Automated ocular artifact removal: comparing regression and component-based methods

Author(s): A.Schlögl(1), A. Ziehe(2), K.-R. Müller(3)

Affiliation:
(1) Graz University of Technology, Graz, Austria
(2) Fraunhofer FIRST, Berlin, Germany
(3) TU Berlin, Computer Science, Berlin Germany.

Address:
Institute for Human-Computer Interfaces, Graz University of Technology,
Krenngasse 37/EG, A-8010 Graz, Austria
e-mail: alois.schloegl@tugraz.at

Abstract

Objective: The aim is to compare various fully automated methods for reducing ocular artifacts from EEG recordings.
Methods: Seven automated methods including regression, six component-based methods for reducing ocular artifacts have been applied to 36 data sets from two different labs. The influence of various noise sources is analyzed and the ratio between corrected and uncorrected EEG spectra, has been used to quantify the distortion.
Results: The results show that not only regression but also component-based methods are vulnerable to over- or under-compensation and can cause significant distortion of EEG. Despite common belief, component-based methods did not demonstrate an advantage over the simple regression method.
Conclusion: The newly proposed evaluation criterion showed to be an effective approach to evaluate 252 results from 36 data sets and 7 different methods.
Significance: Currently, the regression method provides the most robust and stable results and is therefore the state-of-the-art-method for fully automated reduction of ocular artifacts.

Keywords: ocular artifact, electroencephalogram, automated artifact processing

Introduction

The electroencephalogram (EEG) can be corrupted by artifacts, this holds for standard offline EEG analysis studies (Nunez, 1993) as well as online decoding of brain signals (e.g. see Dornhege et al. 2007). The importance of reducing ocular artifacts has been emphasized in several areas e.g. evoked potentials by Donchin (1977), sleep analysis by Anderer et al. (1999) and brain-computer interface research by Fatourechi et al. (2007). Ocular artifact processing might be also important for clarifying the relationship between induced gamma band activity and micro-saccades (Yuval-Greenberg et al. 2008, Fries et al. 2008).

Two main lines of research have tackled removal of the electrooculogram (EOG) from EEG recordings: (a) regression (Gratton et al. 1983, Elbert et al 1985, Semlitsch et al. 1986, Wallstrom et al. 2004, Schlögl et al. 2007) and (b) blind source separation (BSS) (Vigário et al. 1997, Jung et al, 1998, 2000, Ziehe et al. 2000) methods. The idea in (a) is to use the known EOG channel activity for simply regressing their contribution out from all EEG channels. Alternative (b) decomposes the EEG and EOG channels to find the true underlying sources of brain and ocular activity blindly, i.e. without imposing strong modeling assumptions except for 'independence' (Vigario et al. 1997, Jung et al, 1998, 2000, Ziehe et al. 2000). Several works (Jung et al. 1998, p.64, 2000, p.1746, Casarotto et al. 2004), Romero et al. 2004, p.925, Shoker et al 2005, Nicolaou and Nasuto 2005, p. 5991, Phlypo et al. 2007) claim that Independent Component Analysis (ICA) is advantageous over the regression method, because the EOG channels might be contaminated by EEG activity, causing an overcompensation and removal of EEG activity. However this argument assumes that EOG components identified by ICA contain EOG activity only and are not contaminated by EEG, which is not necessarily the case.

Blind source separation (BSS) methods like ICA will always provide a decomposition, no matter, whether the underlying solution is good or not (cf. Meinecke et al. 2002). Vorobyov et al. (2002, p.294) recognize that "ICA cannot guarantee that some individual independent components (ICs) contain *only* noise and do not contain information about useful sources, especially in biomedical applications". Another possible problem of ICA is channel noise (e.g. amplifier and impedance noise). The channel noise is another signal source and effectively doubles the number of sources (see Ziehe et al. 2000, Wübbeler et al. 2000) while the number of sensors (i.e. the number of EEG channels) is the same. Because ICA can identify no more sources than sensors, several components will represent a mixture of signal and noise sources.

There is no consensus whether the approach (a) or (b) is to be preferred under exactly what circumstances: while Blind Source Separation (BSS) is generally recognized as one of the leading methods for artifact removal, ~~it is~~ Wallstrom et al. (2004) ~~that~~ find an advantage of regression and Principle Component Analysis (PCA) methods over the ICA approach. In other words there seems to be no unique view in an important practical issue.

This work therefore aims to contribute to the quest for a practical "reference method" in ocular artifact removal by providing an objective numerical comparison between fully automated methods only. While some of the BSS methods could be hand-tuned by manual component selection, such manual influence is often undesirable from an application point of view, because it requires highly trained staff and it may also introduce subjective influence and personal bias into the evaluation.

First, the paper introduces the automatic EOG removal methods (Table 1 shows an overview of the methods used), then provides their results on a large base of EEG data and puts them into

perspective and finally a brief conclusion is given.

Methods

Linear Superposition Model, Regression method

The EEG and EOG artifacts can be considered as a linear superposition of *N* EEG and *M* EOG ~~channels~~components. This can be written in the from of a regression model

$$\vec{Y}_t = \vec{E}_t + b_{N\times M} \cdot \vec{O}_t \quad (1)$$

Accordingly, the observed EEG at time *t* is a vector $\vec{Y}_t$ with *N* elements, and the observed EOG activity $\vec{O}_t$ at time *t* has *M* elements. The observed EEG data $\vec{Y}_t$ consists of a linear superposition of the true EEG activity $\vec{E}_t$ and the ocular activity $\vec{O}_t$ that propagates through the mechanism of volume conduction to each EEG electrode. The propagation factors are described by the model parameters $b_{N\times M}$ which describe the influence of *M* components of the ocular dipoles to each of the *N* EEG channel. Because the propagation mechanism is simple volume conduction (Nunez, 1981; Malmivuo & Plonsey, 1995; Kierkels et al. 2006, p.246), it depends only on the geometry and the conductivity of the head tissue. It is reasonable to assume that~~,~~ these are constant during the whole EEG recording time T $0<t\le T$ and independent of frequency.

It should be noted that the regression model can be also written in the following form

$$\begin{bmatrix}\vec{Y}_t \\ \vec{Z}_t\end{bmatrix} = \begin{bmatrix} I_{N\times N} & b_{N\times M} \\ 0_{M\times N} & I_{M\times M}\end{bmatrix} \cdot \begin{bmatrix}\vec{E}_t \\ \vec{O}_t\end{bmatrix} \quad (2)$$

where~~as~~ $\vec{Z}_t$ represents the observed EOG channels.

If the EOG activity is measured, its contribution can be removed using the least squares solution of equation (1). A ~~left~~right multiplication of equation (1) with $\vec{O}_t^T$ and applying the expectation operator $\langle .{}_t \rangle$ over time *t* yields

$$\langle \vec{Y}_t \cdot \vec{O}_t^T \rangle = \langle \vec{E}_t \cdot \vec{O}_t^T \rangle + b_{N\times M} \cdot \langle \vec{O}_t \cdot \vec{O}_t^T \rangle \quad (3)\,.$$

Because EEG and EOG can be considered uncorrelated, the term $\langle \vec{E}_t \cdot \vec{O}_t^T \rangle$ becomes zero, the true model coefficients are

$$b = \langle \vec{Y}_t \cdot \vec{O}_t^T \rangle \cdot \langle \vec{O}_t \cdot \vec{O}_t^T \rangle^{-1} \quad (4)$$

and the EEG data can be corrected according to

$$\vec{E}_t = \vec{Y}_t - b \times \vec{O}_t \quad (5)$$

The method is also known as “least squares approach” or “multiple least squares approach” , if more than one EOG component is removed.

This model takes into account only EEG and EOG sources. In practice, other noise sources (e.g. amplifier and impedance noise, electric and magnetic interferences, muscle activity) occur, too. In order to analyze the possible influence of these noise sources on the reduction method, a noisy model has to be considered (see Appendix A:Noisy mixture model). One consequence of this analysis is the fact, that the model estimates $\hat{b} = \beta$ are least biased if the signal to noise ratio between EOG and other noise sources is as large as possible. Therefore, we estimated the model coefficients from data with large ocular activity. Furthermore, it can be advantageous to filter the data (e.g. for removing the very low frequency components of the 1/f amplifier noise, and the very high frequency activity).

Linear Mixture Model, blind source separation

Blind source separation (BSS) methods like ICA are based on a linear mixture model of the EEG sources S_t and the ocular activity O_t

$$\begin{bmatrix} \vec{Y}_t \\ \vec{Z}_t \end{bmatrix} = A \cdot \begin{bmatrix} \vec{S}_t \\ \vec{O}_t \end{bmatrix} = \begin{bmatrix} a_{N\times N} & b_{N\times M} \\ c_{M\times N} & d_{M\times M} \end{bmatrix} \cdot \begin{bmatrix} \vec{S}_t \\ \vec{O}_t \end{bmatrix} \quad (6)$$

where A is the mixing matrix, Y_t and Z_t are the observed EEG and EOG channels, respectively. The inverse of the mixing matrix is called the unmixing matrix W

$$W = \begin{bmatrix} u_{N\times N} & v_{N\times M} \\ w_{M\times N} & x_{M\times M} \end{bmatrix} = A^{-1} = \begin{bmatrix} a_{N\times N} & b_{N\times M} \\ c_{M\times N} & d_{M\times M} \end{bmatrix}^{-1} \quad (7)$$

which is used for reconstructing the EEG S_t and EOG O_t components from the observed data.

$$\begin{bmatrix} \vec{S}_t \\ \vec{O}_t \end{bmatrix} = A^{-1} \cdot \begin{bmatrix} \vec{Y}_t \\ \vec{Z}_t \end{bmatrix} = W \cdot \begin{bmatrix} \vec{Y}_t \\ \vec{Z}_t \end{bmatrix} \quad (8)$$

The unmixing matrix W can be obtained by minimizing an algorithm-specific loss function that enforces independence rsp. uncorrelatedness over time (Hyvärinen and Oja, 2000). Accordingly, the eye activity $\vec{O}_t$ is orthogonal to the EEG activity $\vec{S}_t$, thus $\langle \vec{S}_t \cdot \vec{O}_t^T \rangle = \cdot_{N\times M}$. If the separation between EEG and the EOG provides the correct decomposition, $\vec{E}_t$ and $\vec{S}_t$ describe the same signal subspaceand both are uncorrelated to the EOG activity $\vec{O}_t$. The term „signal subspace“ refers to the fact that the components and $\vec{S}_t$ and $\vec{E}_t$ span a smaller “signal space” with-in a

larger one. In the present case, the signal space is spanned by $\begin{bmatrix}\vec{Y}_t \\ \vec{Z}_t\end{bmatrix}$ and $\begin{bmatrix}\vec{S}_t \\ \vec{O}_t\end{bmatrix}$. Removing some components (e.g. $\vec{O}_t$ or $\vec{Z}_t$) forms a smaller signal space (i.e. subspace). Ideally~~n ideal case~~, both $\vec{E}_t$ and $\vec{S}_t$ describe the EEG activity, but are not necessarily the same because the activity can be described on the surface or the source level.

The ICA approach has become very popular through the software package EEGLab (Makeig et al. 1996, Delorme and Makeig, 2004). Because independence implies uncorrelatedness (Hyvärinen and Oja, 2000), the reconstruction of the EEG without the EOG components (equation 9) is equivalent to removing the EOG components through a least squares approach (applying equation (10) to equation (4) and (5)). In both cases the orthogonal EOG components are removed by projecting the recorded data to a signal subspace orthogonal to the subspace described by the EOG components. The corrected EEG activity is

$$\vec{E}_t = a_{N\times N}\cdot\vec{S}_t = a_{N\times N}\cdot\begin{bmatrix}u_{N\times N} & v_{N\times M}\end{bmatrix}\cdot\begin{bmatrix}\vec{Y}_t \\ \vec{Z}_t\end{bmatrix} \quad (9)$$

and can be also obtained by applying

$$\vec{O}_t = \begin{bmatrix}w_{M\times N} & x_{M\times M}\end{bmatrix}\cdot\begin{bmatrix}\vec{Y}_t \\ \vec{Z}_t\end{bmatrix} \quad (10)$$

to the regression method (see equations 4 and 5).

BSS provides a number of independent components that can be as large as the number of EEG channels. Additional criteria are necessary to identify which component is related to ocular activity. Often, the components are hand selected by trained experts by considering time course, frequency content and topography of a component (cf. Ziehe et al. 2000, Wübbeler et al. 2000). The drawback of this approach is that such manual data processing is time consuming and subjective interpretations may be introduced.

Various automatic attempts have been proposed (Barbati et al. 2004, Joyce et al. 2003, Bouzida et al.2005, LeVan et al. 2006, Li et al. 2006., Vorobyov and Cichocki 2002, Kierkels et al. 2006, Boudet et al. 2007, Romero et al. 2008) that help to overcome manual interference which introduces a subjective component into the data analysis and can be labor intensive. Unfortunately some of them did not report crucial parameters like detection thresholds (Barbati et al. 2004, Bouzida et al. 2005, Romero et al. 2008), or required an expert scoring (LeVan et al. 2006). These methods could not be reproduced and were therefore not considered in the current evaluation. The method of Li et al (2006) removed only a single component and is therefore considered inferior to a two-component regression method. The procedures proposed by Vorobyov and Cichocki (2002), Joyce et al (2003) and Kierkels et al. (2006) were described with sufficient detailed information, in order to re-implement the method in a software-algorithm. The method of Vorobyov and Cichocki (2002) mark a component as blink artifact if the Hurst exponent of this component is between 0.58 and 0.64. The Hurst exponent *H* is an "index of dependency" and is closely related the fractal dimension $D = 2-H$. *H* is in the range of zero to one, a higher value indicates a smoother trend, a lower value more roughness. Joyce et al. (2004) use a five-step procedure which includes normalizing each channel to a variance of 1, computing the correlation between the recorded EOG channels and the IC's should be larger than 0.3, and the R.M.S of the first (temporal) derivative should be smaller than 0.2 (at

500 Hz sampling rate); Kierkels et al. (2006) compared several methods, the best method was SOBI (which is equivalent to TDSEP from Ziehe et al. 2000) and selected the components “if the cross correlation between an extracted component and one of the recorded EOG signals exceeds a threshold value [of 0.7]”.

Recently, the Non-Gaussian Component Analysis (NGCA) has been suggested (Blanchard et al. 2006, Kawanabe et al. 2007) as an alternative BSS method. NGCA identifies sources with a non-gaussian distribution, from a linear mixture of gaussian and non-gaussian sources. Based on the assumptions that (i) the EOG within the training set has a non-gaussian distribution, and (ii) there are three EOG components (x-y-z direction), the heuristic of selecting the first three components has been applied.

Optimal projection

The filtering by optimal projection (FOP) approach in Boudet et al. (2007) is closely related to Common Spatial Subspace Decomposition (CSSD) and the Common Spatial patterns (CSP) (Ramoser et al, 2000, Blankertz et al. 2008) and assumes also a linear mixture model. Unlike BSS methods, this approach uses two data sets (one with and one without EOG artifacts) for identifying the components representing artifacts and for identifying desirable components. Once, the EOG components are known, they can be removed with the least squares approach.

According to Boudet et al. (2007), the covariance matrices C_a (from data with artifacts) and C_r (from data without artifacts) were simultaneously orthogonalized such that

$$C_r \cdot P = C_a \cdot P \cdot D \quad (11)$$

or

$$C_a^{-1} \cdot C_r = P \cdot D \cdot P^{-1} \quad (12)$$

with the eigenvector matrix P and a diagonal eigenvalue matrix D. Like in the other methods, it is assumed that the true EEG and EOG are orthogonal (i.e. uncorrelated). According to Boudet et al. (2007), the eigenvectors with an eigenvalue above a threshold of 0.3 indicate an ocular artifact source. The fixed threshold can result in a varying number of components removed. In the present study no more than five components were removed.

The selected eigenvectors were applied to obtain the EOG components according to equation (10) ; then the obtained EOG activity was removed with regression (i.e. least squares) approach according to equations (4) and (5).

Table 1: Automated methods for reducing ocular artifacts.

Method	Short description
Regression (REG) (Schlögl et al. 2007)	Regression analysis using 2 EOG channels for vertical and horizontal components
JOYCE (Joyce et al. 2002)	ICA components are obtained using SOBI/TDSEP algorithm components of from (common average reference) EEG and two bipolar EOG.
HURST (Vorobyov and Cichocki, 2002)	Select the components with an Hurst exponent between 0.58 and 0.64. from (common average reference) EEG and two bipolar EOG. (Fs=250Hz, Filter 0.1-100Hz)
KIERKELS (Kierkels et al. 2006)	ICA components are obtained using SOBI/TDSEP algorithm components of from (common average reference) EEG and two bipolar EOG. If the cross correlation between an extracted component and one of the recorded EOG signals exceeds 0.7, the component is marked as an ocular component and is removed.
FOP (Boudet et al. 2007)	Filter using optimal projection
PCA-3	Three principle components with largest eigenvalues from (common average reference) EEG and two bipolar EOG.
NGCA Blanchard et al. (2006) Kawanabe et al. (2007)	The first three components from (monopolar) EEG and two bipolar EOG were removed.

Data

Two data sets from two different studies on Brain-Computer-Interfaces (BCI) were used. Data set A contains 19 data sets from 10 different subjects and was recorded in the Graz BCI lab; the EEG was amplified by an amplifier from g.tec (Guger Technology, Graz, Austria), and were filtered between 0.5 and 100 Hz and sampled with 250 Hz. Twenty-two EEG channels, three EOG channels, one ECG and one respiration channel was recorded (for more details see also Schlögl et al. 2007). Both data sets contain recordings of resting EEG with eyes closed and eyes open, recording during deliberate eye movements (the subject was asked to rotate the eyes clockwise and counterclockwise around the whole field of vision, and perform eye blinks). This was followed by a number of experimental EEG recordings for a BCI study.

Data set B consisted of 17 data sets (originally 25, but 6 were not complete and in two cases the EOG channels had large drifts) from 15 different subjects and was recorded in the Berlin BCI lab. The data were recorded with a BrainAmp system (BrainProducts GmbH, Munich, Germany), filtered between 0.016 and 250 Hz, and sampled with 1000 Hz; 54 EEG channels and horizontal and vertical EOG were recorded. Each data set contain a so-called “artifact recording” (containing eye movements, muscle activity and resting EEG) and experimental EEG recording containing three different motor imagery tasks, and consecutive BCI recordings with online feedback. The eye movement data was recorded according a protocol where the technician asked 5 times to move the eyes up-center-down-center-up- ... and left-center-right-center-left- with a pacing of approx. 1 second. Furthermore, several eye blinks were recorded. If not stated otherwise, the data of set B was down-sampled to 250 Hz, by averaging 4 consecutive samples into 1. This procedure attenuates all frequency components above 114 Hz by more the 3 dB (i.e. low pass anti-aliasing filter of 114 Hz).

Accordingly, the data contains horizontal and vertical EOG components as well as blink artifacts. Therefore a single component (like for eye blinks) will not be sufficient, but more components are needed. Based on biophysical consideration, it is expected that at least three components (for the three spatial directions) are necessary (Elbert et al. 1985, Schlögl et al. 2007). Additional components might be needed due to the change of the dipole location (Berg and Scherg, 1991). In this study, the eye movement data and the first experimental recording was used. For one method (FOP), the resting EEG during eyes open was used, too. Note that we will provide the data used within this paper online (ftp://ftp.tu-graz.ac.at/pub/bsdl/EEGwithOcularArtifacts/) as a reference for further studies.

Results

Evaluation

The gold standard for validating EOG correction methods is certainly the visual evaluation by experts (cf. Schlögl et al. 2007). In the present study, 252 results were obtained from 36 recordings and seven different methods. Each of these results consisted of about 6 to 10 minutes of corrected EEG with 20 (data set A) and 54 (data set B) channels. Expert scoring for each of these results (ideally by two or more independent experts) would have been a prohibiting factor for the present study. Therefore, an alternative approach has been used.

In other studies the signal-to-noise ratio was used as an evaluation criterion. This requires the "true" activity, which is only available for simulated data (like in Kierkels et al. 2006). But in the present case of real-world data, the "correct" result is not known. Alternatively, the ratio $r_i(f)$ between the spectral density functions of the corrected $E_i(f)$ and the raw (i.e. uncorrected) EEG $Y_i(f)$ of some channel i was used.

$$r_i(f)=\frac{|E_i(f)|^2}{|Y_i(f)|^2} \quad (13)$$

In the ideal case, the ratio $r_i(f)$ should be always smaller than 1, because the EOG component is removed. In practice, the noisy model (see Appendix A: Noisy mixture model), and therefore $E'(f)$ and $Y'(f)$ must be used. The theoretical analysis of the noisy mixture model shows that the noise on the EOG components is propagated to the corrected signal. This can cause an increase in the noise components, and the ratio $r_i'(f)$ can become larger than 1. Moreover, correlated noise on EEG and EOG channels (e.g. through a common external source like line interference, or the noise on the common reference electrode), can modify the ratio. However, this is a minor difficulty, because line interference can be easily identified by its peak at 50/60 Hz, and the noise between EOG and EEG channels is assumed to be uncorrelated. This assumption is certainly true for the regression method, because bipolar EOG channels were used.

$$r_i'(f)=\frac{|E_i'(f)|^2}{|Y_i'(f)|^2} \quad (14)$$

Figure 1 around here

Figure 1 illustrates the meaning of this ratio. Case A indicates a reduction of EOG activity; case B show no change; and in case D the correction procedure adds some noise components to the EEG.

The desirable case is A. The depth of the ratio is related to the relative amount of removed EOG. Case B shows no reduction, either there is no EOG artifact in the data (this would be acceptable) or the method did not remove any EOG. Case C shows a clear overcompensation, in other words, EEG activity is also removed. In several works, the regression method has been attributed to be vulnerable for this phenomenon. Case D describes the case where the EOG is contaminated by some other noise sources (see Appendix A). In such a case, this noise can propagate to the corrected EEG. In practice, one will observe a combination of these scenarios. But even in those cases. it is possible to determine the dominating mechanism.

In the present study, the Welch method (Oppenheim und Schafer, 1989) with 1024 sample (4.1 s) Hann~~ing~~ window has been used to estimated the power spectral densities. Examples of the frequency dependent ratio $r_i(f)$ of the most frontal channel are shown in Figures 2 and 3. Both models, the linear mixture and superposition model, assume that the propagation of the EOG activity does not depend on frequency. Because the EOG is a low frequency activity, the spectral power should become smaller (ratio smaller than one) only on the low frequency range. In the high frequency range, the spectrum should be not influenced (ratio equal 1), if overcompensation occurs (parts of the EEG are removed), the ratio will become smaller than one also for the high frequency range. If the ratio becomes larger than one, it indicates that some signal component has been added during the correction step.

Figure 2 around here.

Figure 3 around here.

The ratio can be computed separately for every EEG channel i . Because of its proximity to the eyes, the most frontal electrode is most sensitive to overcompensation and is also most vulnerable to noise propagation (because it has usually the largest weighting coefficients $\vec{\beta}_i$). Therefore, the spectral ratio for the most frontal electrode (typically Fpz or Fz) is shown. Moreover, electromagnetic interference from the power line was also present in most data sets. This line interference effect can have an arbitrary effect (it can became smaller or larger depending on its correlation with EOG). Therefore, the comparison of methods ignores changes around 50 Hz.

Regression analysis

Figure 3 shows that regression analysis comes close to the desired result: for both data sets A and B, there is a reduction in the low frequency range indicating a clear reduction of EOG activity. This is also an indicator for the spectral content of the EOG activity, the spectral content of the EOG is 0-5 Hz in data set A, and 0.5 to 10 Hz in data set B. In the high frequency range, the ratio is often close to one indicating no under- or overcompensation of EEG activity.

It is notable that the ratio in data set B is often larger than one in the high frequency range. This is contrary to the common belief, that regression analysis removes EEG activity, too. It can be explained by the fact, that non-EOG noise (amplifier noise, impedance noise) is propagated to the corrected EEG channels. Because the correction coefficients can be quite large for the frontal channels, and if the impedances of the EOG channels are large (providing large impedance noise) the noise contribution can be significant.

Method of Joyce et al.

The method shows for data set A a clear reduction in the low frequency range. Unfortunately, the high frequency range is often also affected. The results for data set B show that the ratio r(f) is almost constant, and often smaller than one. This indicates a significant reduction of EEG activity (overcompensation effect).

ICA component identified by the Hurst exponent.

Vorobyov and Cichocki (2002) suggest an ICA based method, where any ICA "algorithm that ensures robust unbiased estimation of the separation matrix can be employed" (Vorobyov and Cichocki, 2002, p. 297). In the present work, we have used the JADE algorithm (Cardoso and Souloumiac, 1993) on 20 PCA components to identify the ICA components. It is suggested that components with a Hurst exponent between 0.58 and 0.64 represent eye blink components. Since the Hurst exponent might depend on the sampling rate and filter settings, data set B was filtered and resampled with the same settings (filter between 0.1 Hz and 100 Hz, sampling rate 250 Hz) as Vorobyov and Cichocki (2002, p. 297). Data set A has was already sampled with 250 Hz and filtered with 0.5 to 100 Hz; the frequency range could not be widened in order to match the other data.

In all 19 cases of data set A, the spectral ratio between corrected and uncorrected data remained one. The reason is the fact the Hurst exponent of all components from data set A was in the range of 0.29 and 0.52; the criterion of a Hurst exponent between 0.58 and 0.64 did not match this data. Therefore, no EOG components were identified, and no data correction was performed.

Quite the opposite happened for data set B. For a large number of recordings, the ratio became significantly smaller than 1 even in the high frequency range. This indicates that significant parts of the EEG have been removed as well, this can be explained by the fact that some components representing EEG activity have also a Hurst exponent in the range from 0.58 to 0.64. Thus, large overcorrection effects happened, i.e. significant EEG activity was removed in several recordings of data set B.

In other words, the proposed heuristic of a Hurst exponent in the range of 0.58 and 0.64 might have been suitable for the data used by Vorobyov and Cichocki (2002), but it is not a suitable heuristic for the data sets used in this work.

Method of Kierkels et al.

Kierkels et al. (2006) used a simple criterion (correlation to any EOG channel larger than 0.7) for identifying the EOG components; regression analysis and five BSS methods (SOBI, JADE, FastICA, PCA, R-ARE) were compared. SOBI was their best method, SOBI is also equivalent to TDSEP (Ziehe et. 2000). The results show overcompensation (reduction of EEG) as well as an increase in alpha EEG . The latter can be explained by propagation of occipital alpha into the EOG channel due to alpha contamination of the selected ICA components.

Kierkels et al. assumed that the noise component is pure electrode noise and has a flat spectrum. However, the channel noise of real-world EEG recordings is also contaminated by amplifier noise which has typically a 1/f spectral distribution. Furthermore, the simulations assumed that the activity of various dipols are uncorrelated, but the global alpha rhythm suggest a coupling between dipoles. SOBI is using these second-order statistical correlations for its decomposition. As a consequence, the identified components are not pure ocular activity (like in the simulated case) but are corrupted by the brain and noise sources in the real-world case.

Filter of optimal projection (FOP)

The FOP method shows for most cases in data set A similar results than the regression method. However one recording of data set A shows a clear reduction of high frequency activity, indicating an removal of EEG activity (overcompensation). For data set B, the method shows only one result with clear reduction of EOG activity (reduction at low frequencies but no reduction at high frequencies). All other 16 data sets of B show either a broad decrease or increase, indicating that either EEG is removed or some noise terms are added to the signal.

Other BSS methods

PCA shows for data set A a significant reduction of EEG activity. For data set B, the PCA method shows a reasonable reduction of EOG activity (low frequency range) and (with a few exceptions) almost no changes in the high frequency range. However, in a few cases the ratio is larger than one in the alpha range (10-15 Hz), indicating a propagation of the occipital alpha into the frontal channel.

The Non-Gaussian Component Analysis (NGCA) is a promising method , because the spectral ratio is for most recordings very similar to the ideal curve (case A in Fig. 1). Despite a few exceptions (alpha propagation in one case of data set A, and a significant noise propagation in a few cases of data set B), NGCA seems to be the most stable and very promising method among the investigated BSS methods.

Table 2: Median and 1st and 3rd quartile of spectral ratio from 38 data sets. Ideally, the spectral ratio should be 1 for the upper frequency bands (8-13, and 20-40 Hz) and in the range between 0 and 1 for the 1-4 H Hz range. The ~~first and~~ 1st and 3rd quartile span the "inter-quartile range" and contain 50% of the cases, 25 % each are below the 1st and above the 3rd quartile. ✔ and ✘ indicate whether the result is fine or not (see text for details).

	1-4 Hz	8-13 Hz	20-40 Hz
REG	0.519 ✔ [0.402, 0.871]	0.900 [0.846, 1.014]	1.014 [0.924, 1.153]
JOYCE	0.066 [0.026, 0.100]	0.222 [0.125, 0.340] ✘	0.303 [0.152, 0.447] ✘
HURST	1.000 ✘ [0.155, 1.000]	1.000 [0.594 ✘, 1.000]	1.000 [0.475 ✘, 1.000]
KIERKELS	0.519 ✔[0.395, 1.000 ✘]	1.000 [0.601 ✘, 1.167]	0.922 [0.428 ✘, 1.000]
FOP	0.845 ✔[0.597, 0.970]	0.985 [0.906, 1.038]	1.105 [0.989, 1.219 ✘]
PCA-3	0.109 [0.057, 0.662]	0.234 ✘ [0.143, 1.116]	0.213 ✘[0.134, 1.081]
NGCA-3	0.683 ✔[0.444, 0.981]	1.001 [0.928, 1.080]	1.088 [1.005, 1.220 ✘]

In order to summarize the results shown in figure 3~~2~~, the distribution of the spectral ratio in the range 1-4 Hz, 8-13 Hz (alpha) and 20-40 Hz are subsumed in Table ~~1~~2. In the ideal case, the spectral ratio should be close to 1 for the upper frequency bands (alpha from 8-13, and 20-40 Hz).

The method of JOYCE and PCA-3, show for the alpha band and 20 to 40 Hz a median of less than 0.31, this indicates that in at least 50% of the recordings, the spectral power is reduced by a factor of 3. Also in case of the HURST and KIERKELS method, the power ist reduced to 60.1% or less in ~~at~~ 25% (9 out of 38) of the cases. FOP and NGCA-3 show that the power in the 20-40 Hz band is increased by more than 21% in at least 25 % of the cases.

The spectral ratio should be smaller in the range of 1-4 Hz because the EOG activity is removed. JOYCE and PCA-3 show a much smaller factor, but it is reasonable to assume that ~~also~~some EEG activity~~actitivity~~ (like in the higher frequency bands) is removed. The median spectral ratio~~n~~ of 1.0 for the HURST method indicates that in at least 50% of the cases (>18 out of 38) no EOG correction was performed. The median spectral ratio of Kierkels method seems fine, but the upper IQR value of 1.0 indicates that in at least 25 % (9 out of 38) no correction was applied. The median and IQRs suggest a clear EOG reduction for FOP and NGCA-3 method, however, the reduction is less than for the regression method with a median of 0.519.

Discussion and Conclusion

Evaluation criterion

The gold standard of visual scoring of the results from 7 methods on 36 recordings of at least 60 minutes and 22 channels was not feasible in the present study. Therefore, a new criterion based on the ratio between the corrected and uncorrected EEG is suggested for evaluating EOG correction methods. Other criteria need to know the "correct" EEG and are, therefore, only applicable to

simulated data; the proposed criterion is applicable to real world EEG recordings as well.

The theoretical analysis in the appendix details the possible influence of other noise sources. The results of the regression method are not vulnerable to correlated noise, because the noise of the bipolar EOG channels is uncorrelated to the noise on the EEG channels. For the sake of simplicity, it was assumed that the noise on the BSS-based EOG components is also uncorrelated to the noise on the EOG channels. The noise on the common reference electrode might cause a reduction of this noise source, too. Because, the noise on the reference channel is indistinguishable from global EEG activity, the result would look like an overcompensation effect (case C in Fig 1). However, this would also come at the cost that some EOG component might not be properly resolved and some EOG is not removed, because the number of removed components is not increased. In other words, the method might remove some other artifacts (e.g. noise from the common reference channel) but leave some EOG activity in the data. While this could be desirable in some cases, the purpose of this criterion is to quantify the amount of removed EOG.

Comparison of Regression and component-based methods

Regression and several component-based methods are investigated. All methods rely on the assumption of an instantaneous, linear, frequency-independent, mixture of EEG and EOG activity. The difference between the regression approach (linear superposition model) and the component-based approaches (BSS and FOP, the linear mixture model) remains in the manner how the signals describing the EOG activity are obtained. While the regression approach uses the observed EOG activity; the component-based approaches decompose the data into a number of independent (and uncorrelated) components, and different heuristics are used for identifying the EOG components. In order to avoid any subjective influence, only fully automated, and no manual or semiautomated methods were considered.
Despite the expectations found in a number of works (Jung et al. 1998, 2000, Casarotto et al. 2004, Romero et al 2004, Shoker et al. 2005, Nicolaou and Nasuto, 2005, Phlypo et al 2007), an advantage of component-based methods could not be confirmed within this study. A possible explanation has been suggested already by Vorobyov et al (2002, p.294): the identified EOG components can be also contaminated by EEG activity and other noise sources. The main criticism on the regression method was the possible EEG contamination of the EOG channels, which can cause overcompensation (reduction of EEG activity). This criticism is not confirmed by the present results, because almost no reduction (ratio r(f) is close to one) for the high frequency activity can be observed with the regression method. A likely explanation for the belief that the regression method is much more vulnerable to overcompensation than BSS methods, is the incorrect use of monopolar EOG channels as regressors. Monopolar EOG channels with a common reference electrode for EEG and EOG picks up a lot of global EEG activity, which is then also removed. Consequently, it is important that bipolar EOG channels with EOG electrodes close to the eyes are used as regressors.

The fact that some BSS methods use higher order statistical moments (nonlinear information), is only an advantage for the identification step but not for the correction step. The correction step is the same linear operation for all methods (e.g. regression method), because all are based on the same linear model. The higher order information might be advantageous compared to the PCA methods, which is also a blind method. The regression method is not a “blind” source separation method, but utilizes the available information from the EOG channels. In other words, the regression method does not need to rely on some higher-order information for identifying the ocular artifacts, because it is already using the available reference information. The results suggest, it is more difficult to identify the artifact components with BSS methods than with the dedicated channels (like EOG) recording the artifact. Novel methods (like the NGCA algorithm) might eventually overcome the limitation(s) of the current BSS methods.

One could in principle criticize our findings by arguing that (i) the data sets were corrupted by too much noise (non-EEG and non-EOG noise), (ii) the software implementation is insufficient, or (iii) the methods are not good enough. We have addressed criticism in the following ways: (i) The data sets were recorded according to routine procedure by well-trained staff. We have considered saturation artifacts (Schlögl et al. 1999), other artifact types like electrode artifacts or muscle artifacts were ignored in the present study. Even if these artifacts cause the degeneration of several EOG correction results, one must say that regression analysis seems to be most robust.
(ii) The software is available through the open source software library BioSig http://biosig.sf.net and may be checked.
(iii) the main differences between the regression approach and the blind source separation methods is the fact that regression utilizes directly all available information from the EOG channels, whereas the BSS methods ignore this additional information and solve the EOG removal problem indirectly.

The results in Fig 3 show that the correction procedure (regresssion) does not only influence the EOG-related frequencies in the EEG, but also the other noise (e.g. 50 Hz line interference) can be affected. This can be easily explained by the fact that the EOG channels do not only record EOG but also other noise sources like line interference. Accordingly, the noise sources will propagate through the correction step and contaminate the corrected data. This is an experimental confirmation of the theoretical considerations based on the "noisy model". However, if the electromagnetic interference is picked up by the EOG channels, its likely that EEG is contaminated by the same cause; and other means to remove line interferences (e.g. Notch filter) have to be applied anyway.

Out of the 7 investigated methods, only Regression (REG), optimal projection filter (FOP) and NGCA gave reasonable result. However, NGCA and FOP do not show any advantage over the regression method which is easier to a apply and simpler to compute. Therefore, the regression method is the state-of-the-art method (i.e. reference method) for automatic EOG artifact elimination. Future approaches for EOG removal should be compared and validated against the regression method. This finding is in line with the general insight to always try to directly solve a problem given all information available (Vapnik 1995): regression does so by using the EOG channels as additional information, whereas the component-based approaches identify the components blindly and use some heuristics for selecting the ocular components. The heuristics can be more or less successful, currently they are less successful than the two-channel regression method.

The inferior performance of the automated BSS methods is disappointing because it does not seem to provide a solution to the problem of automatically identifying all EOG components. The EOG dipole lives in a three dimensional space (has a x,y and z direction, or VEOG, LEOG and REOG according to Elbert et al. 1985), and therefore removing only two spatial components is not sufficient (Schlögl et al. 2007). Each eye has its own dipole with spatial components (total 6 spatial components). However, the two dipoles are strongly correlated, and therefore a reduction of the number of components might be possible. Whether three components are sufficient, or additional components are needed depend on the application area. E.g. if vergence movements can not be avoided (because of fixating objects at different distances), one additional component is needed. Rapid eye movements during sleep moves the two eyes independently, therefore at least two additional components (horizontal and vertical) are necessary. If a subject can move both eyes independently (or blink independently), three additional components (total 6) will be necessary. Berg and Scherg, (1991) showed that ocular dipole change the location during eye movement, this might also require an additional component. In these cases, it is recommended to record additional EOG channels that capture the horizontal and the vertical components of each eye.

Using additional EOG electrodes can have also another advantage. The theoretical analysis of the

noisy model shows, that other noise sources (amplifier noise, impedance noise etc.) propagates to the corrected data. It can be shown that the use of more (redundant) EOG channels can reduce this additional noise component on the corrected data. Consequently, future work in this area should use more EOG channels.

Appendix A: Noisy mixture model

The models above assume that only EEG and EOG activity is recorded. However, in practice also other components like technical artifacts (amplifier noise, quantization noise, impedance noise, power line interference) and biological artifacts (muscle artifacts, etc.) do occur and are recorded. Therefore, it is reasonable to include a noise term, too. In the following, the effect of additive noise terms on the correction procedure will be analyzed.

Lets assume the recorded EEG and EOG activity is contaminated by noise, Y and O are free of noise, and Y' and O' indicate the noise contaminated, recorded data.

$$\vec{Y}'_t = \vec{Y}_t + \vec{N}_{Y,t} \quad \text{(A1)}$$

$$\vec{O}'_t = \vec{O}_t + \vec{N}_{O,t} \quad \text{(A2)}$$

with signal covariances $\Sigma_Y = \langle \vec{Y}_t \cdot \vec{Y}_t^T \rangle$ and $\Sigma_O = \langle \vec{O}_t \cdot \vec{O}_t^T \rangle$ and noise covariances $\Sigma_{NY} = \langle \vec{N}_{Y,t} \cdot \vec{N}_{Y,t}^T \rangle$ and $\Sigma_{NO} = \langle \vec{N}_{O,t} \cdot \vec{N}_{O,t}^T \rangle$.

Under the assumption that EEG and EOG are uncorrelated, we got the true model coefficients by equation (5). Moreover, it is reasonable to assume that the additional noise terms (caused e.g. by amplifier and electrode impedance) are uncorrelated to the other signals , specifically

$$\langle \vec{Y}_t \cdot \vec{N}_{Y,t}^T \rangle = 0 \quad \langle \vec{O}_t \cdot \vec{N}_{O,t}^T \rangle = 0 \quad \langle \vec{O}_t \cdot \vec{N}_{Y,t}^T \rangle = 0$$

$$\langle \vec{Y}_t \cdot \vec{N}_{O,t}^T \rangle = 0 \quad \langle \vec{E}_t \cdot \vec{N}_{Y,t}^T \rangle = 0 \quad \langle \vec{E}_t \cdot \vec{N}_{O,t}^T \rangle = 0$$

and we get

$$b = \langle \vec{Y}_t \vec{O}_t^T \rangle \cdot \langle \vec{O}_t \vec{O}_t^T \rangle^{-1} = \langle \vec{Y}'_t \vec{O}'^T_t - \vec{N}_{Y,t} \vec{N}^T_{O,t} \rangle \cdot \langle \vec{O}'_t \vec{O}'^T_t - \vec{N}_{O,t} \vec{N}^T_{O,t} \rangle^{-1} \quad \text{(A3)}$$

Typically the noise terms are not known, and the simplified estimator

$$\beta = \langle \vec{Y}'_t \cdot \vec{O}'^T_t \rangle \cdot \langle \vec{O}'_t \cdot \vec{O}'^T_t \rangle^{-1} \quad \text{(A4)}$$

is used. Depending on the fact whether the noise term in the nominator or in the denominator are ignored, the estimates will be to large or to small causing over- or under-compensation for EOG artifacts.

The term $\langle \vec{N}_{Y,t} \cdot \vec{N}_{O,t}^T \rangle$ accounts for the fact of a common noise source (e.g. the noise of a common reference electrode, muscle artifacts, etc.). If this term is non-zero but ignored, the estimates can become larger than the true model parameters $|\beta| > |b|$.

The term $\Sigma_{NO}=\langle \vec{N}_{O,t}\cdot\vec{N}_{O,t}^{T}\rangle$ accounts for the noise covariance of the EOG channels. Omitting this noise covariance can lead to an systematic underestimation of the weighting factors. For example if the EOG noise is as large as the EOG, the denominator will become twice as large, consequently the model estimates will be only half of the true model parameters, causing a sever underestimation of the model coefficients. Consequently, in order to minimize the estimation error $(b-\beta)$, the noise covariances must be known or should be made as small as possible. The corrected data is

$$\vec{E}'_t=\vec{Y}'_t-\beta\cdot\vec{O}'_t=\vec{Y}_t+\vec{N}_{Y,t}-\beta\cdot\vec{O}_t-\beta\cdot\vec{N}_{O,t}=\vec{E}_t+(b-\beta)\cdot\vec{O}_t\cdot+\vec{N}_{Y,t}-\beta\cdot\vec{N}_{O,t} \quad \text{(A5)}$$

and thus

$$\vec{E}'_t=\vec{E}_t+(b-\beta)\cdot\vec{O}_t+\vec{N}_{E,t} \quad \text{(A6)}$$

This equation describes the deviation of the actually corrected data E' from the "true" EEG data E as defined by equation (6). The term $(b-\beta)\cdot\vec{O}_t$ accounts for the fact that the EOG correction is not perfect, because of errors in the estimated model parameters (e.g. over- or under-compensation can occur). The terms $\vec{N}_{O,t}\cdot\beta$ account for the fact that noise on the obtained EOG component is propagated to the corrected EEG data, thus the covariance of the EEG noise becomes

$$\Sigma_{NE}=\langle(\vec{N}_{Y,t}-\beta\cdot\vec{N}_{O,t})\cdot(\vec{N}_{Y,t}-\beta\cdot\vec{N}_{O,t})^{T}\rangle=\Sigma_{NY}+\beta\cdot\Sigma_{NO}\cdot\beta^{T}-\beta\cdot\langle\vec{N}_{O,t}\cdot\vec{N}_{Y,t}^{T}\rangle-\langle\vec{N}_{Y,t}\cdot\vec{N}_{O,t}^{T}\rangle\cdot\beta^{T} \quad \text{(A7)}$$

The last two terms are zero if the EEG noise and the EOG noise are uncorrelated.

The spectral ratio between corrected and uncorrected data of the realistic (noise-contaminated) model is

$$r_i'(f)=\frac{|E_i'(f)|^2}{|Y_i'(f)|^2}=\frac{|Y_i(f)-\vec{\beta}_i\cdot O(f)+N_{Yi}(f)-\vec{\beta}_i\cdot N_O(f)|^2}{|Y_i(f)|^2+|N_{Yi}(f)|^2}$$

$$r_i'(f)=1+\frac{-|\vec{\beta}_i\cdot O(f)|^2+|\vec{\beta}_i\cdot N_O(f)|^2-2\cdot real\{N_{Yi}(f)\cdot\vec{\beta}_i N_O(f)\}}{|Y_i(f)|^2+|N_{Yi}(f)|^2} \quad \text{(A8)}$$

The proof for (A8) is based on the definition of the Fourier-transformation FT{.} and its properties:

$$U(f)=FT\{u_t\},V(f)=FT\{v_t\}$$

$$FT\{a\cdot u_t+b\cdot v_t\}=a\cdot FT\{u_t\}+b\cdot FT\{v_t\}=a\cdot U(f)+b\cdot V(f)$$

$$|U(f)|^2=|FT\{u_t\}|^2=FT\{cov(u_t,u_t)\}$$

$$|FT\{u_t+v_t\}|^2=FT\{cov(u_t,u_t)+2\,cov(u_t,v_t)+cov(v_t,v_t)\}$$

$$|FT\{u_t+v_t\}|^2=|U(f)|^2+|V(f)|^2+2\cdot|U(f)\cdot V(f)|\cdot\cos\phi_{U,V}$$

$$|FT\{u_t+v_t\}|^2=|U(f)|^2+|V(f)|^2+2\cdot real\{U(f)\cdot V(f)\}$$

where $cov(u_t,v_t)$ is the cross-correlation function between u_t and v_t.

Accordingly, the power spectrum density of the corrected EEG is

$$|E'(f)|^2=|Y(f)-\beta\cdot O(f)+N_Y(f)-\beta N_o|^2=$$
$$FT\{cov(Y,Y)+cov(\beta O,\beta O)+cov(N_Y,N_Y)+cov(\beta N_O,\beta N_O)-2\,cov(Y,\beta O)+$$
$$+2\,cov(Y,N_Y)-2\,cov(Y,\beta N_O)-2\,cov(\beta O,N_Y)+2\,cov(\beta O,\beta N_O)-2\,cov(N_Y,\beta N_O)\}$$

Because

$cov(Y,N_Y)=0; cov(Y,\beta N_O)=0; cov(\beta O,N_Y)=0; cov(\beta O,\beta N_O)=0; cov(N_Y,\beta N_O)=0.$

and

$cov(Y,\beta O)=cov(E+\beta O,\beta O)=cov(E,\beta O)+cov(\beta O,\beta O)=cov(\beta O,\beta O)$

becomes

$$|E'(f)|^2=|Y(f)-\beta\cdot O(f)+N_Y(f)+\beta N_O|^2=$$
$$FT\{cov(Y,Y)-cov(\beta O,\beta O)+cov(N_Y,N_Y)+cov(\beta N_O,\beta N_O)-2\,cov(N_Y,\beta N_O)\}=$$
$$|FT\{Y\}|^2+|FT\{N_Y\}|^2-|FT\{\beta O\}|^2+|FT\{\beta N_O\}|^2-2\cdot real\{N_Y(f)\cdot\beta N_O(f)\}$$

Applying this formula to each channel i and normalizing with $|FT\{Y\}|^2+|FT\{N_Y\}|^2$ yields the second line in equation (A8).

For simplicity lets assume the noise terms N_O and N_Y can be neglected. Then, a variation of $r(f)$ for different frequencies can be explained only by differing EEG and EOG spectra, and thus independent activity. The ratio $r(f)$ becomes smaller than one in case of a non-zero ocular activity $O(f)$ and is close to one if no ocular activity occurs. The ratio is also 1 in case of no correction $\beta=0$. The ratio $r(f)$ can become larger than one in case of a significant (non-EEG, non-EOG) noise source N_O contaminating the EOG signal. Furthermore, if $O(f)$ contains EEG activity, a reduction of the ratio in the higher frequency range can be seen. Accordingly, the spectral distribution of the ratio $r_i(f)$ can be used to identify over-compensation as well as propagating noise from the EOG to the corrected EEG channels.

The criterion can not be used to determine whether the correction coefficients β are correct. However, if there is no correlation between the noise of the EEG and noise of the EOG channels, only under-compensation can occur. Therefore, the largest correction (smallest ratio $r(f)$) in the low frequency EOG range is likely to provide the most accurate correction.

Acknowledgment:

This work was supported by the EU "BrainCom" (FP6-2004-Mobility-5 Grant No 024259), BMBF and DFG. The authors thank for valuable discussions with Christine Carl, Matthias Krauledat , Frank Meinecke, Guido Nolte, Gernot Supp and Carmen Vidaurre,as well as Gabriel Curio and Robert Leeb for the data recordings

References:

Anderer P, Roberts SJ, Schlögl A, Gruber G, Klosch, G, Herrmann W, Rappelsberger P, Filz O, Barbanoj MJ, Dorffner G, Saletu B. Artifact Processing in Computerized Analysis of Sleep EEG - A Review. *Neuropsychobiology* 40(3):150-157, 1999.

Barbati G, Porcaro C, Zappasodi F, Rossini PM, Tecchio F. Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals. *Clin Neurophysiol.* 115(5):1220-32, 2004.

Berg P and Scherg M "EOG source dipole modeling: Dipole models of eye movements and blinks", *Electroencephalography and clinical Neurophysiology* 79: 36 – 44, 1991.

Blanchard G, Kawanabe M, Sugiyama M, Spokoiny V, Müller K-R. In search of non-gaussian components of a high-dimensional distribution. *Journal of Machine Learning Research* 7, 247-282, 2006.

Blankertz B, Tomioka R, Lemm S, Kawanabe M, Muller K.-R. Optimizing Spatial filters for Robust EEG Single-Trial Analysis. *IEEE Signal Processing Magazine*,25(1): 41 – 56, 2008.

Boudet S, Peyrodie L, Gallois P, Vasseur C. Filtering by optimal projection and application to automatic artifact removal from EEG. *Signal Processing* 87: 1987-1992, 2007.

Bouzida, N, Peyrodie L and Vasseur C.ICA and a gauge of filter for the automatic filtering of an EEG signal. *Proceedings. 2005 IEEE International Joint Conference on Neural Networks*, 4: 2508 - 2513, 2005

Cardoso JF and Souloumiac A. Blind beamforming for non Gaussian signals. *IEE Proceedings-F*. 140(6): 362-370, 1993.

Casarotto S, Bianchi AM, Cerutti S, Chiarenza GA.
Principal component analysis for reduction of ocular artefacts in event-related potentials of normal and dyslexic children. *Clin Neurophysiol.* 2004 Mar;115(3):609-19.

Delorme A and Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods* 134: 9-21, 2004.

Donchin E, Callaway E, Cooper R, Desmedt JE, Goff WR, Hillyard SA and Sutton S. Publication criteria for studies of evoked potentials (EP) in man. In J. E. Desmedt (Ed.), *Attention, voluntary contraction, and event-related cerebral potentials* (Progress in Clinical Neurophysiology, Vol. 1, pp. 1-11). Basel: Karger, 1977.

Dornhege G, Millan JR, Hinterberger T, McFarland DJ, Müller K-R. (eds.) *Towards Brain-Computer Interfacing*, MIT Press, 2007.

Elbert T, Lutzenberger W, Rockstroh B and Birbaumer N. Removal of ocular artifacts from the EEG - a biophysical approach to the EOG. *Electroencephalography & Clinical Neurophysiology*, 60, 455-463, 1985.

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

Fries P, Scheeringa R, and Oostenveld R. Finding Gamma.
Neuron 58: 303-305, 2008

Gratton G, Coles MG, Donchin E. A new method for off-line removal of ocular artifact. *Electroencephalogr Clin Neurophysiol.* 55(4):468-84, 1983.

Hyvärinen A and Oja E. Independent Component Analysis: Algorithms and Applications, *Neural Networks*, 13(4-5):411-430, 2000

Joyce CA, Gorodnitsky IF and Kutas M. Automatic removal of eye movement and blink artifacts from EEG data using blind component separation, *Psychophysiol.*, 41:313-325, 2004.

Jung TP, Humphries C, Lee TW, Makeig S, McKeown MJ, Iragui V and Sejnowski TJ, Removing electroencephalographic artifacts: Comparison between ICA and PCA Neural Netw. *Signal Process.*, pp. 63–72, 1998.

Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol.* 111(10):1745-58, 2000.

Kawanabe M, Sugiyama M, Blanchard G, Müller K.-R. A new algorithm of non-Gaussian component analysis with radial kernel functions. *Annals of the Institute of Statistical Mathematics*, 59(1):57-75, 2007.

Kierkels JJ, van Boxtel GJ, Vogten LL. A model-based objective evaluation of eye movement correction in EEG recordings. *IEEE Trans Biomed Eng.* 53(2): 246-53, 2006.

Lemm S, Blankertz B, Curio G and Müller KR. Spatio-spectral filters for improving the classification of single trial EEG. *IEEE Trans Biomed Eng.* 52(9):1541-8, 2005

LeVan P, Urrestarazu E and Gotman J. A system for automatic artifact removal in ictal scalp EEG based on independent component analysis and Bayesian classification. *Clin. Neurophysiol.*, 117: 912-927, 2006.

Makeig S, Bell AJ, Jung TP, and Sejnowski TJ. Independent component analysis of electroencephalographic data, in *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, Mass, USA, vol. 8: 145-151, 1996.

Malmivuo J and Plonsey R. *Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetic Fields*, Oxford University Press, New York, 1995.

Meinecke F., Ziehe A. Kawanabe M. and Muller K.-R. A resampling approach to estimate the stability of one-dimensional or multidimensional independent components. *IEEE Transactions on Biomedical Engineering*, 49(12): 1514 – 1525, 2002.

Nicolaou N and Nasuto SJ. Temporal independent component analysis for automatic artifact removal from EEG, in *Proc. 2nd Int. Conf. MDSIP*, pp. 5-8, 2004.

Nunez P. *Electric fields of the Brain: the neurophysics of EEG*. New York, Oxford University Press, 1981.

Phlypo R, Boon P, D'Asseler Y and Lemahieu I. Removing Ocular Movement Artefacts by a Joint Smoothened Subspace Estimator, *Computational Intelligence and Neuroscience*, Article ID 75079, 13 pages, 2007.

Oppenheim, A.V., and R.W. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall, 1989, pp. 730-742.

Ramoser H, Müller-Gerking J, Pfurtscheller G. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans Rehabil Eng.* 8(4):441-6, 2000.

Romero S, Mañanas MA, Riba J, Morte A, Giménez S, Clos S and Barbanoj MJ.
Evaluation of an automatic ocular filtering method for awake spontaneous EEG signals based on Independent Component Analysis. *Proceedings of the 26th Annual International Conference of the IEEE EMBS* San Francisco, CA, USA • September 1-5, 2004

Schlögl A, Keinrath C, Zimmermann D, Scherer R, Leeb R and Pfurtscheller G.
A fully automated correction method of EOG artifacts in EEG recordings. *Clin.Neurophys*. 118(1):98-104, 2007.

Semlitsch HV, Anderer P, Schuster P, Presslich O. A solution for reliable and valid reduction of ocular artifacts applied to the P300 ERP. *Psychophysiology* 23: 695–703, 1986.

Shoker L, Sanei S and Chambers J. Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm, *IEEE Signal Process. Lett.*, vol. 12, pp. 721-724, Oct. 2005.

Vapnik, V. *The Nature of Statistical Learning Theory*. Springer, New York. 1995.

Vigário RN. Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalogr Clin Neurophysiol.* 103(3):395-404, 1997.

Vorobyov S and Cichocki A. Blind noise reduction for multisensory signals using ICA and subspace filtering, with application to EEG analysis. *Biol Cybern.* 86(4):293-303, 2002.

Wallstrom GL, Kass RE, Miller A, Cohn JF and Fox NA. 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, 2004.

Wübbeler G, Ziehe A, Mackert BM, Müller KR, Trahms L, Curio G.
Independent component analysis of noninvasively recorded cortical magnetic DC-fields in humans. *IEEE Trans Biomed Eng.* 47(5):594-9, 2000.

Yandong Li, Zhongwei Ma, Wenkai Lu and Yanda Li. Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach. *Physiol. Meas*. 27: 425-436, 2006.

Yuval-Greenberg S, Tomer O, Keren AS, Nelken I, and Deouell LY. Transient Induced Gamma-Band Response in EEG as a Manifestation of Miniature Saccades.
Neuron 58: 429–441, 2008

Ziehe A, Müller KR, Nolte G, Mackert BM and Curio G. Artifact reduction in magnetoneurography based on time-delayed second-order correlations.
IEEE Trans Biomed Eng. 47(1):75-87, 2000.

Legends:

Fig 1: Ratio between spectra of corrected and uncorrected EEG. Case A indicates a reduction of EOG activity; case B show no change; case C shows a clear overcompensation, in other words, also some EEG activity has been removed; and in case D the correction procedure adds some noise components to the EEG.

Fig 2: Correcting ocular artifacts. One example from data set A is shown. The first plot (a) shows time course of the raw EEG at Fz, the corrected EEG at Fz, and the two bipolar EOG channels. The subject was asked to perform various eye movements at t<0, and did the EEG experiment (cued motor imagery) at time t>0. The subjects were asked to perform the eye blinks only between trials. The first data segment with the eye movements (t<0) was used to estimate the correction coefficients, (b) shows the topographic distribution of the correction coefficients for the two bipolar EOG channels. Then, the correction coefficients were used to correct the EEG data. Based on the experimental EEG data (t>0), the spectra of the raw and corrected EEG and the spectral ratio $r(f)$ was computed (c) using the Welch method with a Hanning window.

Fig 3: Ratio r'(f) between corrected and uncorrected spectral density. Results from 7 methods and 2 different data sets are shown. Each line is the result from the most frontal electrode from a single recording. Data set A was recorded with 0.5-100Hz filter and a sampling rate of Fs=250Hz, data set B used a 0.016 - 250 Hz filter and Fs=1000Hz. Data set B was down-sampled to 250 Hz by averaging 4 consecutive samples. The eye movement data used for estimating the correction coefficients were filtered with 0.5 – 100 Hz. The correction coefficients were applied to the unfiltered data of the experimental EEG data. The spectral ratio was obtained from the experimental EEG data only.

Figure 1

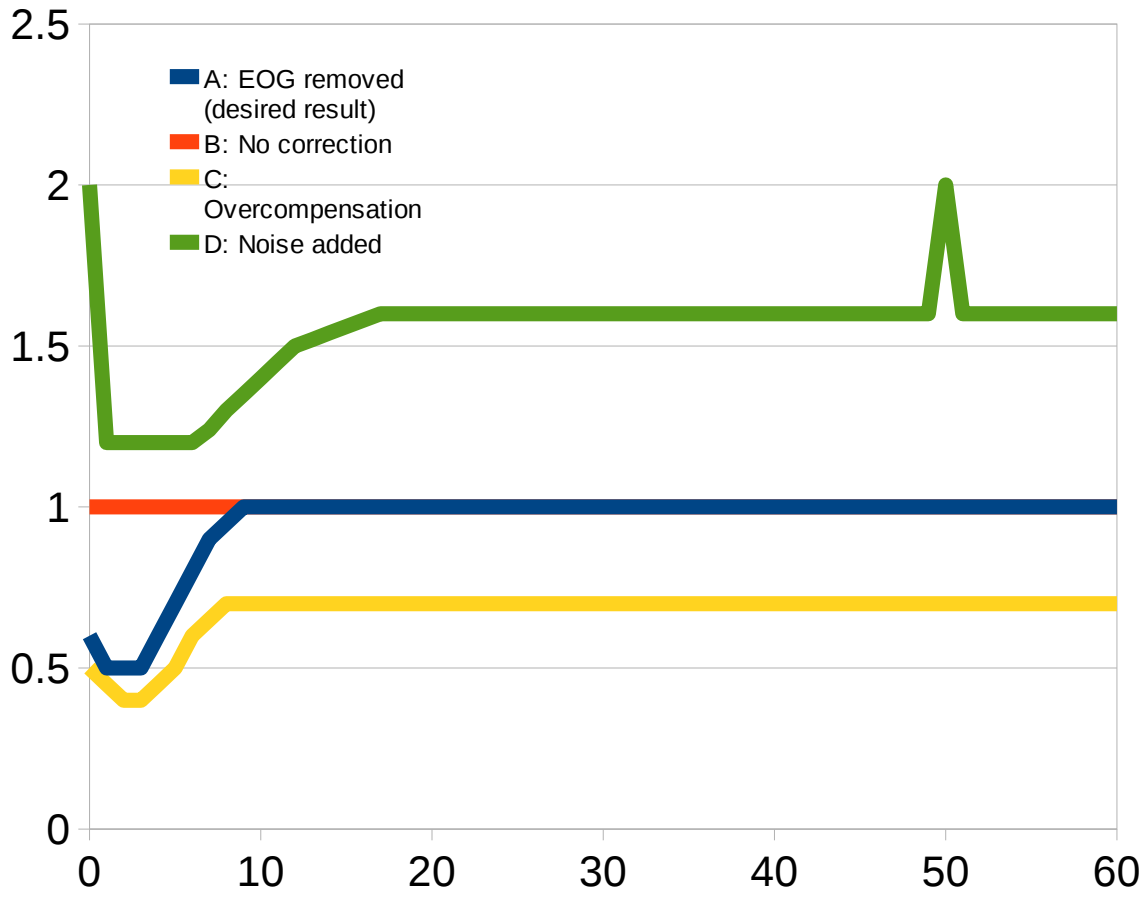

Figure 2

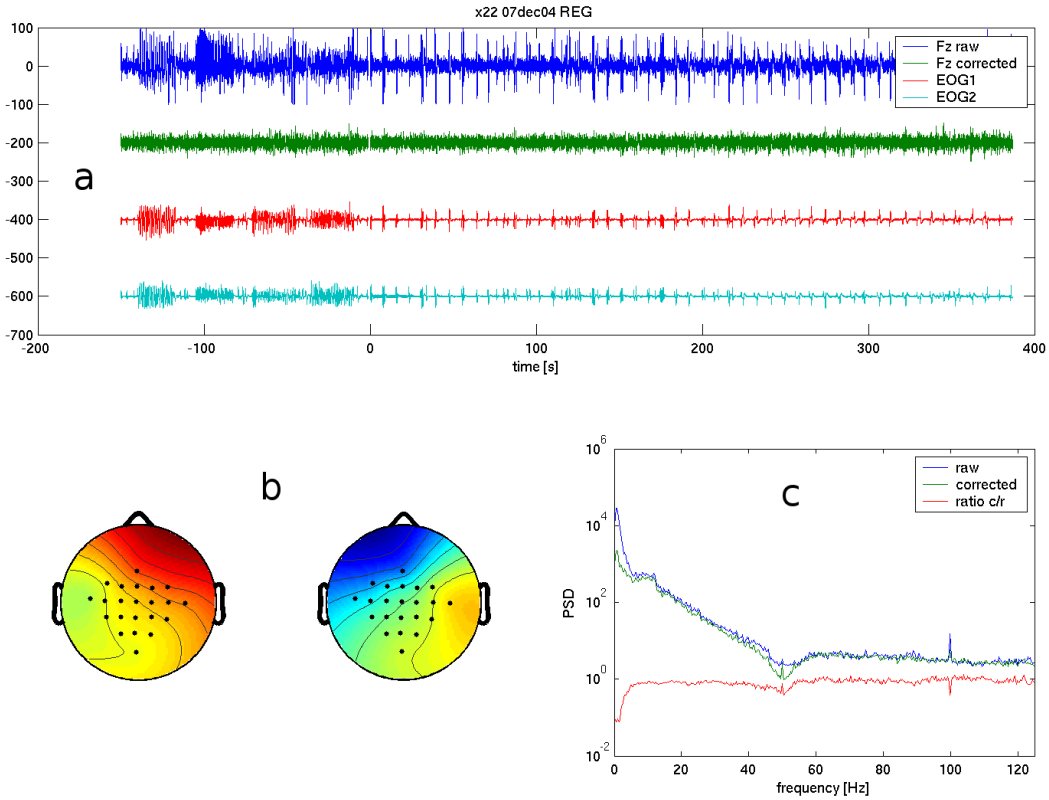

Figure 3