Online detection of auditory attention in a neurofeedback application

Rob Zink[1,2], Annelies Baptist[1,2], Alexander Bertrand[1,2], Sabine Van Huffel[1,2], Maarten De Vos[3]

[1]KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Kasteelpark Arenberg 10, 3001 Leuven, Belgium.
[2]iMinds Medical IT, Leuven, Belgium.
[3]Engineering Department, Oxford University, Oxford, United Kingdom.

Abstract

Auditory attention detection (AAD) holds promising potential for usage in auditory-assistive devices. Being able to train subjects in achieving high AAD performance would increase the application potential of AAD. This requires an acceptable temporal resolution and the analysis should take place online. In the current study, we implemented a fully automated closed-loop system that allows for convenient recording outside a lab environment. We achieved high AAD accuracies with a trial length of 10 seconds and provided subjects with visual feedback on their ongoing performance. This exploratory study proves the feasibility of investigating the effect of neurofeedback in such a setting, paving the way for future studies.

Keywords *Auditory Attention Detection, Mobile EEG, Neurofeedback*

1 Introduction

In the past few years, electroencephalography (EEG) has been employed for auditory attention detection (AAD). A pretrained decoder that makes a linear combination of the EEG and delayed versions of it allows to extract a signal with a significantly higher correlation to the attended speaker's envelope as compared to the unattended speaker's envelope [1]. In dual-speaker scenarios, the subject's attention to a specific speaker could be reliably detected with high accuracy in research labs [2, 3, 4]. In addition, several analysis and preprocessing steps have been evaluated to optimize the estimation made by the decoder [5]. The accuracy was found to be dependent on the considered trial length for calculating the decoder and correlations [1, 4]. The temporal resolution of the attention tracking is dependent on the trial length and can therefore be seen as a trade-off between accuracy and temporal resolution. For applications in real-time, shorter trial lengths such as 10s are preferred over the most reported 60s trial length, for example. Online analysis of the attended audio source holds potential for use in assistive audiological devices, such as hearing aids [3, 4] (e.g. to steer a beamformer to the attended speaker).

The effectivity of EEG-based AAD depends, not only on the temporal resolution and differences in speech (i.e. bilabial sounds), but also on the responses of the test subjects themselves. Large differences in accuracy between subjects were observed in the aforementioned studies. It is known for other cognitive paradigms (e.g. such as P300 oddball studies) that subjects' physiological responses can differ substantially, even within subjects when moving from restricted to more real-life scenarios [6]. Providing feedback about the ongoing EEG signals was shown to be beneficial for other EEG paradigms to strengthen the brain responses (e.g., motor imagery [7]). A similar reasoning can be applied for AAD, i.e. training users to elicit stronger brain responses related to the attended speech stream might increase the accuracy. To this end users might experience positive effects from such a neurofeedback training to increase the performance of EEG-based AAD.

In the current study, we explore the application of the AAD in real time. The subjects were recorded in an office environment with mobile EEG hardware and consumer-grade headphones. This is more convenient for the subjects compared to a lab environment. In addition, we apply a neurofeedback scenario to provide a proof-of-concept for a fully automated closed loop system. We implemented an online AAD analysis with a time resolution of 10 seconds. Half of the subjects received visual feedback about their AAD accuracy per time point. We show competitive results compared to existing studies with offline analyses which paves the way to investigate the effect of long term neurofeedback in future studies (e.g. at subjects' home).

2 Methods

2.1 Participants

Twelve native Dutch-speaking subjects (mean age (SD) 22.4 ($\pm$2.1) years, six women) participated in the current experiment. Subjects reported normal hearing and no past or present neurological or psychiatric conditions. All participants signed informed consent forms prior to participation. The ethics committee of the KU Leuven approved the experimental setup.

2.2 Data Acquisition

The acquisition was conducted with a SMARTING mobile EEG amplifier from mBrainTrain (Belgrade, Ser-

bia, www.mbraintrain.com). This amplifier comprises a wireless EEG system running on a notebook computer using a small 24-channel amplifier with similar characteristics to a stationary laboratory amplifier (<1uV peak to peak noise; 500Hz sampling rate). The EEG was measured using 24 Ag/AgCl passive scalp electrodes (Easycap), placed according to the 10-20 standard system with positions: FP1, FP2, Fz, F7, F8, FC1, FC2, Cz, C3, C4, T7, T8, CPz, CP1, CP2, CP5, CP6, TP9, TP10, Pz, P3, P4, O1 and O2. Impedances were kept below 10 kOhm and an abrasive electrolyte gel was applied to each electrode. EEG data were recorded through Openvibe (and stored for offline (reference) analysis) and streamed from Openvibe to Matlab via the labstreaminglayer interface (LSL). The audio stories were played via Openvibe and pre-loaded into Matlab for the online analysis part and synced via the Openvibe audio triggers. Every ten seconds, the data was retrieved from the LSL stream and analyzed in Matlab. Both online and offline analysis used custom-made Matlab scripts.

2.3 Stimuli and Procedure

Audio stimuli consisted of four stories in Dutch (of approximately 13 minutes length), narrated by four different male speakers. Silences were truncated to 500ms and each story was divided into two parts, resulting in 8 sessions of $\pm$6.5 minutes in length. Subjects listened to two stories presented simultaneously through low-cost consumer headphones (Sennheizer mx475). The two audio streams were filtered by head-related transfer functions leading to more realistic perception [2]. Subjects were asked to pay attention to only one story on the left or right side. Afterwards, multiple-choice questions were presented and subjects indicated the difficulty in listening (i.e., indicated on a ten-point scale) and answering some questions about the story. After listening to one full story (2 x 6.5 minutes), subjects switched attention to the opposite side (i.e. left or right speaker). The experimental setup consisted of two blocks of two stories each. The first half of the recordings (24 minutes) were used for estimating the decoder, the latter for evaluation. Half of the subjects received visual feedback (feedback group) on the laptop screen for the second part of the experiment, whereas the other subjects received no feedback (control group). Order effects were avoided by alternating the listening sides among subjects.

2.4 Preprocessing and Analysis

EEG data were bandpass-filtered at 1-8 Hz and consequently down-sampled to 20Hz for each 10s segment of data. The absolute value of the audio waveforms with power-law compression with exponential 0.6 was taken to obtain the audio envelopes, and then an 8Hz low-pass filter was applied [5]. Envelopes were extracted from the clean audio signals for the separate speakers. Offline analysis was done on trial lengths up to 60s, in steps of 10s. Real-time analysis was done on trial lengths of 10s.

The decoder-construction approach followed similar steps as presented in previous publications [1,3]. In short, seven time-shifted versions of the EEG trial were obtained in a 0-300ms range after stimulus onset. All EEG channels and their 7 delayed versions are then linearly combined using a pre-trained linear decoder w. During training, the decoder is optimized such that the resulting output signal has a minimal mean squared error (MMSE) with the attended speech envelope. Thie linear MMSE decoder w can be computed as $w = R^{-1}c$, where R is the covariance matrix over all the EEG channels, and c the cross-correlation vector between the EEG channels and the attended speech envelope [2]. All the Rs and cs computed over the training trials were averaged to create a single average covariance matrix and cross-correlation vector [5]. For each test trial Pearson's correlation coefficient is employed to quantify the decoders' reconstructed envelope to the attended stimulus (CA) and unattended stimulus (CUA). The highest correlation value determines to which of the two speakers the subject was listening at the current trial. The decoders were trained following an (offline) leave-one-trial-out structure on all data. The decoders for the second half of the experiment were computed solely on the training data of the first half of the experiment.

The subjects who received feedback were presented with a colored circle in the center of the screen. After every test trial of ten seconds, the colored circle indicated the performance of the past ten seconds. Four different colors were used for performance indication: Dark red, light red, light green and dark green. Thresholds determining the colors were based on the training set in such a way that the CA and CUA difference would be equally divided for the correct and incorrect trials. Note that trials with a larger CA as compared to CUA (i.e., correct trials) are always green and the incorrect trials are always red. Decoders for analyzing the second half of the dataset were based on the individual training data.

3 Results

3.1 All sessions

Grand average classification accuracy for the 10s window was 81.9% (SD =5.9%). Increasing the window length in the offline analyses raised the accuracy up to 96.9% (SD=3.5%). Figure 1 displays single-subject and grand-average accuracies at different trial lengths. All subjects scored above chance level (>55% at 10s up to >62% at 60s) for all window lengths.

Stable performance was achieved up to removal of 16 channels that contributed the least in the envelope-reconstruction model performance. With a reduced set of 8 channels, an accuracy of 78.8% (SD = 6.5%) was achieved. If we maintained only seven or fewer electrodes, the performance dropped significantly. This is depicted in Figure 2. To evaluate the most discriminative and most redundant channels, we plotted the average number of removal per channel, down to 8 channels. This is illustrated by the left topoplot in Figure 2. The red/yellow colors depict the channels that were re-

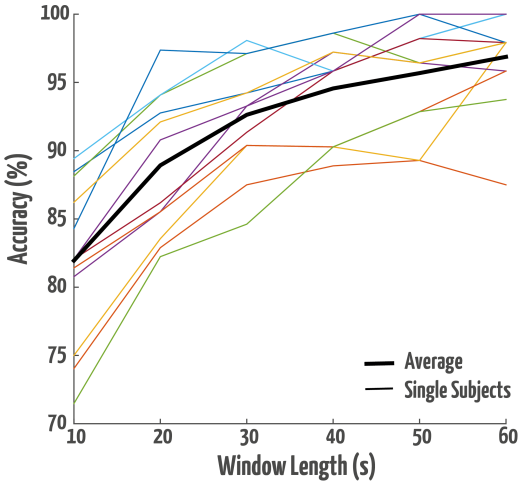


Figure 1: Grand average and subject specific decoding accuracies for different window lengths.

moved without having a large influence on general accuracy. It can be noted that especially the frontal and posterior channels contributed the least to the decoders performance. In contrast, the temporal electrodes (around the ear) are most important, as is depicted by the yellow colors in the right topoplot.

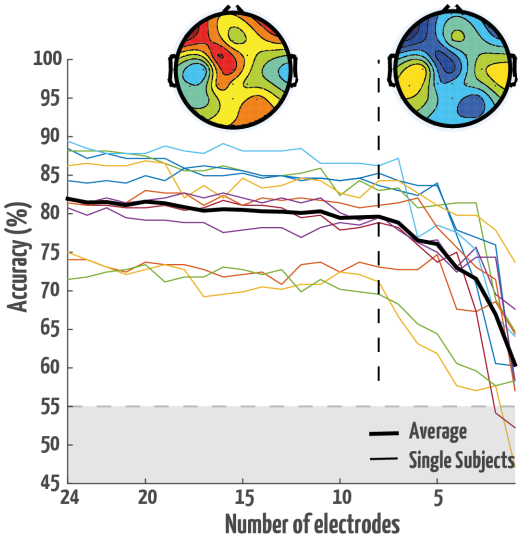


Figure 2: Grand average and subject specific decoding accuracies depending on the number of electrodes for the 10s window length. The channel with the lowest decoder weight (after correction for channel variance differences) is removed in each iteration. Topoplots represent the average distribution for the least discriminative channels on the left and most discriminative on the right. The shaded area indicates the chance level.

3.2 Neurofeedback

Average online classification accuracy in the online decoding of the second half of the data was 79.7% (SD = 7.0%) using the pre-trained decoder. When evaluating the offline leave-one-out decoder, average accuracy was 83.0% (SD = 7.5%). This increase in accuracy was significant (t11=-4.1 $p<0.01$), indicating that additional training data resulted in an increase at the 10s windows. In contrast, this difference was not significant for 60s trial length; both types of decoders achieved equal accuracies: 96.2% and 96.9% for the pre-trained and leave-one-out decoder, respectively.

Comparing the accuracy within the feedback group between the feedback session and the training session revealed a slightly higher accuracy in the feedback session, +4,1 percentage points as calculated with the leave-one-out decoder when averaged over all subjects in the feedback group. Five out of six subjects scored higher in the feedback session, compared to the training session. For the group not receiving feedback, the difference between the second session and the training session was -0.2 percentage points. No effect was found for increased or decreased CA or CUA changes in the feedback group with respect to the training session.

Relative occurrence of the four feedback cues was 41.8% dark green, 40.7% light green, 10.1% light red and 8.7% dark red. We evaluated the temporal patterns at which the cues were evident between the feedback and the hypothetical feedback cues on the training set (i.e., calculated offline with similar thresholds as the feedback session; subjects did not see this feedback.). Entries in Figure 3 indicate, by the colors the relative frequency at which a colorcue in the rows was followed by the colorcues in the columns. It can be noted that, in the feedback session, participants shifted more frequently to dark red after seeing light red. For the training sessions, subjects seem to have less often two consecutive light red trials but more often from light green to light red. Note that the difference in threshold between the light colors is smaller, compared to the dark colors. For example, light green denotes a correct decoding, but with less conficence than in the case of dark green.

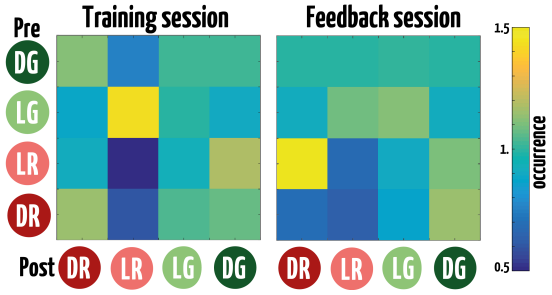


Figure 3: Adjacency matrices for the Training and Neurofeedback session. The entry color indicates the normalized frequencies of a specific color cue (in the rows) that is followed by any other cue (columns). Cues: DG = Dark Green, LG = Light Green, LR = Light Red and DR = Dark Red.

3.3 User Metrics

On average, subjects answered 84.2% (SD = 7.9%) of the questions correctly. We contrasted the number of correct responses of each subject to the general accuracy at the 10s window analysis. This revealed a strong positive correlation ($r = 0.72$, $p < 0.05$). One subject was removed, as its number of correct responses differed more than 2 standard deviations from the mean. Figure 4A displays the individual subjects' accuracy and number of correct responses. A regression line has been added for illustrational purposes. A moderate negative correlation (r

= -0.56, p = 0.059) was found between the average accuracy and the subjects' reported difficulty in answering the questions and overall listening. This correlation is depicted in Figure 4B.

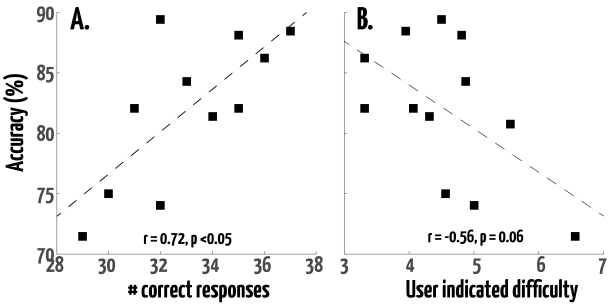


Figure 4: Scatterplots illustrating the correlation between the decoder grand-average accuracy and the number of correct responses after each story in A and the User indicated task difficulty in B. A regression line has been added for illustrational purposes.

4 Conclusion and Discussion

In the current study, we evaluated the possibility of implementing an online closed-loop system for auditory attention detection. With window lengths of 10s, we obtained robust accuracies that, even with a low number of electrodes, are predictive. We provided a fully working feedback system, and the online implementation did not significantly degrade the results. Although a slight improvement in accuracy was observed when using neurofeedback, the effects are not attributed much significance due to the limited number of measurements. Overall the results were similar and competitive to existing lab-studies. This is particularly interesting for future application in long term studies in real-life conditions such as the subjects' home.

The high accuracies in the present work were obtained with 24 electrodes and were found to be stable up to removal of 16 channels. These results are in line with insights presented by [4] and are encouraging for future work in online processing. The neurofeedback results show no clear negative deflection in accuracy due to increased distraction. Nevertheless, subjects who saw light-red feedback were more prone to perform bad in the next trial as well. One explanation may be that when subjects became aware of an error (i.e. shift from green trial to light-red) this leads to a brief surprise effect which lowers the attentional response in the next 10 seconds. In general we conclude that future studies with an increased number of subjects and longitudinal measurements might demonstrate positive effects on the AAD that would be highly valuable for future users of auditory-assistive devices.

A limitation in the current study is the lack of incorporating real-life audio signals. Recently there have been studies evaluating the effect of noisy reference signals and showing a negative impact on performance [3, 8] In addition, real-life scenarios involve frequent switching of attention. This factor is not well reflected in the current paradigm; subjects only switched attention after each story. To this end, it would be interesting to see how a state-space model would perform, as this was shown to have a high temporal resolution [9].

Acknowledgements

Research supported by Research Council KUL: CoE PFV/10/002 (OPTEC); Belgian Federal Science Policy Office: IUAP P7/19/(DYSCO, 20122017), iMinds Medical Information Technologies SBO 2015; Flemish Government, FWO projects: G.0427.10N; EU: (FP7/20072013)/ERC Advanced Grant: BIOTENSORS (nr. 339804). This paper reflects only the authorsFL views, and the Union is not liable for any use that may be made of the contained information.

References

[1] J. A. O'Sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma, and E. C. Lalor. Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral Cortex*, 25(7):1697–1706, 2015.

[2] N. Das, W. Biesmans, A. Bertrand, and T. Francart. The effect of head-related filtering and ear-specific decoding bias on auditory attention detection. *Internal Report, http://homes.esat.kuleuven.be/ abertran/publications.html*.

[3] S. Van Eyndhoven, T. Francart, and A. Bertrand. EEG-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses. *arXiv preprint arXiv:1602.05702*, 2016.

[4] B. Mirkovic, S. Debener, M. Jaeger, and M. De Vos. Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications. *Journal of neural engineering*, 12(4):046007, 2015.

[5] W. Biesmans, N. Das, T. Francart, and A. Bertrand. Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, (in Press), 2016.

[6] R. Zink, B. Hunyadi, S. Van Huffel, and M. De Vos. Mobile EEG on the bike: disentangling attentional and physical contributions to auditory attention tasks. *Journal of Neural Engineering*, 13(4):046017, 2016.

[7] C. Zich, M. De Vos, C. Kranczioch, and S. Debener. Wireless EEG with individualized channel layout enables efficient motor imagery training. *Clinical Neurophysiology*, 126(4):698–710, 2015.

[8] A. Aroudi, B. Mirkovic, M. De Vos, and S. Doclo. Auditory attention decoding with EEG recordings using noisy acoustic reference signals. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 694–698. IEEE, 2016.

[9] S. Akram, A. Presacco, J. Z. Simon, S. A. Shamma, and B. Babadi. Robust decoding of selective auditory attention from MEG in a competing-speaker environment via state-space modeling. *NeuroImage*, 124:906–917, 2016.

Address for correspondence:

Rob Zink — KU Leuven (Belgium), Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics. rob.zink@esat.kuleuven.be