

---

# Optimal latencies for linear cortical auditory attention detection: differences between speech and music listening

Adèle Simon<sup>\*†1,2</sup>, Søren Bech<sup>1,2</sup>, Jan Østergaard<sup>1</sup>, and Gérard Loquet<sup>1</sup>

<sup>1</sup>Aalborg University [Denmark] – Denmark

<sup>2</sup>Bang Olufsen – Denmark

## Abstract

In this study, we explore the differences between linear cortical auditory attention detection when the listener is attentively listening to speech or music. This study focuses on the latencies that can produce a maximal reconstruction of the target signal, thus maximal attention detection success. Differences could suggest different brain processes, specific to music listening or speech listening in presence of distracting sounds.

In recent decades, there has been a lot of interest in detecting auditory attention from brain signals. Cortical recordings have been demonstrated to be useful in determining which speaker a person is listening to a mixed variety of sounds (the cocktail party effect). Linear regression, often called the stimulus reconstruction method, shows that the envelope of the sounds heard can be reconstructed from continuous electroencephalogram (EEG). The target sound, to which the listener is paying attention, can be reconstructed to a greater extent compared to other sounds present in the sound scene, which can allow attention decoding. Reconstruction can be obtained with EEG signal that are delayed compared to the audio signal, to take into account the time for neural processing. It can be used to identify latencies where the reconstruction is optimal, which reflect cortical process specific to the type of audio heard. With this method, several previous studies highlighted an optimal reconstruction around 200 milliseconds following the stimulus, where an attentional process can also be observed, suggesting a neural process involved in cocktail party resolution at this latency. However, most of these studies used only speech signals and did not investigate other types of auditory stimuli, such as music.

From existing literature and preliminary results, we hypothesized that a maximal correlation, influenced by the direction of the attention of the listener, will be visible at middle or late latencies, for both speech and music listening. However, differences are expected to be seen between conditions where the target of attention was music compared to conditions where the target is speech: in the reconstruction accuracy, the timing of the latencies, and weighting of the models.

In the present study, we applied this stimulus reconstruction method to decode auditory attention in a cocktail party scenario that includes both speech and music. Participants were presented with a target sound (either speech or music) and a distracter sound (either

---

\*Speaker

†Corresponding author: amds@es.aau.dk

speech or music) while continuously recording their cortical response during the listening with 64-channels EEG system. From these recordings, we reconstructed the envelope of the stimuli, both target and distracter, by using linear ridge regression decoding models at individual latencies. Multiple information can be extracted from this analysis: identify the latency where the reconstruction accuracy is maximal; compare target-trained and distracter-trained models to identify latencies affected by auditory attention; exploration of the weight of each decoder to identify the brain region involved in the stimulus reconstruction. The analysis is still ongoing, and the results will be presented in the paper.