Effects of eye artifact removal methods on single trial P300 detection, a comparative study

Foad Ghaderi, Su Kyoung Kim, Elsa Andrea Kirchner

Robotics Group, University of Bremen, Bremen, Germany

Robotics Innovation Center, German Research Center for Artificial Intelligence (DFKI GmbH), Bremen, Germany

Abstract

Electroencephalographic signals are commonly contaminated by eye artifacts, even if recorded under controlled conditions. The objective of this work was to quantitatively compare standard artifact removal methods (regression, filtered regression, Infomax, and second order blind identification (SOBI)) and two artifact identification approaches for independent component analysis (ICA) methods, i.e. ADJUST and correlation . To this end, eye artifacts were removed and the cleaned datasets were used for single trial classification of P300 (a type of event related potentials elicited using the oddball paradigm). Statistical analysis of the results confirms that the combination of Infomax and ADJUST provides a relatively better performance (0.6% improvement on average of all subject) while the combination of SOBI and correlation performs the worst. Low-pass filtering the data at lower cutoffs (here 4 Hz) can also improve the classification accuracy. Without requiring any artifact reference channel, the combination of Infomax and ADJUST improves the classification performance more than the other methods for both examined filtering cutoffs, i.e., 4 Hz and 25 Hz.

Keywords: Electroencephalogram, eye artifact removal, regression, Infomax, SOBI, ADJUST.

1. Introduction

Because of the high temporal resolution and ease of use (compared with the other acquisition techniques), electroencephalographic (EEG) signals have been used widely in brain computer interfaces (BCIs). Among the others, P300 based BCI systems exploit the brain responses to infrequent stimuli interspersed with frequent stimuli. The essential task of such BCIs is online detection of P300. However, EEG data recorded for clinical, research or BCI purposes are always contaminated by different artifacts, which can drastically affect further analysis results. The artifacts can be divided into two general categories of non-physiological and physiological artifacts (Fatourechi et al., 2007). Non-physiological artifacts include the power-line interference, noise in the environment, changes in the electrode locations and impedances. On the other hand, physiological artifacts originate from sources inside the body of the subjects, e.g., electromyogram (EMG) and electrocardiogram (ECG).

A major source of artifacts in EEG and event related potentials (ERPs) is the eye activity. Blinks and eye movements are in most cases unavoidable during the recording sessions. The frequency content and the amplitude of the eye artifacts vary for different subjects and tasks (Gratton, 1998), and on different recording electrodes. The amplitude of the artifacts in frontal electrodes can reach to several hundred microvolts which is much higher than that of EEG (typically less than 50 microvolts) (Gratton, 1998). Therefore, removing the eye artifacts is an essential step before further processing the EEG signals.

The most common opinion about the origin of eye electrical activities, known as electrooculogram (EOG), is that they are generated as the result of the difference in electric potential between the cornea and the retina. It is believed that the eye dipole is mostly due to the polarization of the retina (Berg, 1989) and that the electric voltage at the eye is the result of eyelid movements over the eyeball, even when there is no rotation of the eyeball (Croft and Barry, 2000). The electric field generated by each eyeball is considered to be a dipole field with almost constant amplitude. Such a dipole has different effects on each recording electrode. The effects depend on the changes in the orientation of the eyeball (eye movements), the location of the electrode on the scalp, and also changes in the propagation path of the electric field across the head (e.g., blinks change the path through which the electric field propagates to the surface of the head) (Gratton, 1998).

Different methods have been proposed so far to remove the eye artifacts. The simplest method, which is widely used in neuroscience research, is based on rejection of the segments in which the amplitudes measure higher than a predefined threshold. However, rejecting artifacts does not necessarily mean that the remaining parts are clean. Furthermore, rejection bears the disadvantage that data is lost.

Linear regression is one of the most common EOG artifact removal methods (Gratton et al., 1983; Croft and Barry, 2000; Romero et al., 2009). It is assumed that the artifact component in each EEG lead is a multiplier of the pure electrical eye activity. Using this idea and recording vertical and horizontal EOG channels, the eye artifacts are removed in the least squares sense. It is shown in (Kenemans et al., 1991; Pham et al., 2009) that frequency domain and multi-lag time domain regressions do not provide significant change over the standard time domain regression method. Different adaptive filtering approaches have also been proposed for EOG artifact removal, e.g., LMS (Puthusserypady and Ratnarajah, 2006), RLS (He et al., 2007), H^∞ (Puthusserypady and Ratnarajah, 2006).

Independent component analysis (ICA) methods are a set of algorithms that having mixtures of unknown source signals estimate the underlying sources. In the ICA methods the main assumption is the statistical independence of the sources. Therefore, assuming that the brain signals and the artifacts have independent origins, different ICA based methods have been used to estimate and remove the artifacts from EEG signals (Delorme et al., 2007; Kachenoura et al., 2008; Albera et al., 2012; Nazarpour et al., 2008; Mennes et al., 2010). The effects of the ICA and regression based methods on topographic and spectral distribution of the cleaned data are compared in (Romero et al., 2009; Wallstrom et al., 2004). ICA methods have also been used in BCI systems to separate the sources of interest and hence improve the overall performance. In (Naeem et al., 2006; Winkler and Tangermanna, 2011) the effect of ICA algorithms in motor imagery tasks has been analyzed. Investigations about using ICA algorithms in P300 based BCIs are reported in (Li et al., 2009; Serby et al., 2005).

A major challenge with component based methods is correct identification of the artifacts in the derived independent components (ICs). In some papers e.g., (Lagerlund et al., 1997) the artifact channels are selected by visual inspection which is not a suitable approach for on-line applications. As an alternative, one can record extra EOG channels to compare with the ICs and select those with the highest correlation. For the sake of fully automatic solutions, different temporal, spatial, spectral and statistical properties of the components are exploited to identify artifact channels (Wallstrom et al., 2004; Romero et al., 2009; Mognon et al., 2011). Classification methods have also been used to improve the selection procedure (Halder et al., 2007). Recently a novel method for automatic identification of artifacts is proposed (Mognon et al., 2011). The method, called ADJUST, can identify the artifacted ICs using the artifact-specific spatial and temporal features.

Apart from the research works that focused on developing artifact removal methods, several studies have been reported that compare the methods in different performance aspects, e.g., least squares linear regression (Croft et al., 2005), spectral distortion and time domain mean error (Wallstrom et al., 2004), correlation and transfer function estimation (Kenemans et al., 1991), and signal to interference-plus-noise ratio (Kachenoura et al., 2008). However, to the best of our knowledge, no research on comparison of the effects of eye artifact removal methods on single trial detection of P300 has been published so far .

The objective of this paper is to quantitatively compare the effects of artifact removal methods on the performance of single trial P300 detection. To this end, we used datasets (Tabie and Kirchner, 2013; Kirchner et al., 2013) recorded previously in a controlled condition in order to develop/evaluate single trial EEG analysis methods in the context of BCI control. The main constraint in selecting the removal methods was the possibility of automatic operation and also possible extension to on-line applications. Considering the previous reports on the performance of the ICA methods, we selected Infomax (Hyvärinen et al., 2001) and SOBI (Belouchrani et al., 1997) methods among the others. We also used two regression based methods due to their simple structure and approved performance in removing artifacts. It is also known that the spectral filtering can affect the P300 detection performance (Jansen et al., 2004; Ghaderi, 2013). The interaction of removal methods and filtering cutoff frequencies was investigated by repeating the experiments for data filtered at two different cutoff frequencies. For comparison, the accuracy of single trial P300 classification was reported for all experiments.

2. Methods

2.1. Experimental setup

To evaluate the eye artifact removal methods, a dataset that was recorded under controlled conditions was selected. Here, controlled conditions refers to the setup in which first, all experiments were conducted in a shielding cabin. By doing so, we reduced external, non-physiological noise that would otherwise contaminate the EEG signals. Second, subjects had to perform a well defined motion task. By this, differences between subjects that are common for complex interaction scenarios could be reduced. Third and the most important, subjects were asked to fixate a fixation cross all through the experiment. By asking the subjects to fixate a certain point, the amount of saccades triggered by presenting a visual command that require directed movement was reduced. This was important because we wanted to prevent eye artifacts, especially saccades, from being strongly correlated with the evoked event related activity in the EEG.

Data from six healthy male subjects (age: 26.5 ± 3.8 years; right-handed; normal or corrected-to-normal vision) was used in this study. 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

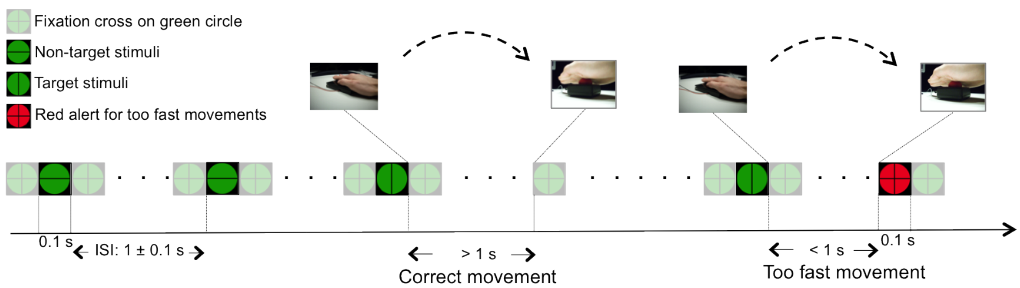


Figure 1: Schema of the experiment. Subjects were asked to move their right hand from a flat micro switch board to a buzzer upon replacement of the fixation cross with a vertical line. No movement was required on the presentation of horizontal lines. For presentation purposes the fixation cross and its green background are slightly faded in this figure. However, in the real experiments this was not the case.

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 keep their eyes fixed on the fixation cross all through the experiment. Schema of the experiments is illustrated in Fig. 1.

After presentation of a target stimulus (visual task-relevant command), the subjects had to perform a slow movement of the right arm between the two input devices. For this, the subjects had to move their hand from the left to the right input device, push the device and return their hand to the left device. Each input device had a micro switch that was used to monitor the begin (left device) and end (right device) of a performed movement. Events that were detected by the devices (pressing/releasing) were marked in the EEG and EOG data. Each movement to the right had to take at least 1 second. Visual commands were given in an oddball fashion. Here, task-irrelevant frequent standard stimuli (change of fixation cross to horizontal line) mixed with infrequent target stimuli (change of fixation cross to vertical line) with a ratio of 8 : 1 and inter stimulus intervals (ISI) of 900 to 1100 ms were presented to the subjects. Duration of stimulus presentation was 100 ms (see Fig. 1).

For each subject we recorded 3 runs. Too early movements or false movements on standard stimuli were reported to the subjects by changing the color of the green circle behind the fixation cross to red for 100 ms. Such wrong movement trials were not used for later data analysis. Each run ended after 40 correctly performed movements. The experiment was designed with Presentation software [Neurobehavioral Systems, Inc., Albany, USA]. 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.

Data Acquisition

EEG and EOG signals were acquired with 5 kHz, filtered between 0.1 to 1 kHz using BrainAmp DC amplifiers [Brain Products GmbH, Munich, Germany] and saved to a computer. A 128-channel actiCap system was used (referenced at FCz). Four of the active electrodes, i.e., I1, OI1h, I2, and OI2h were used to record EOG data. The first pair of electrodes were placed above and below right eye to obtain the vertical eye movements (vEOG). The other pair of electrodes were placed at the left and right outer canthi to record horizontal eye movements. The hEOG and vEOG data were obtained by subtracting the data from the corresponding channels.

2.2. Artifact removal methods

In this section, we first address the effects of spectral filtering on P300 detection. Then theories behind regression, filtered regression, Infomax and SOBI methods are presented. Finally, deflation approach for removing the artifacted ICs and the artifact identification methods (ADJUST and correlation) are introduces.

2.2.1. Filtering effect

EEG channels have to be low-pass filtered before further processing to remove the high frequency noise and artifacts. Selecting appropriate cutoff frequency has an important impact on the overall performance of BCI systems (Ghaderi, 2013). Considering that P300 is mostly associated with activities in the delta band, in (Jansen et al., 2004) it is suggested to use the data in the 0-4 Hz band. However, P300 can be found in other spectral bands as well (Kolev et al., 1997). In P300 classification applications, different low-pass cutoff frequencies are reported, ranging from 4 Hz (Jansen et al., 2004) to 30 Hz (T. Kaufmann, 2011). There is no specific guideline for selecting the appropriate spectral filtering band in the literature. Therefore we repeated the experiments with all artifact removal methods for two low-pass cutoff frequencies, i.e., 4 Hz and 25 Hz.

2.2.2. Regression

The assumption that the EOG artifacts are added linearly to EEG data in different channels is widely accepted in the community. Under this assumption regression is known as one of the standard methods for eye artifact removal (Gratton et al., 1983; Kenemans et al., 1991; Croft and Barry, 2000). The governing model is as follows:

$$y_i(t) = x_i(t) + \alpha_i \sum_j r_j(t) \tag{1}$$

where $y_i(t)$ is the recorded EEG signal at electrode i, $x_i(t)$ is the pure EEG signal at the same electrode, each $r_j(t)$ is one of the artifact channels, and α_i is the propagation factor at electrode i. Having the recorded data at each electrode, and also the reference channel for vertical and horizontal EOG, the propagation factors can be estimated using the least squares method (Croft et al., 2005).

2.2.3. Filtered regression

Main concern about removing eye artifacts using regression method is the so called bidirectional contamination, i.e., at the same time that EEG recordings are contaminated by eye activities, the data recorded at the EOG electrodes are contaminated by the brain activity (mostly from frontal and temporal lobes) (Wallstrom et al., 2004). Therefore, removing the EOG artifacts would also remove parts of the EEG data.

In a derivation of the regression method, called filtered regression, the EOG reference channels are low-pass filtered before regression to overcome the bidirectional contamination problem (Gasser et al., 1992; Romero et al., 2009). This idea is based on the studies that show the high frequency components in EOG channels are generated from brain activity (Gasser et al., 1992). Different cutoff frequencies between 6 to 9.5 Hz have been used. A typical value for the cutoff frequency (which we use here) is 7.5 Hz (Romero et al., 2009).

2.2.4. Infomax

This algorithm maximizes the output entropy or information flow of a neural network with nonlinear outputs. Assume $\mathbf{x}$ is the input vector to the artificial neural network and $\mathbf{y}$ is the output vector. The following equation describes the relation between the inputs and the outputs of neurons:

$$y_i = \phi_i(\mathbf{w}_i^T \mathbf{x}) + n_i \tag{2}$$

where ϕ_i is a nonlinear scalar function, $\mathbf{w}_i = [w_{ij}]$ is the weight vector corresponding to the ith output channel, n_i is the additive Gaussian white noise vector at sensor i, and superscript T denotes the transpose of a vector or matrix. The entropy of the output is expressed by

$$H(\mathbf{y}) = H(\phi_1(\mathbf{w}_1^T \mathbf{x}), \ldots, \phi_n(\mathbf{w}_n^T \mathbf{x})) \tag{3}$$

For a typical invertible transformation of the random vector $\mathbf{x}$, i.e., $\mathbf{y} = \mathbf{f}(\mathbf{x})$, the relationship between the entropies of $\mathbf{y}$ and $\mathbf{x}$ can be expressed as

$$H(\mathbf{y}) = H(\mathbf{x}) + E\{\log|\det J\mathbf{f}(\mathbf{x})|\} \tag{4}$$

where $J\mathbf{f}(.)$ is the Jacobian matrix of the function $\mathbf{f}(.)$ (Hyvärinen et al., 2001).

Using (4) and assuming that $\mathbf{y} = \mathbf{f}(\mathbf{x}) = [\phi_1(\mathbf{w}_1^T\mathbf{x}), \ldots, \phi_n(\mathbf{w}_n^T\mathbf{x})]$ denotes the nonlinear function defined by the artificial neural network, the transformation of the entropy in (3) is obtained as

$$H(\mathbf{y}) = H(\mathbf{x}) + E\{\log|\det \frac{\partial \mathbf{f}}{\partial \mathbf{W}}(\mathbf{x})|\} \tag{5}$$

The second term in the right hand side of (5) can be easily derived and simplified as follows:

$$E\{\log|\det \frac{\partial \mathbf{f}}{\partial \mathbf{W}}(\mathbf{x})|\} = \sum_i E\{\log \phi_i'(\mathbf{w}_i^T\mathbf{x})\} + \log|\det \mathbf{W}| \tag{6}$$

The $\mathbf{w}_i$ vectors are estimated using the stochastic gradient of the log-likelihood expression in (6) and using appropriate non-linear functions ϕ_i, (Hyvärinen et al., 2001).

2.2.5. SOBI

If $\mathbf{x}$(t) is the input vector, and $\mathbf{z}$(t) $=\mathbf{W}\mathbf{x}$(t) is the estimate of underlying sources, the SOBI method estimates $\mathbf{W}$ by simultaneously diagonalizing the covariance matrices of $\mathbf{z}$(t) calculated at different time delays, i.e., $\mathbf{C}_\tau^z = E\{\mathbf{z}(t)\mathbf{z}(t-\tau)\}$, where τ is a typical time delay (Belouchrani et al., 1997). The objective of the method is to minimize the value of the following cost function:

$$\mathcal{J}(\mathbf{W}) = \sum_{\tau \in \mathcal{T}} \mathrm{off}(\mathbf{W}\mathbf{C}_\tau^z\mathbf{W}^T) \tag{7}$$

where $\mathcal{T}$ is the set of delays and off(.) is the sum of squared off diagonal elements. Minimizer of $\mathcal{J}(\mathbf{W})$ is found using an extension of the Jacobi method (Belouchrani et al., 1997).

2.2.6. Deflating independent components

In the ICA based artifact removal methods, the independent component (IC) channels that are more likely to be artifacts are deflated. That is, they are set to zero and the remaining data is projected back to the sensor space. There is no general solution for selecting the appropriate channels, and setting the selection criteria is an application specific task. The need to identify the artifacts in these methods is one of the main challenges of using the ICA methods for artifact removal, specially in applications that are supposed to be automatic. Here we use two artifact identification methods as follows:

Correlation

In some experimental setups there exist reference channels, here hEOG and vEOG channels, that represent some similarities with the artifacts. In such cases, the IC(s) that have the highest correlation with the reference can be selected as an artifact channel and be deflated consequently. The disadvantage of this method is that recording extra channels that contain enough information about the artifacts, is not always practical.

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 (Mognon et al., 2011)).

2.3. Data processing

The analysis was done in offline mode, i.e., datasets were first cleaned using the removal methods and the cleaned data were used for further processing. It is worth to mention that in order to prevent losing the data, we used all the epochs and did not reject any part of the EEG recordings. Data was down-sampled to 100 Hz and high-pass filtered at 0.1 Hz. Continuous data from each dataset was further low-pass filtered at 4 Hz and 25 Hz using 2nd order Butterworth filter. In order to avoid phase shifts, forward-backward filtering technique was used. All artifact removal methods were applied independently to the continuous data of each filtered dataset.

The EEGLAB (Delorme and Makeig, 2004) implementations of the Infomax and SOBI methods were used and for both methods we assumed that the number of ICs is equal to the number of EEG channels. For Infomax, maximum number of training steps and the initial learning rate was set to 512 and $0.00065/log(n)$, respectively, where n is the number of samples. The learning rate is updated during execution of the algorithm. For SOBI, the number of delayed covariance matrices and the convergence threshold were set to $\min(100, m/3)$ and $\frac{1}{100\sqrt{m}}$, respectively, where m is the number of channels. In order to identify and deflate artifact channels, we used ADJUST and correlation measures. The ADJUST method provides a sorted set of 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. For correlation, all IC channels were compared with the recorded EOG channels and the two with the largest (absolute value) correlations were selected as hEOG and vEOG channels. Next, the artifact channels were deflated one by one. This way, three cleaned datasets (1. hEOG deflated, 2. vEOG deflated, and 3. hEOG and vEOG deflated) were obtained.

To evaluate the effect of removal methods on single trial ERP detection, the ERP samples in each cleaned dataset were cut 100-900 ms after the onsets of the stimuli. For each channel, the mean was subtracted and the variance was normalized to one. Feature vectors were generated by concatenating samples form all channels and support vector machines (SVM) classifier (C-SVC with linear kernel and complexity equal to 1) was used to discriminate the two ERP classes. We used LIBSVM (Chih-Chung and Chih-Jen, 2011) implementation of the classifier. ERP samples in each dataset were randomly divided into 3 separate splits, i.e., roughly $\frac{40}{3}$ targets and at least $\frac{320}{3}$ non-targets in each split (the experiment in each run was continued until 40 valid movements were captured). The methods were cross-validated using the leave-one-out technique. This procedure was repeated for 30 independent evaluations, i.e., for each evaluation, ERP samples in each dataset were randomly divided into three groups with identical number of targets and non-targets in each group. Results of the experiments are reported in terms of true positive rate (TPR), true negative rate (TNR), and balanced accuracy (BA) of classification. BA is the average of TPR and TNR and therefore unaffected by unbalanced class distributions.

For each filtering frequency (4 Hz and 25 Hz) the results of seven different approaches are reported, i.e., Noop (no artifact removal applied), Reg (regression), F_Reg (filtered regression), A_Info (Infomax and ADJUST), C_Info (Infomax and correlation), A_SOBI (SOBI and ADJUST), and C_SOBI (SOBI and correlation). For those cases that correlation coefficients were used for artifact identification we only used the best result from three cleaned datasets.

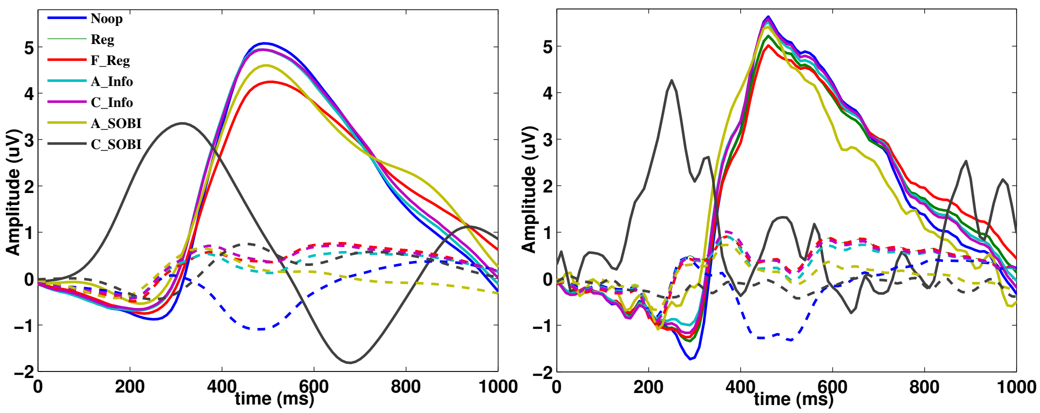
2.4. Statistical analysis

The classification performances obtained from artifact removal methods were statistically analyzed in two steps. First, in order to find the best method, the results were analyzed by repeated measures ANOVA with three within-subjects factors: a) filter types (2 levels: 4 Hz, 25 Hz), b) removal method types (7 levels), and c) subjects (6 levels). Next, only the best artifact removal method (based on the results of the first analysis) was compared with the baseline (Noop). To this end, the classification performances were analyzed by repeated measures ANOVA with filter types (2 levels), removal methods (2 levels: A_Info, and Noop), and subjects (6 levels) as within-subjects factors. If necessary, the Greenhouse-Geisser correction was performed and the corrected p-values were reported. For multiple comparisons, the Bonferroni correction was applied.

3. Results

Figure 2 depicts the grand average ERPs at electrode Pz evaluated separately for each removal method at the two low-pass cutoff frequencies, i.e., 4Hz and 25Hz. Illustrated are the averages of ERPs from all subjects after correcting the baseline (subtracting mean of 100-0 ms prior to stimulus onset). The C_SOBI method has changed the ERP waveforms, in a way that both amplitude and latency of the P300 are considerably different from those of the outputs of the other methods. Table 1 shows the classification performances obtained by cleaning the data using different artifact removal methods and the baseline (no artifact removal method used). The true positive and negative rates, and also the balanced accuracies are reported.

The A_Info was the only approach that performed better than Noop [main effect of *method type*: $F(6,534) = 11876.64$, $p < 0.001$, multiple comparisons: Noop vs. A_Info: $p < 0.001$]. The C_Info and C_SOBI methods showed lower classification performances compared with Noop [Noop vs. C_Info: $p < 0.001$, Noop vs. C_SOBI: $p < 0.001$]. Other methods, Reg, F_Reg, and A_SOBI were not significantly better than Noop [Noop vs. other remaining methods: $p = n.s.$]. This pattern was observed for both filter types [interaction between *filter type* and *method type*: $F(6,534) = 167.08$, $p < 0.001$, multiple comparisons: Noop vs.



(a) Data low-pass filtered at 4 Hz

(b) Data low-pass filtered at 25 Hz

Figure 2: Grand average ERPs at electrode Pz extracted from data cleaned using different EOG artifact removal methods. Average ERPs were calculated over all subjects. Baseline was corrected by subtracting mean of 100-0 ms prior to stimulus onset. Solid line and dashed lines show the targets and non-targets, respectively.

A_Info: $p < 0.05$ for $4Hz$, Noop vs. A_Info: $p < 0.004$ for $25Hz$, Noop vs. C_Info: $p < 0.001$, Noop vs. C_SOBI: $p < 0.001$ for both filter types, Noop vs. all other method types: $p = n.s.$ for both filter types].

Analyzing the results, it is confirmed that higher classification performances were achieved by filtering the data at $4Hz$ compared with filtering at $25Hz$ [main effect of *filter type*: $F(1,89) = 1686.51$, $p < 0.001$, pairwise comparisons: $4Hz$. vs. $25Hz$: $p < 0.001$]. Such higher classification performance was observed for all method types [interaction between *filter type* and *method type*: $F(6,534) = 167.08$, $4Hz$. vs. $25Hz$: $p < 0.001$ for all method types] and for all subjects [interaction between *filter type* and *subject*: $F(5,445) = 58.80$, $p < 0.001$, $4Hz$. vs. $25Hz$: $p < 0.001$ for all subjects]. However, not all subjects showed a higher classification performance for data filtered at 4 Hz for all method types [interaction of *filter type* with *method type* and *subject*: $F(30,2670) = 50.67$, $p < 0.001$, multiple comparisons, see below]. Subject 3, 4, 5, and 6 showed this pattern for all method types [$4Hz$. vs. $25Hz$: $p < 0.001$ for Subject 3, 4, and 5; $p < 0.006$ for Subject 6]. However, Subject 1 and Subject 2 showed a higher classification performance with the data filtered below $4Hz$ for some method types [Subject 1: $4Hz$. vs. $25Hz$: $p < 0.001$ for A_Info, C_Info, A_SOBI, otherwise: $p = n.s.$; Subject 2: $4Hz$. vs. $25Hz$: $p = n.s.$ for C_Info, otherwise: $p < 0.001$]. Also not all subjects showed a higher classification performance for A_Info compared to

Table 1: Classification performance for 6 different artifact removal approaches including the baseline (no artifact removal method applied). The results are reported for each filter type (4 Hz, 25 Hz) separately and averaged over 6 subjects, 3 splits, and 30 evaluations. Reported is averages and standard deviations of true positive, and true negative rates and the balanced accuracies for different artifact removal methods.

		Artifact removal method						
		Noop	Reg	F_Reg	A_Info	C_Info	A_SOBI	C_SOBI
4 Hz	TPR	0.893 ± 0.035	0.884 ± 0.036	0.883 ± 0.036	0.895 ± 0.035	0.672 ± 0.037	0.894 ± 0.035	0.502 ± 0.068
	TNR	0.985 ± 0.002	0.992 ± 0.002	0.993 ± 0.003	0.993 ± 0.002	0.981 ± 0.003	0.992 ± 0.003	0.965 ± 0.017
25 Hz	TPR	0.845 ± 0.034	0.840 ± 0.035	0.838 ± 0.036	0.852 ± 0.035	0.687 ± 0.034	0.845 ± 0.036	0.372 ± 0.035
	TNR	0.986 ± 0.003	0.990 ± 0.002	0.991 ± 0.002	0.992 ± 0.002	0.98 ± 0.002	0.99 ± 0.002	0.978 ± 0.003
BA	4Hz	0.938 ± 0.019	0.938 ± 0.019	0.938 ± 0.019	$\mathbf{0.944 \pm 0.019}$	0.826 ± 0.019	0.943 ± 0.019	0.733 ± 0.035
	25Hz	0.915 ± 0.019	0.915 ± 0.019	0.915 ± 0.019	$\mathbf{0.921 \pm 0.019}$	0.833 ± 0.019	0.917 ± 0.019	0.674 ± 0.020

Noop for all filter types [interaction of *method type* with *filter type* and *subject*: $F(30,2670) = 50.67$, $p < 0.001$, Noop vs. A_Info: $p < 0.022$ for Subject 1 and Subject 3; otherwise: $p = n.s.$].

For more in depth investigations, only the results of A_Info method (which performed better than Noop) were further analyzed. Figure 3 illustrates the classification performances of A_Info and Noop for two filter types and six subjects. A higher classification performance was obtained for A_Info compared to Noop [main effect of *method type*: $F(6,534) = 11876.64, p < 0.001$, A_Info vs. Noop: $p < 0.001$]. This pattern can be observed for both filter types [interaction between *method type* and *filter type*: $F(1,89) = 0.645$, $p = 0.424$, pairwise comparisons: A_Info vs. Noop: $p < 0.003$ for $4Hz$, A_Info vs. Noop: $p < 0.001$ for $25Hz$].

Higher classification performance was achieved for the data filtered below $4Hz$ compared to the data filtered below $25Hz$ [main effect of *filter type*: $F(1,89) = 1686.56, p < 0.001$, pairwise comparisons: $4Hz$ vs. $25Hz$: $p < 0.001$]. This pattern could be observed for both method types [interaction between *method type* and *filter type*: $F(1,89) = 0.645$, $p = n.s.$, pairwise comparisons: $4Hz$ vs. $25Hz$ for A_Info and Noop, respectively]. However, only one subject (Subject 1) did not show such a filter-specific difference [$4Hz$ vs. $25Hz$: $p = n.s.$ for Noop and $p = n.s.$ for A_Info, all other remaining subjects: $p < 0.001$ for both methods]. Also not all subjects showed a higher classification performance for A_Info compared to Noop [interaction of *method type* with *filter type* and *subject*: $F(5,445) = 2.38$, $p < 0.047$, multiple comparisons, see Fig. 3]. Figure 3 shows that the higher classification performance was not consistently obtained by the A_Info for all subjects.

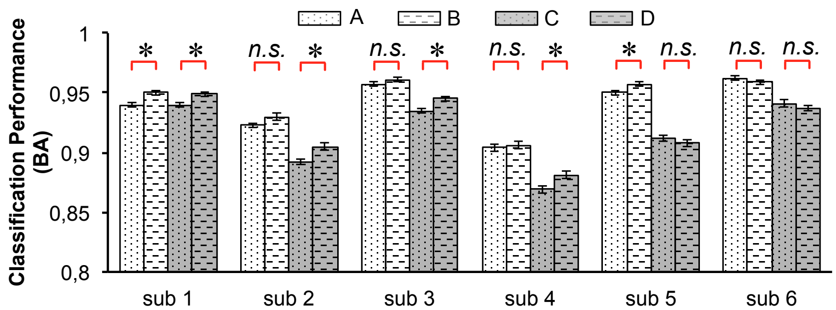


Figure 3: Classification performances for different subjects obtained from Noop and A_Info applied to the data filtered at 4 Hz and 25 Hz. Mean classification performance and standard error of mean (SEM) for each subject, each filter, and each method are depicted (*=p <0.05; n.s.=p >0.05). A: (4 Hz, Noop), B: (4 Hz, A_Info), C: (25 Hz, Noop), B: (25 Hz, A_Info).

4. Concluding discussions

Evaluating the effects of standard artifact removal methods on single trial P300 classification indicates that using the combination of Infomax and ADJUST algorithms leads to a relatively better performance compared with the other approaches. The Infomax algorithm is based on maximizing the joint entropy between the source signal estimates and has approved performance for separating biomedical signals. In comparison with regression and filtered regression, the better performance of Infomax is the result of taking advantage of multi-channel analysis and the assumption that the underlying sources are independent. Considering the theory behind the regression based methods, high quality reference data is always required. However, because of the bidirectional contamination problem, the recorded EOG channels also contain some sort of brain activity.

Successful applications of the SOBI method in different fields have been reported so far, however the combination of this method with correlation was not successful in separating the eye artifacts from EEG signals without distorting the sources of interest. In contrast to Infomax method that assumes underlying sources are statistically independent, the SOBI method is based on joint diagonalization of a number of delayed covariance matrices. The lower performance of this approach (compared with Infomax) indicates that exploiting the time structure of the data might not be suitable for separating the underlying sources in EEG recordings, especially when correlation is used to identify the IC channels. Also, it might be possible that the time delays used for calculating covariance matrices are not selected properly. Although from the grand averages in Fig. 2 it is not possible to infer about the the effects of the removal methods on the single trial ERP waveforms, the worst approach (SOBI, correlation) has clearly affected the signals. These results are in line with the previous research on application of ICA methods in BCI systems (Naeem et al., 2006).

The ADJUST method and correlation with the recorded vEOG and hEOG channels were used for identifying the artifacted IC channels. Results confirm that ADJUST outperforms the correlation coefficients, both for SOBI and Infomax methods. ADJUST exploits temporal and spatial properties of the components and therefore is more suitable than temporal correlation coefficients for artifact identification. Considering the contamination of EOG channels with brain activity, correlation based artifact identification suffers from the same problem as regression based methods do. Furthermore, acquisition of EOG channels usually requires extra effort during recording sessions and is not a comfortable experience for subjects.

The objective of the current study was not to deeply investigate the effects of different cutoff frequencies on P300 detection performance. However, the results show that low-pass filtering the data at 4 Hz results in a better performance compared with filtering at 25 Hz for all artifact removal methods ($p < 0.001$) except for one subject. This is because the main frequency components of P300 are in lower ranges (Jansen et al., 2004). On the other hand, filtering at 4 Hz removes significant parts of non-physiological and physiological artifacts in the data.

The data used in this study was recorded in a controlled condition. The subjects were asked to perform defined hand movements and to fixate on a certain point on the screen. This way, arbitrary eye movements were avoided to a high degree. Even under such a controlled condition, eye artifacts are still induced. Our analysis confirms that using ICA for removing the artifacts induced under such condition can improve the accuracy of single trial P300 classification. However, the absolute values of performance improvements were not so big here. This can be the result of controlled condition in the experiments, which yields high classification performance even without removing the artifacts. Artifact removal methods can increase the classification performance if the artifacts are not correlated with the tasks. In other words, depending on the behavior of the subjects, the eye movements can be correlated with the evoked event related activity in the EEG. In such cases, the classifiers would easily learn the eye artifacts properties because of their high amplitude. Therefore, cleaning the artifacts may eventually decrease the classification accuracy.

Movement related potentials, directly generated by brain, can affect P300 signals. In other words, P300 can be correlated with movement. High classification performances in our study might be the result of correlation between P300 and movement potentials. However, this does not violate the artifact removal evaluation procedure. Movement related potentials and P300 are brain potentials, but EOG artifacts are related to eye movements. Here, we showed that if eye artifacts are not correlated with tasks, removing the artifacts can improve the performance even in systems with high classification rates. It is also very unlikely that movement artifacts were classified in our experiments, because the EMG artifacts are maximal at frequencies above 30 Hz (Fatourechi et al., 2007). In our experiments these components were filtered out using 4 Hz and 25 Hz low-pass filters. Obtaining higher performances from the data filtered at 4 Hz is another clue to conclude that EMG artifacts were probably not used by the classifier.

Because of the uncorrelated eye and hand movements in our experiments, the effectiveness of the artifact removal methods does not depend on the hand movements of the subjects. The hand movements can improve the overall classification performance, but not the artifact removal performance. In other words, if there is no movement in a paradigm, e.g., BCI applications, similar artifact removal performance patterns are expected.

The follow up of this work will focus on artifact removal methods for on-line single trial P300 detection systems. The main limitation of using Infomax in such applications is the high computational costs of the algorithm. However, applying the method to short EEG windows can reduce computational costs significantly. Adaptive ICA algorithms can also be used to overcome this problem.

References

Albera, L., Kachenoura, A., Comon, P., Karfoul, A., Wendling, F., Senhadji, L., Merlet, I., 2012. ICA-based EEG denoising: a comparative analysis of fifteen methods. Bulletin of the Polish Academy of Sciences: Technical Sciences 60 (3), 407–418.

Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., Moulines, E., Feb. 1997. A blind source separation technique using second-order statistics. IEEE Transactions on Signal Processing 45 (2), 434–444.

Berg, P., 1989. Comments on EOG correction methods. J Psychophysiology 3, 41–44.

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

Croft, R., Barry, R., 2000. Removal of ocular artifact from the EEG: a review. Neurophysiologie Clinique/Clinical Neurophysiology 30 (1), 5–19.

Croft, R. J., Chandler, J. S., Barry, R. J., Cooper, N. R., Clarke, A. R., 2005. EOG correction: A comparison of four methods. Psychophysiology 42 (1), 16–24.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods 134 (1), 9–21.

Delorme, A., Sejnowski, T., Makeig, S., 2007. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. NeuroImage 34 (4), 1443–1449.

Fatourechi, M., Bashashati, A., Ward, R. K., Birch, G. E., 2007. EMG and EOG artifacts in brain computer interface systems: A survey. Clinical Neurophysiology 118 (3), 480–494.

Gasser, T., Ziegler, P., Gattaz, W., 1992. The deleterious effect of ocular artefacts on the quantitative EEG, and a remedy. European Archives of Psychiatry and Clinical Neuroscience 241, 352–356.

Ghaderi, F., 2013. Joint spatial and spectral filter estimation for single trial detection of event-related potentials. In: Proceedings of the IEEE international workshop on machine learning for signal processing, MLSP.

Gratton, G., 1998. Dealing with artifacts: The EOG contamination of the event-related brain potential. Behavior Research Methods 30, 44–53.

Gratton, G., Coles, M. G., Donchin, E., Apr. 1983. A new method for off-line removal of ocular artifact. Electroencephalography and clinical neurophysiology 55 (4), 468–484.

Halder, S., Bensch, M., Mellinger, J., Bogdan, M., Kübler, A., Birbaumer, N., Rosenstiel, W., April 2007. Online artifact removal for brain-computer interfaces using support vector machines and blind source separation. Computational Intelligence and Neuroscience 2007, 8:1–8:9.

He, P., Wilson, G., Russell, C., Gerschutz, M., 2007. Removal of ocular artifacts from the EEG: a comparison between time-domain regression method and adaptive filtering method using simulated data. Medical and Biological Engineering and Computing 45, 495–503.

Hyvärinen, H., Karhunen, J., Oja, E., 2001. Independent Component Analysis. Wiley-Interscience.

Jansen, B., Allam, A., Kota, P., Lachance, K., Osho, A., Sundaresan, K., 2004. An exploratory study of factors affecting single trial P300 detection. Biomedical Engineering, IEEE Transactions on 51 (6), 975–978.

Kachenoura, A., Albera, L., Senhadji, L., Comon, P., 2008. ICA: a potential tool for BCI systems. Signal Processing Magazine, IEEE 25 (1), 57–68.

Kenemans, J. L., Molenaar, P. C., Verbaten, M. N., Slangen, J. L., 1991. Removal of the ocular artifact from the EEG: A comparison of time and frequency domain methods with simulated and real data. Psychophysiology 28 (1), 114–121.

Kirchner, E. A., Albiez, J. C., Seeland, A., Jordan, M., Kirchner, F., 2013. Towards assistive robotics for home rehabilitation. In: Proceedings of the International Conference on Biomedical Electronics and Devices. SciTePress, pp. 168–177.

Kolev, V., Demiralp, T., Yordanova, J., Ademoglu, A., Isoglu-Alka, Ü., May 1997. Time-frequency analysis reveals multiple functional components during oddball P300. NeuroReport 8 (8), 2061–2065.

Lagerlund, T. D., Sharbrough, F. W., Busacker, N. E., 1997. Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition. J Clin Neurophysiol 14, 73–82.

Li, K., Sankar, R., Arbel, Y., Donchin, E., 2009. Single trial independent component analysis for P300 BCI system. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC. pp. 4035–4038.

Mennes, M., Wouters, H., Vanrumste, B., Lagae, L., Stiers, P., 2010. Validation of ICA as a tool to remove eye movement artifacts from EEG/ERP. Psychophysiology 47 (6), 1142–1150.

Mognon, A., Jovicich, J., Bruzzone, L., Buiatti, M., 2011. ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. Psychophysiology 48 (2), 229–240.

Naeem, M., Brunner, C., Leeb, R., Graimann, B., Pfurtscheller, G., 2006. Seperability of four-class motor imagery data using independent components analysis. Journal of Neural Engineering 3 (3), 208.

Nazarpour, K., Mohseni, H. R., Hesse, C. W., Chambers, J. A., Sanei, S., January 2008. A novel semiblind signal extraction approach for the removal of eye-blink artifact from EEGs. EURASIP J. Adv. Signal Process 2008, 98:1–98:12.

Pham, T. T. H., Croft, R. J., Cadusch, P. J., 2009. Is ocular voltage propagation to the electroencephalogram frequency dependent? Psychophysiology 46 (5), 949–956.

Puthusserypady, S., Ratnarajah, T., 2006. Robust adaptive techniques for minimization of EOG artefacts from EEG signals. Signal Processing 86 (9), 2351–2363.

Romero, S., Maanas, M., Barbanoj, M., 2009. Ocular reduction in EEG signals based on adaptive filtering, regression and blind source separation. Annals of Biomedical Engineering 37, 176–191.

Serby, H., Yom-Tov, E., Inbar, G., 2005. An improved P300-based brain-computer interface. Neural Systems and Rehabilitation Engineering, IEEE Transactions on 13 (1), 89–98.

T. Kaufmann, E.M. Hammer, A. K., 2011. ERPs contributing to classification in the P300 BCI. In: 5th International BCI Conference 2011. pp. 136–139.

Tabie, M., Kirchner, E. A., 2013. EMG onset detection - comparison of different methods for a movement prediction task based on EMG. In: Proceedings of the International Conference on Bio-inspired Systems and Signal Processing. SciTePress, pp. 242–247.

Wallstrom, G. L., Kass, R. E., Miller, A., Cohn, J. F., Fox, N. A., 2004. Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods. International Journal of Psychophysiology 53 (2), 105–119.

Winkler, I., Tangermanna, M., 2011. Artifact-insensitivity of CSP in motor imagery BCI. International Journal of Bioelectromagnetism 13 (2), 72–73.