Biomedical Engineering, Vol. 11, No. 1, 2014

ON THE EFFECTIVENESS OF ICA BASED EYE ARTIFACT REMOVAL FROM EEG WINDOWS OF DIFFERENT LENGTHS

Foad Ghaderi[1], Elsa Andrea Kirchner[1,2]
[1] Robotics Group, Faculty of Mathematics and Informatics, University of Bremen, Bremen, Germany
[2] Robotics Innovation Center, German Research Center for Artificial Intelligence (DFKI GmbH), Bremen, Germany
fghaderi@uni-bremen.de

ABSTRACT
Eye artifacts, i.e., blinks and saccades, are usually non-avoidable when recording electroencephalogram (EEG) data. These artifacts can affect the performance of classifying the EEG patterns especially in real world applications, e.g. brain computer interfaces. To evaluate the effectiveness of independent component analysis (ICA) based eye artifact removal methods, the data are analyzed in batch and window-based modes in this paper. Despite the improvements achieved in the batch mode, it turns out that applying the removal methods to overlapping windows of the EEG data stream does not improve the classification performance.

KEY WORDS
Electroencephalogram, eye artifacts, independent component analysis, artifact removal.

1 Introduction

Eye activities, major sources of artifacts in EEG and event related potentials (ERPs), are barely avoidable in practical recording sessions. These activities can affect the performance of systems that rely on single trial ERP detection, e.g., brain computer interface (BCI) systems. Considering the characteristics of EEG signals, different methods have been proposed or adapted so far to remove the artifacts while preserving the signals of interest [1, 2].

In a previous study [3], a number of artifact removal methods were compared. The main criterion for algorithm selection was their possibility to be used in online scenarios. It was shown that a combination of infomax [4] and ADJUST [5] was providing better performance compared with other methods. The Infomax method was used for decomposing the multichannel EEG data into independent components (ICs), and ADJUST was used to identify the artifacted ICs. The study in [3] is based on removing the artifacts from the whole EEG dataset recorded in an oddball paradigm and then classifying the target and non-target ERPs. However, in some applications it is necessary to process the data in an online manner. Therefore, artifact removal approaches that rely on processing the whole dataset are not applicable anymore because in an online scenario only the previously recorded samples are available. On the other hand, taking into account that in online scenarios the response time of the system have to be in a limited range, one can not use very long segments of data as the input to artifact removal algorithms.

A common approach for online artifact removal is using adaptive filters [1]. However, using methods like adaptive filters and regression need separate references (i.e. electrooculogram (EOG) channels), which are not always available. Considering the fact that the ICA methods have good removal performance when applied to the whole data, they have also been used for online artifact removal. In [6] an ICA based method for offline and online artifact removal was proposed. The artifact channels were identified using support vector machines. The ICA and canonical correlation analysis where compared in [7] and it was shown that the latter outperforms the former in online muscle artifact removal. The effect of ICA based artifact removal methods on different frequency ranges of EEG were also investigated in [8].

ICA methods usually optimize cost functions, which are defined based on statistical properties of the data. In theory, these properties are defined when infinite number of samples are available. Using data with short lengths would cause the statistics to be sensitive to outliers, and hence the ICA methods will overlearn the data. As a general rule to avoid overlearning, it has been proposed in [9, 10, 11] to use $k \times n^2$ samples for a successful ICA decomposition, where, n is the number of sensors and k is empirically set to a value between 5 to 32. It is also shown in [9, 12] that for very short windows of data (in the extreme case the number of observations equal the number of sensors) the ICA methods overlearn the data by estimating spikes or bumps as independent sources. The proof for the case of fastICA is presented in [9]. However, using simulated and real EEG data, it is shown in [13] that the fastICA performs almost in the same range when using data with different lengths. The evaluation was conducted using a defined performance index.

Considering the outcomes of previous studies, it turns out that more comprehensive investigations about the effects of the length of the data on performance of ICA methods is necessary. The purpose of this paper is to study the effectiveness (in terms of single trial ERP classification performance) of the ICA methods for removing eye artifacts from EEG data in an online manner. To this end, we use the infomax (as reported in [3]) and the fastICA [14] (com-

DOI: 10.2316/J.2014.216.818-0078

monly used for analyzing multi-channel biological signals) methods.

The paper is structured as follows. In the next section the methods for eye artifact removal are described. Data and processing flow for ERP detection is addressed in Section 3. Experimental results are presented in Section 4. Concluding discussions come in the final section.

2 Methods

2.1 ICA methods

Two commonly used ICA algorithms have been used in this study, i.e., infomax and fastICA. The infomax algorithm maximizes the output entropy or information flow of a neural network with nonlinear outputs [14]. The fastICA method maximizes the statistical independence of the estimated sources by maximizing the non-Gaussianity [4]. FastICA is inspired from the central limit theorem which states that the distribution of sum of independent random variables tends to have Gaussian distribution. The ICA methods generate a number of ICs, however, in order to identify the artifact ICs one has to check the ICs visually or use automated techniques like the ADJUST method.

2.2 ADJUST

This method exploits the combination of temporal course and spatial distribution of the independent components. Three different classes of eye artifacts are considered in this method, i.e., blinks, vertical and horizontal eye movements. First, ICA method is applied to the EEG data. For each artifact class, a detector is implemented by computing a class-specific set of spatial and temporal features on all independent components. For each feature, a threshold, which separates artifacts from non-artifacts is estimated on the whole set of ICs by the Expectation-Maximization automatic thresholding method. If all artifact-specific spatial and temporal features of a detector are larger than their respective thresholds, the IC is classified as an artifact channel. This way, for each artifact class a sorted list of channel indexes is returned (for more details of the method see [5]).

2.3 Online artifact removal

In an online application of single trial ERP classification, EEG data has to be cleaned as new samples arrive to be used for further processing. A common approach for online artifact removal using ICA methods is to apply the algorithm to some intervals of the EEG data, suspicious to be contaminated by artifacts [8]. Alternatively, the removal procedure can be used to clean all the intervals of the data considering the fact that eye blinks and saccades occur spontaneously all over the time. To do so, we apply the ICA to overlapping windows of data cut from the EEG stream.

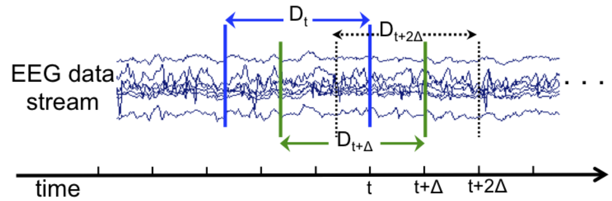


Figure 1: Windows from the EEG stream are cut and passed to the artifact removal method. This technique is used for online removal of eye artifacts.

Every Δ second, a window of data, with the length l ($l > \Delta$), is cut and passed to the artifact removal procedure. Figure 1 shows three windows cut at times t, $t+\Delta$ and $t+2\Delta$. Since the windows, $\mathbf{D}_{t_i}$, are overlapped, only the last Δ second of the cleaned data will be used by the next modules after removing the eye artifacts. Based on the arbitrary processing flow in the online application, these intervals can be used individually or in combination with previously cleaned intervals. For each window of the data, first the mean of all channels are subtracted and after removing the artifact ICs, the mean values are added to the cleaned data.

In order to identify and deflate artifact channels, the ICs and the ICA transformations are passed to the ADJUST algorithm. This technique provides a sorted set of IC channel indices that are likely to be blink, horizontal or vertical EOG. In a conservative approach we only used the first channel in the union of the three sets [3].

3 Data and processing

3.1 Data

Data from eight healthy male subjects (age: 29.9 ± 3.3 years; right-handed; normal or corrected-to-normal vision), recorded in a previous study, was used (see [3] for more details).

Experiments were performed in a shielding cabin to reduce the effect of non-physiological artifacts. The subjects were seated in a comfortable chair in front of a table. Two input devices were placed on the table at a distance of approximately 30 cm from each other. A monitor was used to give commands and feedback to the subjects. If no instructions or feedback was given a black fixation cross was presented in the middle of the screen on a green circle and the subjects had to put their right hand on the left input device. Subjects were instructed to continuously keep their eyes fixed on this cross and to executed slow movements of the right arm between two input devices on the table (used to label the begin and end of performed movements in the EEG) on command, i.e., whenever a target event (cross changed to a vertical line) was presented. Infrequent task-relevant target stimuli were interleaved with frequent task-irrelevant non-target stimuli (cross changed to a horizontal line) in an oddball fashion (ratio of $1:8$)

with an inter stimulus interval of 1000 ± 100 ms. Stimuli were shown for 100 ms. Too fast arm movements (duration < 1 second) and commission errors (movements on non-targets), were reported back to the subjects. A run ended after 40 correctly performed movements. Each subject performed 3 runs. The study was conducted in accordance with the Declaration of Helsinki and approved with written consent by the ethics committee of the University of Bremen.

The experiments were designed and run using Presentation software [Neurobehavioral Systems, Inc., Albany, USA]. EEGs were acquired with 5 kHz and reference at FCz (using a 128-channel actiCap system and BrainAmp DC amplifiers [Brain Products GmbH, Munich, Germany]) and were filtered between 0.1 to 1 kHz before saving to a PC. We used only the data from the first 64 channels (10-20 system).

3.2 Data processing

In order to use the whole existing data, we used all the epochs and did not reject any part of the EEG recordings. Data is processed in a flow of consecutive components as follows:

Down-sampling / filtering

This module down-samples the data to 100 Hz and filters the data between 0.1 Hz to 7 Hz. In order to avoid phase shifts, forward-backward filtering technique is used.

Artifact removal

As mentioned earlier the output of the online artifact removal method is an EEG window with the length $\Delta = 1s$ (regardless of the actual size of the input window, l). Considering the fact that the ERP instances are distributed in the EEG stream randomly, the cleaned data are concatenated. The ERP windows will be taken from this data.

In order to investigate about the effects of the artifact removal methods on the overall performance, they were used in three different modes:

- No artifact removal: EEG data are passed to the next processing component without any processing.
- Batch removal: removing the artifacts from the whole dataset.
- Window based removal: removing the artifacts from windows of length l, as illustrated in Fig. 1.

ERP windower

ERP instance windows were cut 0-1000 ms after the onsets of each stimulus.

Feature generation

Feature vectors were generated from the slopes of the lines fitted to overlapping windows cut from the retained channels of the spatial filter (each line fitted to a 400 ms segment with 120 ms overlap) and then passed to the classifier. For the features of each channel the variance was normalized to one, after subtracting the mean.

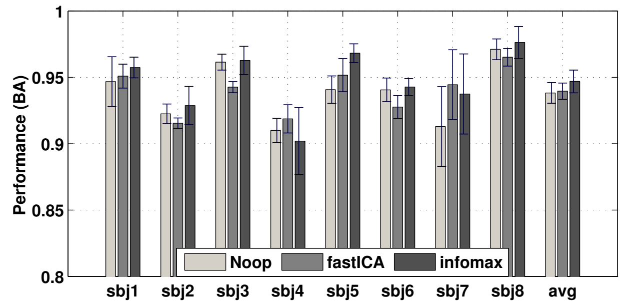


Figure 2: Classification performances obtained from infomax and fastICA methods applied in batch mode. Noop represents the case that no artifacts were removed.

Classification

The support vector machines (SVM) classifier (C-SVC with linear kernel) was used to discriminate the two ERP classes: targets vs. non-targets. We used LIBSVM [15] implementation of the classifier. The optimum complexity value for the classifier was found using grid search [16]. ERP samples in each dataset were randomly divided into 5 separate splits and evaluation performed by cross-validation using the leave-one-out technique. Results of the experiments are reported in terms of the balanced accuracy (BA) of classification, which is the average of TPR and TNR and therefore unaffected by unbalanced class distributions [17].

4 Results

The artifact removal methods are compared when applied in the batch mode and the results are presented in Fig. 2. The figure depicts the results obtained from infomax and fastICA combined with ADJUST. As a ground truth, the experiments were repeated when no artifact removal methods were used (Noop in the figure). Statistical tests (one-tailed paired t-test) of the batch mode results confirm that (compared with Noop) the infomax improves the classification performance ($p < 0.05$), while the achieved performance for fastICA does not significantly differ ($p = 0.41$). Also the results show that infomax outperforms fastICA ($p < 0.1$).

In order to investigate the effects of window length on the artifact removal performance in the online application, experiments with different EEG window lengths were conducted following the procedure explained in Section 2.3. According to the rule of requiring $k \times 64^2$ samples, and for $k = 5$, we need at least 205 seconds of data to be sure that the ICA algorithms do not overlearn the data. Considering the high computational costs, we evaluated the algorithms for windows of lengths $l \in \{1, 2, 3, 4, 5, 7, 10, 15, 20, 30, 40, 50, 60, 70, 80\}$ seconds.

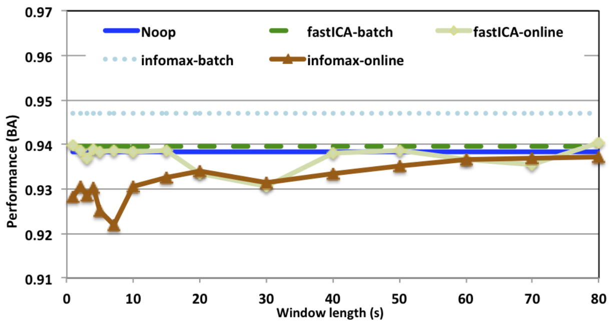
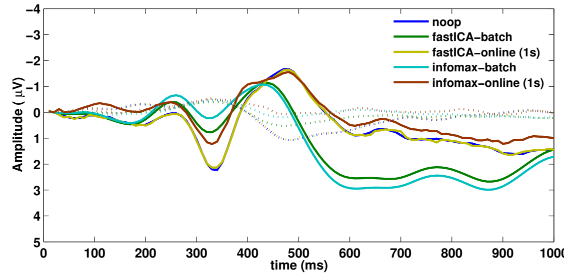


Figure 3: Classification performances obtained from the artifact removal methods based on the infomax and fastICA methods. The algorithms have been used in batch and online modes, i.e., the artifacts were removed from the windows cut from the EEG stream.

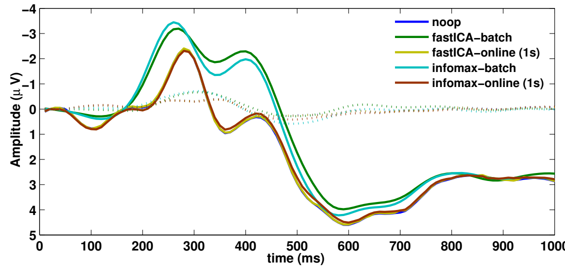
In our experiments we figured out that for very short windows (below 10s, i.e. $n_s < 0.25n^2$, where n_s is the number of samples) the algorithms can not converge and in particular the outcome of the infomax algorithm was complex number transformations. The problem can be resolved by reducing the dimensions of the data [9]. Therefore, in all the experiments we only kept the first r principal components, where r is the rank of the data matrix.

Figure 3 illustrates the performance of the methods in the online mode. Reported is the averaged balanced accuracies from all datasets of all subjects obtained from using the removal methods when applied to the whole data points in a single dataset and when applied to the EEG windows with different lengths. The results obtainecd from the batch application of ICA and also Noop are also reported. As shown in Fig. 2, the infomax algorithm outperform the fastICA when applied to the whole dataset. However, application of the methods to the short windows shows different results. The performances of both methods are below or not significantly better than that of Noop. This effect is more clearly visible for the infomax method, since for windows of less than 10 seconds the performance is almost 1% and 2% less compared with the results of the Noop and batch infomax, respectively. Although for the online version of fastICA the performance is close to that of the batch application of the method, this is not the case for infomax. Even with 80s of data, i.e., $n_s \approx 1.95n^2$ the performance is almost 1% below that of the batch infomax and close to the Noop performance. That means in this case removing the artifacts using online ICA based methods does not improve the performance considerably in single trial ERP classification.

The average waveforms of the targets and non-targets of one of the datasets are illustrated in Fig. 4. The outputs of different artifact removal approaches plus the ERP windows of the raw data from Fp1 and Pz electrodes are illustrated. For the online methods the outputs of the case that 1s windows were used are shown. As it can be seen, the effects of the removal methods are different for different electrodes. Batch methods have the largest impact on the waveforms.



(a) Fp1 electrode



(b) Pz electrode

Figure 4: Averaged targets and non-targets obtained from the application of different methods to a single dataset. The solid and dotted lines represent the targets and non-targets, respectively.

5 Conclusion

Two ICA-based eye artifact removal methods were compared in batch and online scenarios. The motivation behind using the latter approach is the potential to use it in online scenarios in which the data is buffered and after removing the artifacts will be used for ERP detection. The infomax and fastICA methods were used to decompose the EEG data into ICs and the ADJUST method was utilized for identification of artifact ICs. The methods were evaluated in terms of obtained single trial ERP classification performances.

Experiments show that infomax outperforms the fastICA method in batch mode artifact removal. However, when applying the methods to the windows from stream of data the outcome is different. The ICA methods are sensitive to the data length, in the sense that with very short data lengths the algorithm will not converge to the optimum solution unless the dimensions of the data is reduced.

In general, comparing the performances of the methods for different window sizes show that the online algorithms do not improve the performance over Noop. It is also worth to mention that considering the computational costs of the ICA methods, using very long windows is not practical.

When talking about removing artifacts one has to take into account the recording paradigm and the conditions under which the data were recorded. The datasets that we used here were recorded in a controlled condition, i.e., the correlated artifacts (eye artifacts correlated with the task) were minimized. However, blinks and uncorrelated artifacts exist in the data, which makes the recorded data a good candidate for evaluating the artifact removal methods.

Acknowledgement

This work was supported by the German Bundesministerium für Wirtschaft und Technologie (BMWi, grant FKZ 50 RA 1012 and grant FKZ 50 RA 1011).

References

[1] S. Sanei, *Adaptive Processing of Brain Signals*, Wiley, 2013.

[2] F. Ghaderi, K. Nazarpour, J.G. McWhirter, and S. Sanei, "Removal of ballistocardiogram artifacts using the cyclostationary source extraction method," *IEEE Trans Biomed Eng*, vol. 57, no. 11, pp. 2667–2676, 2010.

[3] F. Ghaderi, S. K. Kim, and E. A. Kirchner, "Effects of eye artifact removal methods on single trial P300 detection, a comparative study," *J NEUROSCI METH*, vol. 221, no. 0, pp. 41–47, 2014.

[4] H. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley-Interscience, 2001.

[5] A. Mognon, J Jovicich, L Bruzzone, and M Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, no. 2, pp. 229–240, 2011.

[6] S. Halder, M. Bensch, J. Mellinger, M. Bogdan, A. Kübler, N. Birbaumer, and W. Rosenstiel, "Online artifact removal for brain-computer interfaces using support vector machines and blind source separation," *Intell. Neuroscience*, vol. 2007, pp. 8:1–8:9, Apr. 2007.

[7] J. Gao, C. Zheng, and P. Wang, "Online removal of muscle artifact from electroencephalogram signals based on canonical correlation analysis," *Clin EEG Neurosci*, vol. 41, no. 1, pp. 53–59, 2010.

[8] F. Matsusaki, T. Ikuno, Y. Katayama, and K. Iramina, "Online artifact removal in EEG signals," in *World Congress on Medical Physics and Biomedical Engineering May 26-31, 2012, Beijing, China*, Mian Long, Ed., vol. 39 of *IFMBE Proceedings*, pp. 352–355. Springer Berlin Heidelberg, 2013.

[9] J. Särelä and R. Vigário, "Overlearning in marginal distribution-based ICA: Analysis and solutions," *J. Mach. Learn. Res.*, vol. 4, no. 7-8, pp. 1447–1469, Oct. 2004.

[10] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[11] J. Onton and S. Makeig, "Information-based modeling of event-related brain dynamics," in *Event-Related Dynamics of Brain Oscillations*, Christa Neuper and Wolfgang Klimesch, Eds., vol. 159 of *Progress in Brain Research*, pp. 99 – 120. Elsevier, 2006.

[12] A. Hyvärinen, J. Särelä, and R. Vigário, "Bumps and Spikes: Artefacts Generated by Independent Component Analysis with Insufficient Sample Size," in *Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99)*, 1999, pp. 425–429.

[13] G. Korats, S. Cam, R. Ranta, and M. Hamid, "Applying ICA in EEG: Choice of the window length and of the decorrelation method," in *Biomedical Engineering Systems and Technologies*, vol. 357 of *Communications in Computer and Information Science*, pp. 269–286. Springer Berlin Heidelberg, 2013.

[14] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[15] C. Chih-Chung and L. Chih-Jen, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[16] M. M. Krell, S. Straube, A. Seeland, H. Wöhrle, J. Teiwes, J. H. Metzen, E. A. Kirchner, and F. Kirchner, "pySPACE - a signal processing and classification environment in Python," *Frontiers in Neuroinformatics*, vol. 7, no. 40, Dec 2013.

[17] S. Straube and M. M. Krell, "How to evaluate an agent's behaviour to infrequent events? – reliable performance estimation insensitive to class distribution," *Frontiers in Computational Neuroscience*, vol. 8, no. 43, 2014.