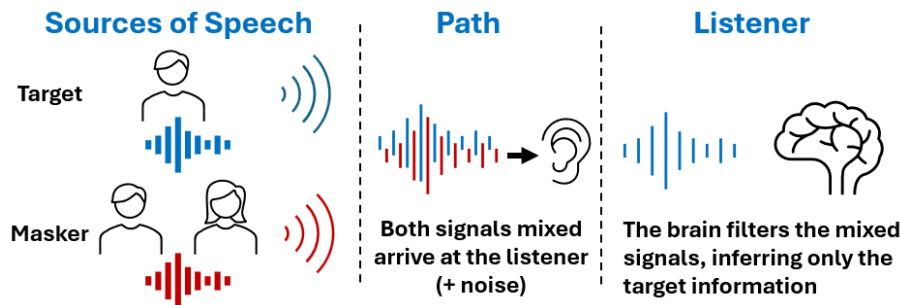
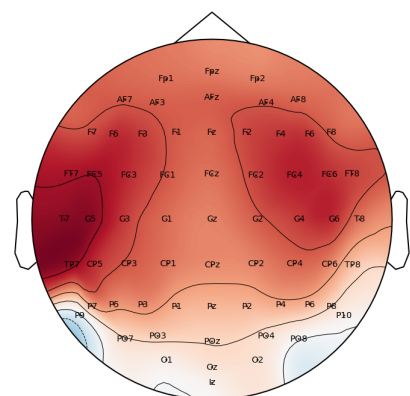


Illustration of the *Cocktail party problem*.

Our lives are very noisy: whether we are at the mall buying groceries or enjoying fika at the cafe, most of the time we are surrounded by noise and chatter. Yet, even with all this disturbance, our brain is incredibly good at focusing on a particular speaker as soon as they capture our attention. This ability is unfortunately severely degraded in hearing loss patients, who have a hard time understanding others in this scenario, known as the *cocktail party problem*. In an effort to solve these issues, we have tried to apply Whisper, an advanced speech-to-text algorithm, to the difficult task of predicting the attention of a test subject between two competing sources of speech, given real-time *electroencephalography* data recorded through an electrode cap.

We trained a machine learning model to extract relevant information from the internal states of Whisper and compare it to the recorded brain responses to the same stimuli, confirming that Whisper's processing of speech follows similar patterns to actual brain processes, and can be used to explain the EEG data of a patient listening to the same audio much more accurately when the patient is paying attention to it. Finally, we applied this model to an *attention classification* task, where the similarity between the recorded brain signals and our predictions for each of the competing speech segments was fed to a final model to be classified into *attended* and *ignored* speech. The resulting performance is promising, ranging from around 80% to 98% accuracy when basing the decision on 5 or 30 seconds of data respectively, a result significantly better than what we were able to achieve using simpler inputs, such as the audio itself.



Whisper's predictive power (red areas are better predicted).

While the road ahead is still long, we envision our work to be one of several stepping stones towards a future of intelligent hearing aids capable to decode attention directly from the brain and using it to greatly enhance the lives of hearing loss patients, who have to contend daily with the *cocktail party problem*.