EEG-based Decoding of Auditory Attention using a Deep Attention Network: Revealing neural commonalities of selective attention across individuals

Gabriel Ivucic[1,*], Saurav Pahuja[1,2], Siqi Cai[3], Haizhou Li[2,3], Tanja Schultz[1]

1) Cognitive Systems Lab, University of Bremen, Bremen, Germany
2) Machine Listening Lab, University of Bremen, Bremen, Germany
3) Department of Electrical Engineering, National University of Singapore, Singapore

*CSL, University of Bremen, Enrique-Schmidt-Str. 5, Bremen, Bremen, Germany. Email:ivucic@uni-bremen.de

Introduction: Focusing on specific sound sources in cluttered environments is crucial for daily communication. However, this ability poses a great challenge for persons that are dependent on hearing aids, as the devices do not possess information about which speech sources are interesting to the user. To solve this problem, approaches from the field of auditory attention detection (AAD) are trying to develop cognitive models of auditory selective attention using electroencephalography (EEG). Here, subject-independence (SI) is useful for EEG-applications in AAD because it eliminates the need for pretraining on specific individuals, making the model more flexible and adaptable to a wider range of users. This allows the model to be applied to new persons without the need for additional data collection and training, making it more efficient for practical applications. Such models could expand our understanding of the cognitive processes involved in selective attention. Further, an integration into hearing aids in future applications would allow individuals with hearing impairments to regain a level of normalcy in their daily activities.

Materials, Methods and Results: This study aims to investigate subject-independent auditory attention decoding using EEG and Deep Neural Networks (DNN). The EEG data set in this work is publicly available and widely used in the AAD community [4.]. Participants were presented with two simultaneous but spatially separated speech stimuli, with the instruction to focus on one of the speech streams while their 64-channel EEG signals were recorded. The decoding task is a binary classification of the attended speaker in a given time window. To achieve this, the data was preprocessed and analyzed using a Deep Attention Network [1.], which is designed to be a lightweight and efficient architecture to process raw windows of EEG signals. The network uses spatial and temporal attention modules to extract EEG-channel interactions and temporal dynamics at different frequencies. The EEG-data was lightly processed by common-average referencing and filtering the signal between 1-32 Hz, followed by segmenting in 1 second non-overlapping windows for each of the 16 participants. The network was trained to classify the attention states of the participants based on the EEG data in a leave-one-subject-out cross-validation. The results show an accuracy of 72% (STD: 11%) over all 16 participants with all but 1 participant significantly outperforming the baseline of 50%. Excluding the 6 subjects below 70% as a threshold of practical performance, the remaining 10 subjects average an 80% accuracy (STD: 6%). The extraction of the spatial maps of the network allows an insight into the importance of each channel for the classification model. The averaged electrode weights for participants reveal strongly localized activations in the prefrontal and temporal lobes (AF7, AFz, AF8, T7, T8) with an average standard deviation of 5% of the mean between the participants, for all channels.

Discussion: The attention network significantly outperforms former DNN approaches for subject-independent auditory attention ($p < .01$) [3.] by an absolute of 7% (accounting for all participants) and has a lower variance between participants. While the spatial weights only reflect a part of the DNN-models, they imply a shared neural processing between the individuals in the prefrontal and temporal lobes. These areas are known to play a crucial role in speech tracking during selective listening [2.], and are likewise used by the neural network to discriminate between the different speech streams.

Significance: The attention network reaches state-of-the-art performance for subject-independent auditory attention decoding with lower variability and less than 4000 parameters while allowing an intuitive visualization and interpretation of modules of the model. The ability to decode auditory attention in a subject-independent manner is crucial for the development of cognitive models that can be applied to a wide range of individuals, including those with hearing impairments.

References:

[1.] E. Su, S. Cai, L. Xie, H. Li and T. Schultz, "STAnet: A Spatiotemporal Attention Network for Decoding Auditory Spatial Attention From EEG," in IEEE Transactions on Biomedical Engineering, vol. 69, no. 7, pp. 2233-2242, July 2022, doi: 10.1109/TBME.2022.3140246.

[2.] Puschmann S, Steinkamp S, Gillich I, Mirkovic B, Debener S, Thiel CM. The Right Temporoparietal Junction Supports Speech Tracking During Selective Listening: Evidence from Concurrent EEG-fMRI. J Neurosci. 2017 Nov 22;37(47):11505-11516. doi: 10.1523/JNEUROSCI.100717.2017. Epub 2017 Oct 23. PMID: 29061698; PMCID: PMC6596752.

[3.] Servaas Vandecappelle, Lucas Deckers, Neetha Das, Amir Hossein Ansari, Alexander Bertrand, Tom Francart (2021) EEG-based detection of the locus of auditory attention with convolutional neural networks eLife 10:e56481 https://doi.org/10.7554/eLife.56481

[4.] Das, Neetha, Francart, Tom, Bertrand, Alexander. (2020). Auditory Attention Detection Dataset KULeuven (1.1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.3997352