Deep learning and eye-tracking for accurate EOG rejection

Scott Huberty
Keck School of Medicine
University of Southern California
Los Angeles CA 90027
seh33@uw.edu

Christian O'Reilly
Department of Computer Science
Artificial Intelligence Institute
Carolina Autism and Neurodevelopment Research Center
Institute for Mind and Brain
University of South Carolina
Columbia SC 29201.
christian.oreilly@sc.edu

Abstract

Electroencephalography (EEG) is a neuroimaging technique used to record the electrical activity generated by the brain. EEG recordings are often contaminated by various artifacts, notably those caused by eye movements and blinks (EOG artifacts). Independent component analysis (ICA) is commonly applied to isolate EOG artifacts and subtract the corresponding independent components from the EEG signals. However, ICA is an unsupervised technique that contains no knowledge of the eye movements during the task or the generative process by which these movements result in EOG artifacts. It is generally difficult to assess whether subtracting EOG components estimated through ICA removes some neurogenic activity. To address this limitation, we developed a deep learning model for EOG artifact removal that exploits information about eye movements available through eye-tracking. We leveraged the *Large Grid* task from the open-source EEGEyeNet dataset to develop and validate this approach. In this task, 30 participants looked at a series of dots appearing at 25 predetermined positions on the screen (810 trials/participant). EEG and eye-tracking were collected simultaneously. In this paper, we show that we can train a long short-term memory (LSTM) model to predict the component of EEG signals predictable from eye-tracking data. We further used this eye-tracking-informed evaluation of EOG artifacts to investigate the sensitivity and specificity of ICA, currently the dominant approach for EOG artifact correction. Our analysis indicates that although ICA is very sensitive to EOG, it has a comparatively low specificity.

1 Introduction

Electroencephalography (EEG) is a non-invasive neuroimaging technique used to record the electrical activity generated by the brain. It involves placing electrodes on the scalp to detect the tiny voltage fluctuations resulting from the synchronized firing of neurons. However, the eyes, heart, and muscles also generate small electric currents that can be captured by EEG electrodes. This is also true of small currents due to the electromagnetic fields generated by electrical devices. All these electrical currents travel across the tissues of the head and the EEG electrode array captures this mixture of neural signals mingled with biological and non-biological artifacts. Although these artifacts may have higher amplitude than the neural signal and overlap with the frequency content of the neural activity, they often have highly characteristic time and frequency properties that can be used for their detection and removal.

Preprint. Under review.

EOG (electrooculography) artifacts in EEG recordings refer to the electrical signals generated by eye movements and eye blinks. EOG artifacts arise due to the potential difference between the cornea and the retina, which acts as a dipole. When the eyes move, this dipole also moves, generating a changing electrical field that propagates instantaneously across the volume of the head and reaches the EEG electrodes [3]. Thus, EOG artifacts are omnipresent in the EEG signal and are particularly problematic in EEG tasks requiring eye movements, as they can obscure the neural activity related to the experimental paradigm. Therefore, developing artifact-reduction techniques for removing EOG is an important topic in EEG research.

Researchers apply various strategies to prevent and remove artifacts from EEG recordings. To prevent EOG artifacts, the researcher may design experiments such that the participant must gaze at the center of the screen and keep their eyes fixed so that the electrical dipole of the eye is kept in a fixed orientation throughout the experiment. This, however, limits the design of naturalistic paradigms where the participant can visually explore their surroundings. Such limitations are also inadequate for brain-computer interface (BCI) and mobile EEG setups.

Various strategies and techniques have been proposed to reduce the impact of artifacts on EEG recordings. For one, the researchers may remove periods of the EEG recording containing artifacts, likely at the expense of a large amount of data loss [7]. Alternatively, various techniques can be used to separate and remove the artifactual components from the neural components of the EEG signals. Among that class of algorithms, Independent Component Analysis (ICA) has become arguably the most widely adopted approach [8]. This blind source separation technique can be used to reduce stationary and stereotypical artifacts (ones with a stable spatial projection pattern across the electrodes) from EEG signals [1]. This approach has the benefit of retaining the full EEG time course but may fail to fully remove the artifact (insufficient sensitivity) or may distort the neural signals (insufficient specificity). ICA returns a set of components representing different sources of the EEG signal. Researchers then make topographic plots of these components and assess their spatial projections to decide whether each component is an artifact, neural signal, or a mixture of both. EOG artifacts, for example, usually account for a large amount of variance in the data and have greater amplitude in the channels near the eyes. Since ICA is an unsupervised algorithm, it can be difficult to confirm that neural signal is not included in the components labeled as artifact and, thus, some neural signal can be lost when these components are subtracted from the EEG data. Further, it should be noted that ICA assumes that all sources in the data are independent and that their spatial representation does not vary with time. A large amount of data (several electrodes, and tens of minutes of data) is also usually required to achieve a stable ICA decomposition. Finally, since ICA is an unsupervised algorithm, expert judgment is still needed to decide which components are artifacts.

While ICA can be used to identify and remove various artifacts (EOG, ECG, power line noise), there are alternative artifact reduction techniques for removing EOG artifacts specifically. EOG electrodes (peripheral reference electrodes placed near the eyes) can be used in a linear regression to estimate the corresponding EOG artifact in the EEG electrodes, which can then be subtracted from the signal [2,6]. It should be clarified that EOG electrodes differ from EEG electrodes only in that they are placed near the eyes, and, thus, are also susceptible to picking up neural signals or other types of artifacts (e.g., muscle contraction, sweating) that can diminish their utility as a reference signal for EOG activity.

Integrated EEG and eye-tracking acquisition systems are becoming increasingly common. In these setups, EEG is acquired, while at the same time, an eye-tracker is used to record the participant's gaze, generally in pixels or visual angles relative to a screen used to display stimuli. These systems can measure precisely what the eye is doing at any time, whether the participant is fixating to a point on the screen, shifting their gaze (referred to as a saccade), or blinking. Importantly, the eye-tracking signal is electrically independent of the EEG signal, so there is no risk of neural signal or physiological artifact bleeding into the eye-tracking signal.

Previous studies have explored leveraging eye-tracking data to enhance EOG artifact reduction in simultaneously collected EEG data, mostly by improving ICA decomposition or classification. For example, Plöch and colleagues proposed classifying independent components as EOG based on their variance during saccade periods, under the assumption that components with greater variance during saccades are associated with EOG activity [14]. In 2020, Dimigen proposed a method for fitting ICA specifically to periods of the EEG data where the participant was blinking or moving their eyes (as determined by the eye-tracking signals) to improve the decomposition of EOG activity into independent components [3].

However, in situations where EEG and synchronized eye-tracking data are available, given the ground-truth information on eye movements that these systems provide, it is no longer necessary to rely on unsupervised techniques like ICA to reduce EOG artifacts. Instead, the knowledge of the eye movements should be fully leveraged during EOG artifact reduction in a supervised or self-supervised manner. To this end, using eye-tracking signals for a linear regression as previously described is not expected to work well for estimating EOG artifacts because the relationship between the gaze (as recorded by the eye tracker) and the EOG signal is expected to be nonlinear. Thus, we propose a deep-learning model for reducing EOG artifacts using eye-tracking signals. Not only do deep learning algorithms work well with nonlinear relationships, but model architectures like recurrent neural networks (RNN) may be particularly suited for this application since they can account for temporal correlations within time series. Further, to support the improvement of EOG-removing techniques for cases where synchronized eye-tracking data is not available, we use this approach to evaluate the performance of ICA and characterize its sensitivity and specificity to EOG artifacts.

2 Method

2.1 Datasets

To develop and test the proposed model, we leveraged the open-access EEGEyeNet benchmark dataset [15]. The complete EEGEyeNet dataset contains recordings from 356 healthy adults, including simultaneously collected high-density 129-channel EEG data (see Figure 5 for EEG electrode locations) synchronized with video-infrared eye-tracking. The eye-tracking data include three channels, two for the position in X and Y, and one for the pupil size. Both raw and preprocessed data are included in the EEGEyeNet dataset. The experiment includes three tasks: a pro- and antisaccade task, a visual symbol search task, and the *Large Grid* task. For our study, we used the Large Grid task [16], which has been performed by 30 participants. In this task, the participants are asked to look at dots appearing at 25 different positions, distributed across the whole surface of a screen. Each dot is presented for 1.5 to 1.8 seconds, in a pseudo-randomized order (see [15] for details on this pseudo-randomization). The central dot is presented 3 times, while the other dots are presented one time each per block. Five blocks were used by run, and six runs were recorded for a total of 810 stimuli per participant. Each run is saved as a separate recording, providing 177 recordings (i.e., 3 were missing).

2.2 Outlier rejection

Before running analyses, we excluded recordings with noisy or unreliable eye-tracking data, which could have been due to many reasons, the most likely being poor eye-tracker calibration. We identified outliers by first computing the mean squared difference between the event-related response (ERR) of every recording and the ERR averaged across recordings:

$$\mathbf{e} = \overline{(X - \overline{X}^{r})^2}^{t} \tag{1}$$

where X is a matrix of x and y coordinate gaze values, and the bars with r and t represent the averaging across the recording and time dimensions, respectively. These errors were computed for each eye-tracking channel and dot, in each recording. We then computed the 25th, 50th, and 75th quantiles of these error terms. To determine whether an individual recording was an outlier, we used the classic outlier rejection formula, $e > Q_{50} + k(Q_{75} - Q50)$ with $k = 6$ that we rewrote

$$\frac{\mathbf{e} - Q50}{Q75 - Q50} > k \tag{2}$$

to average the left-hand term across channels and dots (i.e., event type) before comparing it with the threshold. Using this criterion, 8 recordings were deemed outliers and were excluded from further analyses, most of them being the different runs of the same participant.

We confirmed participant compliance with the large grid task instructions and the quality of the eye-tracking data for the remaining recordings by displaying the kernel density estimation of the distribution of the X/Y pixel coordinates for every dot in the large grid (supplementary figure 6). In

this Figure, the gaze position has been determined as the averaged position in the [0.3, 1.0]s time window.

2.3 Preprocessing

Per the description of the authors of the EEGEyeNet paper [15], this dataset has been processed *minimally* and *maximally*, as defined in the *Automagic* open-access toolbox they adopted [12]. For our analyses, we used the signals preprocessed minimally, which includes the detection and interpolation of bad channels, and the filtering to the 0.5-40 Hz band. This minimal preprocessing does not include ICA artifact rejection as this step would remove the EOG artifacts necessary for our study. The authors synchronized the EEG and eye-tracking signals and confirmed no synchronization error exceeded 2 ms.

To make our analysis more computationally efficient, we filtered to the 1-30 Hz band before down-sampling the signals to 100 Hz using MNE-Python [5]. Although the dataset is recorded with a sampling rate of 500 Hz, EOG signals are limited to relatively low frequencies due to natural biomechanical constraints imposed on the kinematics of eye movements. Thus, such a high sampling rate significantly increases the network size (i.e., multiply by a factor of five the LSTM input shape and the associated weights to learn) without adding relevant information.

For machine learning, recordings were epoched into contiguous 1s segments (this should not be confused with the concept of training epochs in deep learning). We also set an average reference. For our comparison with ICA, we used the Extended Infomax approach [9] as implemented in MNE-Python. EOG-associated components were detected automatically as those labeled as "eye blink" by MNE-ICALabel [10], which ports to Python the functionalities of ICLabel [13] (i.e., ICLabel has seven classes of independent components: "brain", "muscle artifact", "eye blink", "heart beat", "line noise", "channel noise", and "other").

2.4 Deep-learning Model

The overarching idea of this approach is to teach an RNN how to predict EEG signals only from eye-tracking signals. Of course, only a small portion of the EEG signals will be predictable from eye-tracking signals, and this predictable portion will be due to EOG artifacts and potentially some neural and non-neural correlates of eye movements (e.g., electromyographic signals due to the activation of the muscle required for eye movements). More formally, for the EEG signal matrix Y (EEG channels X time) and the eye-tracking matrix X (eye-tracking channels X time), we model this relationship as

$$Y = f(X) + R \tag{3}$$

where R is a matrix of the residual (i.e., the neural signals, in this case) and f is a nonlinear function that we want to learn by adjusting the RNN weights to minimize the mean square amplitude of R. For this task, we used a 2-layer LSTM with 3 features corresponding to eye-tracking channels and 64 hidden states whose outputs get pruned with a 0.5 dropout layer, and fed to a fully connected layer which maps these internal states to 129 outputs corresponding to the EEG channels. This model has been implemented in PyTorch and fitted using the ADAM optimizer with a 0.01 learning rate and a mean-square-error (MSE) loss function. We found that 1000 training epochs (not to be confused with the EEG 1s epochs) were enough to reach stable training loss. For this analysis, we did not attempt to test for generalizability across participants or recordings but rather wanted to test if it was possible to learn within a participant the mapping between eye-tracking signals and their impact on EEG. Thus, no hold-out or cross-validation was implemented. That is, the mapping was learned independently for all 177 recordings, and each of these mappings was directly used to clean the EEG of the corresponding recording only.

2.5 Analysis

EOG signals are known to affect mostly frontal EEG channels. To validate the capability of the neural network to detect and remove EOG noise, we computed the percentage of signal removed per

channel as

$$\Delta = 1 - \frac{R}{Y} \tag{4}$$

Further, we validated the specificity of EOG correction with our method by looking at how much the EEG signals are distorted by the EOG removal before any eye movement happens. This period during which we would like to see little or no distortion of the EEG can be defined by a baseline period (chosen to start 0.2s before the onset of the stimuli) up to the reaction time measured as the moment at which the gaze position has changed by 5% of the amplitude of the response.

Based on that reaction time, we used an approach similar to the one adopted in [17] to characterize the sensitivity and specificity of removing eye artifacts based on eye-tracking versus automated ICA. In this approach, we defined a segment where no noise is expected, here defined as the window from the start of the baseline period (-0.2s) to the reaction time (RT), and a window where an artifact is expected (RT to 1s). Defining the signal (S) as the root mean square (RMS) amplitude of the original recording (Y), and the noise (N) as the RMS amplitude of what has been removed by the artifact removal approach (i.e., $N = Y - R$), we can define the classic measure of SNR in dB as

$$SNR = 10 * log_{10}\left(\frac{S}{N}\right) \tag{5}$$

A large SNR before the reaction time is indicative of a good specificity (i.e., clean signals do not get distorted by artifact removal) whereas low SNR after the reaction time is indicative of a good sensitivity (i.e., more noise has been detected and removed by the cleaning approach).

3 Results

A demonstration of cleaned EEG signals using the proposed model and ICA compared to the original EEG signals is shown in figure 1.

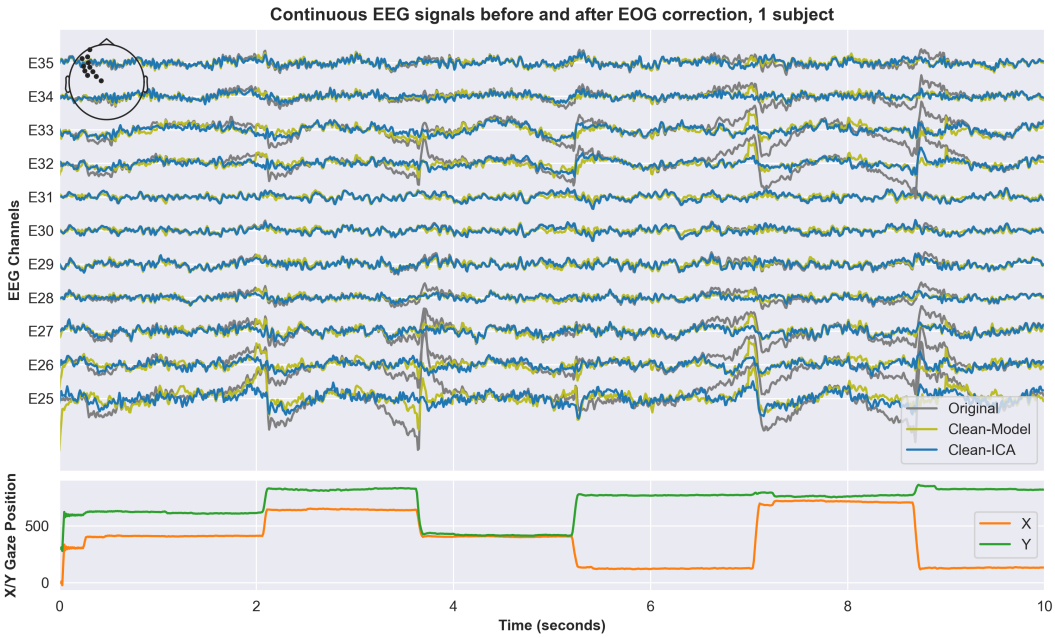

Figure 1: Top panel: EEG signals for a typical recording before (gray) and after EOG artifact reduction using the proposed model (green) and ICA (blue). Bottom panel: X and Y gaze positions (reported in pixel coordinates) during the same period as the top panel.

To validate our approach more systematically, we looked at the distribution of the predicted noise over the scalp (Figure 2), averaged across participants. This distribution clearly shows the bias toward the more frontal regions, with about 70% of the amplitude of the recorded signals in prefrontal and

frontal channels being due to EOG artifacts. Further, the proposed model predicted less EOG in the pre-stimulus period as compared to ICA. Given that the participants' gaze was generally fixed to the center of the screen (i.e. not moving) during this period (see the bottom right panel of figure 3 for a representation of this), this suggests that the proposed model does not distort the signal as much as ICA.

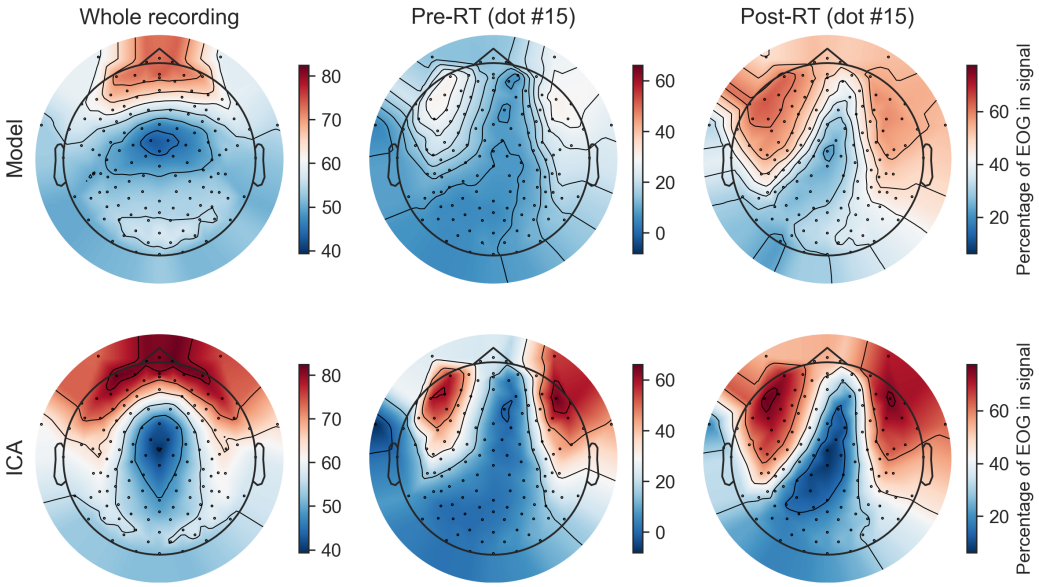

Figure 2: Topographic plots showing the spatial projection of the predicted EOG signal (across participants), as predicted by the model (top row) and by ICA (bottom row). The spatial projection is shown for the full time-course of the recording (left column), as well as the pre-RT period (middle column), and the post-RT period (right column). For the pre-RT and post-RT plots, only trials where dot 15 was presented were used.

Next, we assessed the evoked EEG time series (averaged for each dot) before and after EOG artifact reduction, once using the proposed deep learning model, and once using ICA. As can be seen in Figure 3, the proposed deep learning model appears to be more conservative in EOG reduction than ICA and distorts the signal less in the pre-stimulus period.

To quantify the specificity and sensitivity of the proposed model for reducing EOG artifact, we computed the SNR as described in section 2.5. The SNR values for the period before (specificity) and after (sensitivity) the RT were averaged within subjects (across runs of the task), and a paired t-test was computed to compare the SNR values for the proposed model versus ICA. An example of our results is displayed for channel E25 and dot 25 in Figure 4.a. For this combination of channel and dot, the average values for the Model and ICA would indicate higher specificity for the model but a higher sensitivity for ICA. This process has been repeated for all channels and dots and aggregated to be displayed on topomaps for specificity (Figure 4.b) and sensitivity (Figure 4.c). The results suggest that this tendency toward higher specificity but lower sensitivity for the model compared to ICA can be generalized across the scalp, except for a higher sensitivity of the model in the posterior region of the scalp.

4 Discussion

This study demonstrated a novel approach for EOG artifact rejection in EEG signals recorded simultaneously with eye-tracking. We made the code implementing this approach available on GitHub at https://github.com/lina-usc/eog-learn. The emergence and popularization of such recordings [4] has opened new possibilities by providing reference signals closely associated with EOG generation.

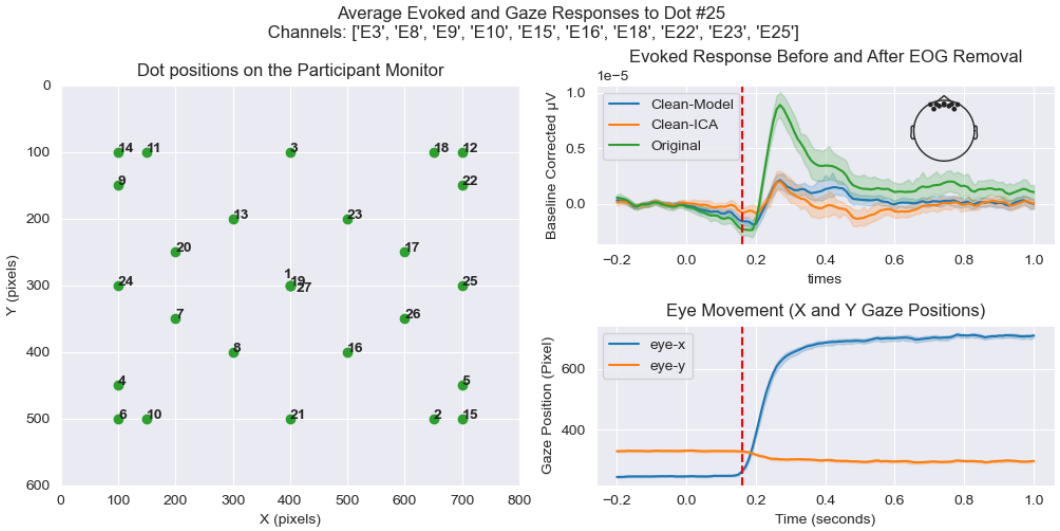

Figure 3: The position of each presented dot during the Large Grid task (left panel), the evoked response to dot 25 before and after EOG removal (top-right panel), and the average eye movement (X/Y pixels) during those trials (bottom-right panel). The vertical red dashed line represents the average reaction time across participants, i.e. the moment they began to shift their gaze toward the dot displayed at the beginning of the trial (i.e. at t= $0s$).

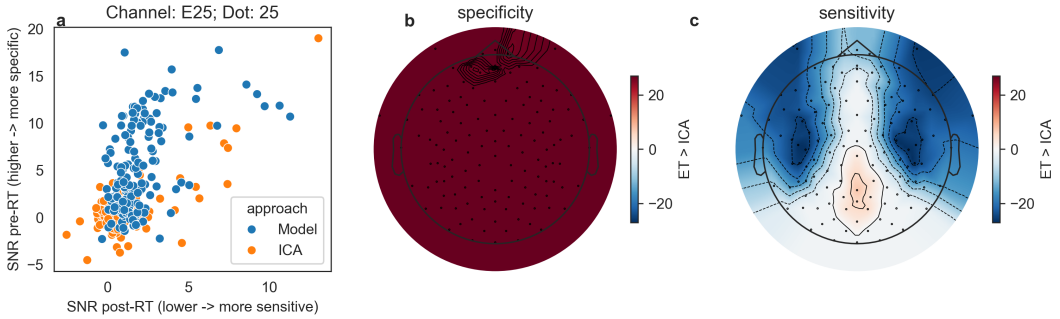

Figure 4: Performance of our approach based on eye-tracking signals (ET) versus ICA. a) Example of analysis of sensitivity and specificity for a specific channel (E25, a frontal channel) and dot 25. b) Comparative performance of the specificity for the two approaches across the scalp. For each channel and dot, specificity is determined as illustrated in panel a. SNR values are then averaged within subjects (across runs) and a paired t-test is computed to compare the SNR values for ET vs ICA. Cases where the specificity is larger for ET than ICA are counted as a +1, cases where ICA is more specific than ICA are counted as -1, and cases that are not statistically significant ($p > 0.05$) are counted as a 0. The sum of these scores is computed across the 27 dot conditions and displayed as a topomap. c) Same as for b, but for sensitivity.

One of the constant challenges of EOG rejection is the absence of a ground truth for evaluating the effectiveness of proposed approaches. This common situation limits the options available to investigators to the following: 1) use synthesized recordings with a known ground truth but questionable face and ecological validity, or 2) use recorded EEG with unknown ground truth. The addition of eye-tracking data, although not providing ground truth for EOG artifacts *per se*, partly mitigates this thorny issue by providing reliable information on eye motions usable for inferring EOG artifacts.

Taking benefit from these signals, we adopted a *data-driven black-box* approach to EOG artifact removal from EEG signals. The data-driven qualifier comes from using eye-tracking signals and deep learning to empirically map the association between the movement of the eyes and its impact on EEG signals. We use the black-box qualifier to contrast with an approach using eye-tracking signals and a

physiological model of how eye movements generate EOG artifacts. Our approach does not consider any knowledge specific to the application at hand. All the intelligence of the solution is generic to the task of learning a relationship between an input and an output given enough data. Thus, conceptually, this approach may be suitable for other applications (e.g., removal of electrocardiogram artifacts in EEG, correction for the effect of motion on electrocardiogram signals), so long as the source of the contamination is due to a process for which we have a separate reference signal.

Crucially, the approach we adopted provides an opportunity to more objectively assess the performances of blind source separation conducted with ICA. This technique currently dominates the field [8]. It has been shown in most cases to perform very well, and our analyses support in many ways this assessment. However, ICA for EOG rejection is known to have limitations [17]. More importantly, although its capacity to remove EOG artifacts can be readily evaluated on noisy EEG signals (e.g., see Figure 1), the degree to which it may distort neural signals is more difficult to establish. For example, ICA tends to distort the phase of EEG signals [11,18], a key element in neural dynamics and a property essential for all functional connectivity metrics based on phase consistency (e.g., coherence, phase locking value, phase lag index). In our experiment, we demonstrated that eye movement information can be used to effectively remove EOG artifacts while distorting neural signals significantly less. Although the lack of a ground truth creates some uncertainty in the interpretation of the SNR-based measures of sensitivity and specificity, our approach has shown a much higher specificity than ICA. In general, ICA has shown a higher sensitivity. However, it is interesting to note that our approach was more sensitive to parts of the scalp that typically are less impacted by EOG artifacts (i.e., central/occipital regions; see Figure 4.c). Our comparison with ICA relied on the automated classification of independent components associated with eye movement artifacts. The classifier we used (i.e., ICLabel) has been designed using machine learning and a large dataset of independent components annotated by experts. This classification is, therefore, vulnerable to bias associated with our current understanding of how independent components associated with EOG artifacts should look. EOG artifacts are known to have the most impact on the frontal region. However, although electrical dipoles generated close to the forehead may have their strongest effect on that part of the scalp, their field wraps around virtually the whole head (with decreasing amplitude due to attenuation). Our results suggest that independent components selected for rejection may tend to undercorrect for these more distant effects. This observation highlights the most significant contribution of this approach: using eye-tracking for assessing the impact of eye movement on EEG may provide us with a more reliable and objective assessment. We can then use this assessment to develop new methods or correct existing methods that do not require the availability of eye-tracking data.

5 Limitations

The need for synchronized EEG and eye-tracking recordings constitutes the most obvious limitation of the approach we proposed in this paper. However, although we may use this approach directly for cleaning EEG in such datasets, this application was not the main reason for this study. We argue that using such a dataset to develop methods that can highlight limitations and suggest possible ways to remedy current techniques not requiring eye-tracking data is of greater interest. In our analysis, we focused our comparisons on the automated ICA approach because it is currently the most widely used approach for EOG removal.

For our specific application, we attempted to learn the predictable part of the EEG based on eye-tracking information. This part is arguably a minor portion of the EEG makeup. The learning task is, therefore, complicated by the relatively low percentage of predictable information in the signals. We demonstrated that even with a relatively small amount of data (i.e., single recordings), it is possible to learn that relationship with a satisfying degree of precision. However, we could possibly improve the sensitivity and specificity of EOG removal by using a larger training set. The degree to which the performance is saturated with the current size of the training data would need further investigation.

The data used for this study (i.e., looking at many predefined targets on a screen) were particularly well-suited for our analysis and for learning the mapping between eye movement and EOG artifacts. However, we used information on the structure of the task only for performance analysis. Our training did not rely explicitly on the properties of the experimental protocol (i.e., the whole recording was segmented in 1s epochs and passed to the training routine without any information on the stimuli),

but implicit characteristics (e.g., the fact that eye movements were systematically covering the whole screen area) may have been beneficial.

It is also worth considering that the relationship learned between eye movement and the predictable part of the EEG is made up of various contributions, including a component due to the artifact generated by the movement of the eye (the component we generally want to remove) and the neural activity systematically correlated with the eye movement, such as the neural signals controlling the movement of the eye and the change in neural activity created by the change in visual stimuli as the line of sight shifted. Depending on the investigated hypotheses, the analyst may or may not want to remove these latter components. The approach considered in this study cannot disentangle these different components. We would argue, however, that it offers a powerful framework for investigating these different components, for example, by using virtual reality to experimentally control changes in the visual field as a function of eye movement.

Lastly, we decided in this study to perform learning and testing on single recordings. One advantage of this approach is that it is constrained to the recordings themselves (e.g., it is independent of the subject sample size). This approach is therefore tailored to the specificity of the participants (e.g., the specific shape and electrical properties of the head of the participant affect how electrical currents generated by the movement of the eyes travel and are recorded in EEG) and of the recording (i.e., effects of the experimental protocol, environment factors, etc.). Thus, our study did not aim for the generalizability and reusability of these deep learning models. Future work could consider training a single model across a large sample to target generalizability and reusability. It is also possible that by benefiting from a larger sample for training, the mapping learned would be more precise. Whether or not the gain obtained from training across subjects would offset the loss of precision due to interindividual and within-individual/between-recordings variability is currently unknown.

6 Conclusion and Future Works

We presented a deep-learning approach that leverages eye-tracking information in synchronized eye-tracking/EEG recordings to remove EOG artifacts from EEG recordings. Most importantly, we demonstrated how this objective source of information can be harnessed to benchmark existing approaches and better understand their limitations. In future work, we plan to address the limitations associated with this black-box approach by designing a generative model of how EOG artifacts are generated from eye movements relying on physiological knowledge. Provided that 1) we dispose of enough information to develop a faithful generative model of EOG artifact from measurements of the eyes position and 2) given that precise eye-tracking information is available, this source of artifact could be removed accurately from the EEG. Furthermore, in combination with the approach we presented here, the EOG and the neural component associated with eye movements could then be disentangled.

References

[1] Maximilien Chaumon, Dorothy V.M. Bishop, and Niko A. Busch. A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*, 250:47–63, July 2015.

[2] R.J. Croft and R.J. Barry. Removal of ocular artifact from the eeg: a review. *Neurophysiologie Clinique/Clinical Neurophysiology*, 30(1):5–19, February 2000.

[3] Olaf Dimigen. Optimizing the ica-based removal of ocular eeg artifacts from free viewing experiments. *NeuroImage*, 207:116117, February 2020.

[4] Olaf Dimigen and Benedikt V Ehinger. Regression-based analysis of combined eeg and eye-tracking data: Theory and applications. *Journal of vision*, 21(1):3–3, 2021.

[5] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, et al. Meg and eeg data analysis with mne-python. *Frontiers in neuroscience*, 7:70133, 2013.

[6] Gabriele Gratton, Michael G.H Coles, and Emanuel Donchin. A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 55(4):468–484, April 1983.

[7] Riitta Hari and Aina Puce. *MEG - EEG Primer*. Oxford University PressNew York, September 2023.

[8] Xiao Jiang, Gui-Bin Bian, and Zean Tian. Removal of artifacts from eeg signals: a review. *Sensors*, 19(5):987, 2019.

[9] Te-Won Lee, Mark Girolami, and Terrence J Sejnowski. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural computation*, 11(2):417–441, 1999.

[10] Adam Li, Jacob Feitelberg, Anand Prakash Saini, Richard Höchenberger, and Mathieu Scheltienne. Mne-icalabel: Automatically annotating ica components with iclabel in python. *Journal of Open Source Software*, 7(76):4484, 2022.

[11] Rodrigo Montefusco-Siegmund, Pedro E Maldonado, and Christ Devia. Effects of ocular artifact removal through ica decomposition on eeg phase. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 1374–1377. IEEE, 2013.

[12] Andreas Pedroni, Amirreza Bahreini, and Nicolas Langer. Automagic: Standardized preprocessing of big eeg data. *NeuroImage*, 200:460–473, 2019.

[13] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198:181–197, 2019.

[14] Michael Plöchl, José P Ossandón, and Peter König. Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Front. Hum. Neurosci.*, 6:278, October 2012.

[15] Martyna Beata Płomecka, Ard Kastrati, and Nicolas Langer. Eegeyenet. 2023.

[16] Jake Son, Lei Ai, Ryan Lim, Ting Xu, Stanley Colcombe, Alexandre Rosa Franco, Jessica Cloud, Stephen LaConte, Jonathan Lisinski, Arno Klein, et al. Evaluating fmri-based estimation of eye gaze during naturalistic viewing. *Cerebral Cortex*, 30(3):1171–1184, 2020.

[17] Diksha Srishyla, Sara Jane Webb, Mayada Elsabbagh, Christian O'Reilly, and BASIS Team. Eye-movement artifact correction in infant eeg. *bioRxiv*, pages 2024–03, 2024.

[18] RW Thatcher, EP Soler, DM North, and G Otte. Independent components analysis "artifact correction" distorts eeg phase in artifact free segments. *J Neurol Neurobiol*, 6(4):5–7, 2020.

Supporting information

EEG Sensor Montage

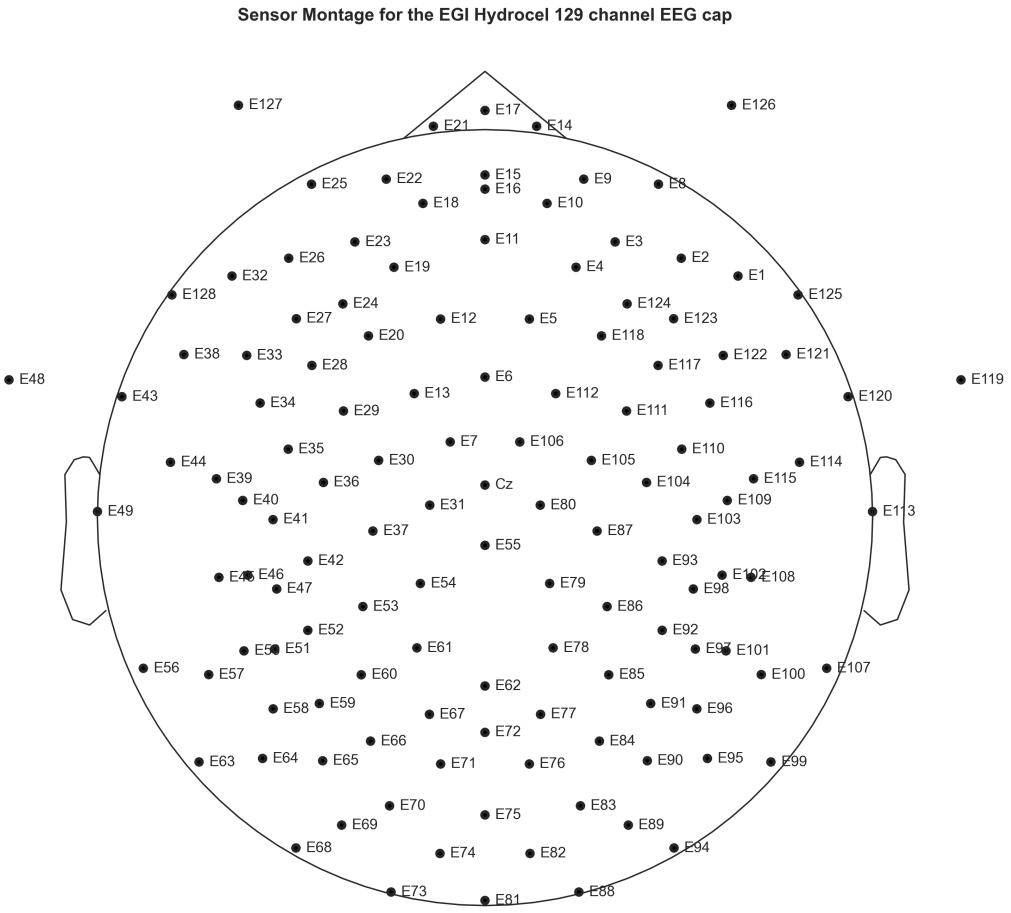

Figure 5: Montage displaying the idealized sensor locations of the EGI 129 channel EEG cap.

Eye-tracking Validation

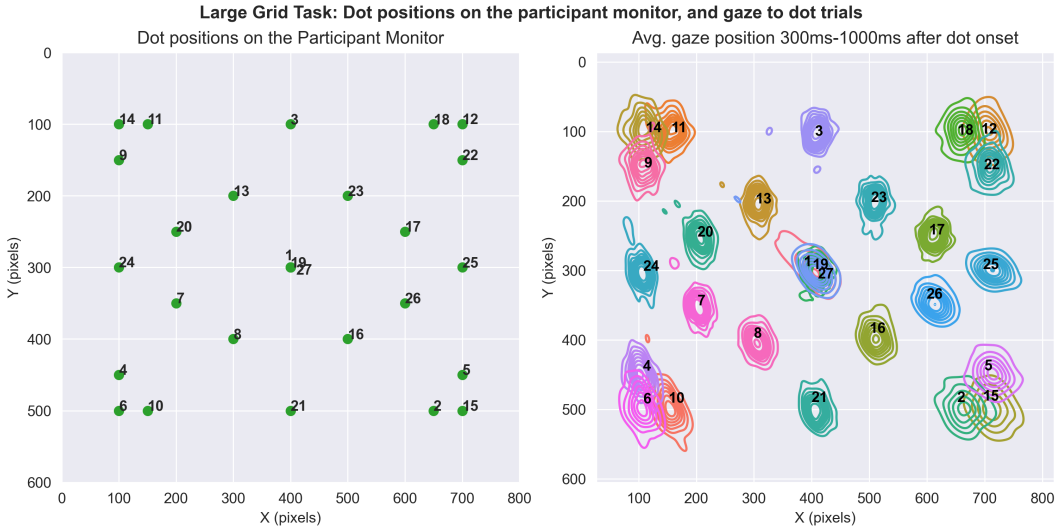

Figure 6: Pixel coordinates on the participant screen where each dot was displayed (left), and the average gaze position in response to the dot onsets, across participants.

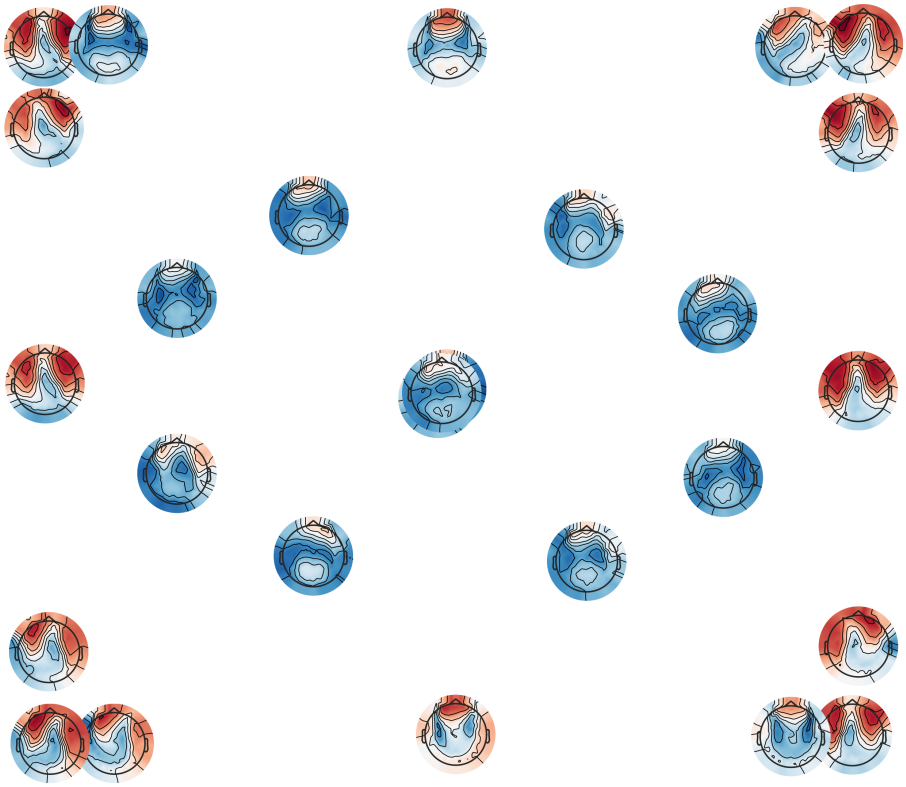

Figure 7: Topographic representation of the predicted EOG after gaze to each dot, averaged across participants.