# A Model-Based Objective Evaluation of Eye Movement Correction in EEG Recordings

Joep J. M. Kierkels*, Geert J. M. van Boxtel, and Leo L. M. Vogten

***Abstract*—We present a method to quantitatively and objectively compare algorithms for correction of eye movement artifacts in a simulated ongoing electroencephalographic signal (EEG). A realistic model of the human head is used, together with eye tracker data, to generate a data set in which potentials of ocular and cerebral origin are simulated. This approach bypasses the common problem of brain-potential contaminated electro-oculographic signals (EOGs), when monitoring or simulating eye movements. The data are simulated for five different EEG electrode configurations combined with four different EOG electrode configurations. In order to objectively compare correction performance for six algorithms, listed in Table III, we determine the signal to noise ratio of the EEG before and after artifact correction. A score indicating correction performance is derived, and for each EEG configuration the optimal correction algorithm and the optimal number of EOG electrodes are determined. In general, the second-order blind identification correction algorithm in combination with 6 EOG electrodes performs best for all EEG configurations evaluated on the simulated data.**

***Index Terms*—Artifact removal, electroencephalography, eye movements, modeling.**

## I. Introduction

THIS paper introduces a method to objectively assess the performance of eye movement artifact correction algorithms used in electroencephalographic (EEG) measurements. The EEG is a recording of potential changes on the scalp caused by brain activity. It is often used in clinical situations, for instance to diagnose sleep disorders or epilepsy, because it reveals important information about a person's mental condition. The EEG can be distorted by numerous other sources of electrical activity, called artifact sources. Before the information in the EEG can be retrieved, however, any artifacts should be removed. Eye movement artifacts can have a large disturbing effect on EEG recordings because the eyes are located close to the brain. The front of the eye (cornea) is positively charged with respect to the back (retina) and, thus, the eye can be seen as a dipole [1]. Rotation of this dipole, caused by eye movements, changes the electric field in the tissues surrounding the eye. This change in electric field will be picked up by EEG electrodes as a change in electric potential. The magnitude of this potential change can be as much as ten times the change due to brain activity at frontal electrodes. Eye blinking also causes artifacts. When the eyelid is moved during blinking, the electric field surrounding the ocular dipole changes and the EEG electrodes record blink artifacts. In this paper, blinks will not be examined because the model that is needed to simulate the blink artifact is different from the model for eye movement artifacts. The blink model would be an extension of the model presented in this paper. The data we recorded for this study did contain blink artifacts but for performance validation we only selected intervals without blinks.

An EEG electrode is positioned near some part of the brain and mainly, but not exclusively, records cerebral signals. The EEG electrode also records some ocular signals. Likewise, an electro-oculographic (EOG) electrode is positioned close to the eyes and mainly, but not exclusively, records ocular signals. It also records small cerebral signals. This "double signal" recorded at every electrode, called cross-over, is the main reason why artifact correction is so difficult. An electrode, positioned anywhere on the scalp, will record a signal $\mathrm{E(t)}$ that is a combination of a brain-related potential $\mathrm{B(t)}$, an artifact-related potential $\mathrm{O(t)}$ and electrode-noise $\mathrm{N(t)}$ that is assumed here to have a white spectrum.

Since the volume conduction in biological tissues can be considered instantaneous [2], the summation of $\mathrm{O(t)}$, $\mathrm{B(t)}$, and $\mathrm{N(t)}$ is also instantaneous.

Because it is not possible to "turn off" either the ocular sources or the cerebral sources it is not possible to record either $\mathrm{B(t)}$ or $\mathrm{O(t)}$ alone. The eye movement artifact, thus, has to be removed from the combined recording by means of signal processing. This has led to the development of several correction algorithms. A description of the differences between these algorithms can be found in reviews [3]–[6]. Which of these algorithms performs best is difficult to determine because "clean" $\mathrm{B(t)}$ and $\mathrm{O(t)}$ signals are not available. Once the recorded signal is corrected optimally, the corrected EEG, $\mathrm{C(t)}$, should be the same as $\mathrm{B(t)}$. But $\mathrm{B(t)}$ cannot be measured, so there is no performance measure that can validate the results of a correction algorithm applied on real data. Several papers have, therefore, made a comparison between some of the algorithms based on data simulated by (randomly) mixing simulated sources. Recently, regression based algorithms, the Principal Component Analysis (PCA) algorithm and the Independent Component Analysis (ICA) algorithm were evaluated in [7]. Results of that study showed that PCA and regression-based methods are favored over ICA methods.

By using the conductivity properties of the human head and eye tracker data to simulate ocular movements we are able to simulate ocular movement related artifacts, $\mathrm{O^S(t)}$, in the EEG.

Manuscript received October 20, 2004; revised May 4, 2005. This work was supported by a grant from the Co-operation Centre Tilburg and Eindhoven Universities. *Asterisk indicates corresponding author.*

*J. J. M. Kierkels is with the Electrical Engineering Department, Eindhoven University of Technology, P.O. Box 513, 5600 MB, Eindhoven, The Netherlands (e-mail: j.j.m.kierkels@tue.nl).

G. J. M. van Boxtel is with the Psychology Department, University of Tilburg, 5000 LE Tilburg, the Netherlands. (e-mail: g.j.m.v.boxtel@uvt.nl).

L. L. M. Vogten is with the Electrical Engineering Department, Eindhoven University of Technology, 5600 MB, Eindhoven, The Netherlands (e-mail: l.l.m.vogten@tue.nl).

Digital Object Identifier 10.1109/TBME.2005.862533

These artifacts will be superimposed on the simulated, artifact-free, EEG, $\mathrm{B^S(t)}$. The use of the eye tracker gives the unique opportunity to have a very realistic eye movement in our simulations. Moreover the eye tracker does not pick up any brain related potentials and therefore is very useful to simulate clean $\mathrm{O^S(t)}$ data. Since we are now able to simulate $\mathrm{O^S(t)}$ and $\mathrm{B^S(t)}$ separately and without artifacts, we can use them as references when evaluating how well a correction algorithm has removed eye movement artifacts. The model used in this paper is based on the boundary element method (BEM) [8], which models the shapes of the main tissues in the head as different boundaries. The differences in conductivity between these tissues can be modeled by assigning a different conductivity value, $\sigma$, to each tissue. Brain activity and ocular activity can be modeled separately. The method we propose can be used to evaluate objectively which algorithm performs best in a specific electrode configuration. We use a signal-to-noise ratio (SNR) measure to determine the performance of the algorithms for various EEG and electro-oculographic (EOG) electrode configurations. On real data, the correction performance might deviate from the results shown here for simulated data. Any possible deviancies are due to features from reality that were omitted in the model and, thus refining the model can minimize this deviance, but at the cost of more complexity.

## II. Simulations

A requirement in choosing a simulation model for this study is that the model should be able to simulate separately $\mathrm{B^S(t)}$ and $\mathrm{O^S(t)}$. Also electrode-noise, $\mathrm{N^S(t)}$, should be simulated. A model should, thus, simulate the cerebral and ocular sources and also the transfer from source to electrode resulting in $\mathrm{B^S(t)}$ and $\mathrm{O^S(t)}$. As mentioned before, $\mathrm{B^S(t)}$ and $\mathrm{O^S(t)}$ originate from electrically active tissues inside the head, either brain or eye. In the model, such a structure is called a "source-dipole." Each source-dipole has a specific location, determined by the position in the head, and a specific activity determined by the electrical activity it generates. For the calculation of the potential at the electrode positions the BEM, as described in [8], is implemented in MATLAB 6.5. The BEM, often used in biopotential studies [9], calculates the potential at any position in an arbitrarily shaped volume. This potential can be a weighted summation of multiple electrical source-dipoles in that volume. In Fig. 1, details are shown of our model to simulate the $\mathrm{B^S(t)}$, $\mathrm{O^S(t)}$, and $\mathrm{N^S(t)}$. The blocks in Fig. 1 will be explained in this section.

The lower part of Fig. 1 illustrates the steps needed to obtain a mesh of the boundaries between different tissues in the head. Such a mesh is required as an input for the BEM. For this mesh we simplify the head to 4 different tissues: scalp, skull, brain and eye. Conductivity values of these tissues are taken from literature, with $\sigma = 0.3$, 0.04, 0.25, and 1.0 ($\Omega^{-1}\mathrm{m}^{-1}$), respectively [10], [11]. Points that specify the boundaries between these tissues are imported from the "ASA 3.0.0.7. Signal & Source" software package [12]. By linking these points systematically, a closed triangular mesh is generated as follows. Points are grouped in transversal slices according to their coordinates. Each point is clustered with the nearest clockwise point in that group, and with the nearest points from the groups both below and above the transversal slice of the group. The resulting

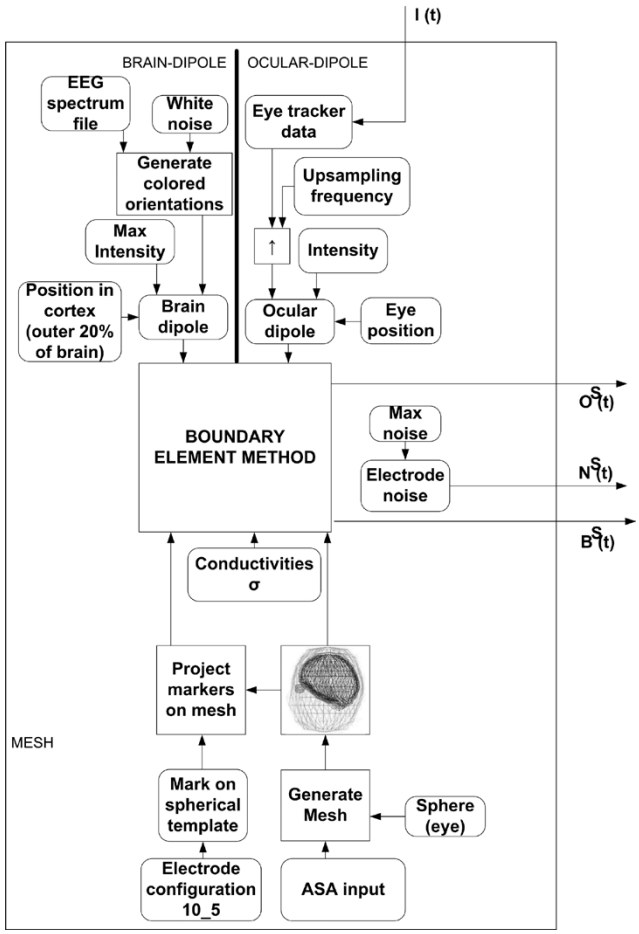


Fig. 1. Model used to simulate $\mathrm{O(t)}$ and $\mathrm{B(t)}$. Input of the model is eye tracker data, $\mathrm{I(t)}$. Other important parameters are: EEG spectrum, volume specification (mesh), electrodes positions and conductivities. The upper part of the figure models the dipole properties, the lower part models the volume. These are combined in the BEM to simulate the output $\mathrm{B^S(t)}$, $\mathrm{N^S(t)}$ and $\mathrm{O^S(t)}$.

TABLE I
Number of Points on the Boundaries Between Tissues Used in This Paper

| Boundary | Number of points |
|---|---|
| Skull-Brain | 860 |
| Scalp-Skull | 616 |
| Air-Scalp | 510 |
| Scalp-Eye | 151 |

cluster contains two triangles that are part of the triangular mesh. On the top and the bottom of the head, one point is added to assure the full triangular mesh represents a closed surface.

The number of points, listed in Table I, determines the complexity of this mesh.

The boundary between scalp and eye is not obtained from ASA, because no eye boundaries are described in this package. This boundary is added by sampling a sphere, with a radius equal to an eye, and positioning these samples in the other meshes. This eye positioning is done based on an MR cross section of a head. Landmark points are marked in this image and angles and (relative) distances between the eyes and these landmark points are determined and then used to position the eyes in the model. The resulting triangular mesh of the head is

shown in one of the boxes of Fig. 1. The next step adds electrode positions to this mesh. The positions, as specified in e.g., the 10–5 system [13], are based on relative distances between landmark points on the scalp (nasion and inion). To map the electrode positions we use a spherical template with nasion and inion on opposite sides and with a radius equal to the mean radius of the outer boundary of the head. According to the 10–5 system specifications we place one marker for each position on the spherical template. Because the outer boundary of the head is not spherical, not all of the markers on the template are on the triangular mesh. The last step in defining the electrode positions is, thus, to project these "off mesh" markers onto the nearest triangle of the mesh.

The upper part of Fig. 1 illustrates the simulation of electric dipoles inside the mesh. These dipoles simulate electric activity in the head. Each dipole has its own specific location, orientation, and intensity. In this paper, the dipoles were of two types. The first type simulated the ocular activity. Each eye has a fixed location, so the dipole modeling this eye is also fixed. The potential difference between the front and the back of the eye is constant over a short period, assuming illumination is constant [14]. Therefore, the intensity of the dipole is also chosen constant. The orientation of an eye changes with every eye movement. We use an eye tracker system to monitor these eye movements and use the eye tracker data to determine the orientation of the dipole. In the simulations, one dipole of this type is used for each eye. The second type of dipole simulates the brain activity. Electrical brain activity originates from all neurons in the brain, approximately $10^{11}$, and the EEG is a result of the summation of these activities. Representing each neuron by a dipole is impossible, so we model large groups of neurons as one "equivalent dipole." One group would model approximately $10^{10}$ neurons. Because the EEG signal originates mainly from pyramidal neurons in the cortex, each equivalent dipole is randomly positioned in the outer 20% of the brain tissue. In this paper, the positions of these equivalent dipoles are fixed, but their orientations and intensities vary. Whereas one pyramidal cell has a fixed position and orientation, the combined average orientation and position of a group of pyramidal cells may vary, given that not all cells are innervated simultaneously. For simplicity, the model uses a varying orientation but a fixed position. The number of equivalent dipoles needed to simulate EEG recordings depends on the required simulation accuracy. In our simulations, we want the average absolute potential to be equal for all electrode positions because this is also observed in EEG recordings. With a small number of dipoles (e.g., 2) the electrodes close to the source-dipoles record a much stronger EEG signal than other electrodes further away. Increasing the number of dipoles smoothes these differences. For this study, 10 equivalent dipoles are used. The frequency spectrum of an EEG is not white, but consists of characteristic frequency bands containing most of the EEG power spectrum. These bands range from 0 to 30 Hz. The EEG spectrum is assumed to be unrelated to the direction of the ocular dipole [15]. For each participant we recorded the EEG at Cz position during a period of no eye movement. This was verified in the simultaneously recorded EOG and eye tracker data. In this period, 25 blocks of 512 samples are selected, Fourier transformed, and averaged. The resulting average spectrum of the EEG is stored and used in the model to simulate EEG signals for the specific participant.

TABLE II
ELECTRODE CONFIGURATIONS

| EEG configuration | EEG electrodes |
|---|---|
| 1 | Cz |
| 2 | Midline: Nz, Fpz, AFpz, AFz, AFFz, Fz, FFCz, FCz, FCCz, Cz, CCPz, CPz, CPPz, Pz, PPOz, POz, POOz, Oz, OIz, Iz. |
| 3 | 10-20 system (21 electrodes) |
| 4 | 10-10 system (85 electrodes) |
| 5 | 10-5 system (340 electrodes) |
| **EOG configuration** | **EOG electrodes** |
| 2 | Outer canthi left (EO5) and right (EO6) eyes |
| 4 | Above (EO2) and below (EO4) the right eye, + EO5 + EO6 |
| 6 | Above (EO1) and below (EO3) the left eye, + EO2 +EO4 + EO5 + EO6 |
| 8 | Inner canthi left (EO7) and right (EO8) eyes +EO(1-6) |

The locations of the simulated brain dipoles are chosen randomly as described earlier in this section. The orientation and intensity of each dipole need to be chosen in a way that the resulting $\mathrm{B^S(t)}$ resembles $\mathrm{B(t)}$ in spectral shape. The orientation of a single neuron in the brain does not change, nor does the direction of the electric current generated by the neuron. However, a large group of neurons, which we represent by one equivalent dipole, will have a mean current-direction that changes with time. Therefore, the orientation of the simulated brain dipole should vary in time. To simulate this property the brain-dipole is split in three orthogonal vectors, each of them having a time-varying length. This varying length is simulated by passing a white noise through a filter with the spectral shape as in the stored spectrum for this participant. Combining the three vectors results in a brain dipole that has a time varying orientation and intensity. The simulated $\mathrm{B^S(t)}$ resembles $\mathrm{B(t)}$ in spectral shape at any electrode position.

An EEG electrode configuration is often selected based on the aim of the study and/or on the hardware available. The number of electrodes used in recordings has increased in the last decades, resulting in new standards for electrode positioning. The first real standard was introduced with the 10–20 system [16], later the 10–10 [17] and the 10–5 system filled the need for a standard that could be used with a larger number of electrodes. The performance of an artifact correction algorithm might depend on the EEG configuration. Therefore, we determine the performance of correction algorithms for 5 different configurations, including the 10–20, 10–10, and 10–5 systems, as shown in Table II.

For each of the EEG configurations the number of EOG electrodes can be varied. Usually an even number of EOG electrodes is used and pairs of two electrodes are combined in horizontal- and vertical EOG (H-EOG, V-EOG). We combine 2, 4, 6, or 8 EOG electrodes with each of the EEG configurations. Which EOG electrodes are added is also shown in Table II.

At every EOG electrode position, a monopolar EOG is recorded. For some algorithms, it is common to use bipolar EOG recordings measuring the difference between the two monopolar recordings known as the V-EOG or H-EOG. For the algorithms that require bipolar recordings V-EOG or H-EOG was calculated from the electrodes in that EOG configuration,

e.g., V-EOG = EO1-EO3. Other algorithms do not have this convention and use monopolar recordings. In this paper, we only simulate even numbers of EOG electrodes because of the bipolar measurement used in some algorithms.

To determine the intensities of the equivalent dipoles and the ocular dipoles, we use the ratio between $\mathrm{B(t)}$ and $\mathrm{O(t)}$ in recorded data. This ratio indicates how large the ocular artifact is with respect to the artifact-free EEG. The ratio, determined by dividing the mean of $\mathrm{B(t)}$ over the mean of $\mathrm{O(t)}$ for a fixed time period, is different for every electrode position because the amplitude of $\mathrm{O(t)}$ decreases as the distance to the eyes increases. We determined the ratio at the Fpz position for all participants and use the average ratio of 1:3 in our simulations to scale the intensities of the dipoles in a way that the ratio between $\mathrm{B^S(t)}$ and $\mathrm{O^S(t)}$ is also 1:3 at Fpz. Finally, $\mathrm{E^S(t)}$ should have the same maximum amplitude as $\mathrm{E(t)}$ and therefore we scale all intensities in a way that the resulting $\mathrm{E^S(t)}$ has a maximum amplitude of 150 $\mu$V at Fpz position. A small part of this 150 $\mu$V is due to $\mathrm{N^S(t)}$ which is set to have a maximum amplitude of 1 $\mu$V.

## III. Data Acquisition

EEG, EOG and eye tracker measurements are gathered from 9 participants aged 19–21, 5 male and 4 female. The participants perform a task involving fast eye movements. During this task, a dot appears in the middle of the upper edge of a 19-in monitor. After 1.33 s the dot disappears and immediately reappears in the middle of the next counterclockwise edge of the screen. The participant is asked to keep his eyes on the moving dot. An eye tracking system is positioned directly below the monitor and records the position of the participants left eye. It uses an infrared light and from the light reflected by the eye the position of the center of the pupil is determined. EEG measurements are performed with 21 EEG electrodes positioned according to the 10–20 system. Another 6 electrodes are used to record the EOG. They are positioned above and below both eyes, left of the left eye and right of the right eye. Recordings for all electrodes are offline referenced to linked mastoids. EEG and EOG are recorded at 256 Hz using the BioSemi ActiveTwo system with sintered Ag/AgCl electrodes using a lowpass filter of 67 Hz. Eye tracker data are recorded at 50 Hz using the Senso-Motoric Instruments RED eye tracker with an angle resolution better than 0.1°. The eye tracker data are up-sampled from 50 to 256 Hz afterwards because the EEG is simulated at 256 Hz, illustrated by the upward arrow in Fig. 1. During the task, the participant sits comfortably in front of a monitor at 0.8-meter distance with the head supported and eyes in line with the center of the screen.

## IV. Validation of Correction

With this model any artifact correction algorithm can be tested on the simulated data. In the lower half of Fig. 2, the corrected EEG, $\mathrm{C^S(t)}$, can be compared to the $\mathrm{B^S(t)}$ to validate the correction procedure.

Unfortunately this is not possible for the upper half of Fig. 2 because there is no knowledge of $\mathrm{B(t)}$. To quantify the resemblance between $\mathrm{C^S(t)}$ and $\mathrm{B^S(t)}$ a correction performance measure is defined. Because we are looking for the overall correction, our performance measure treats all frequencies in the signal

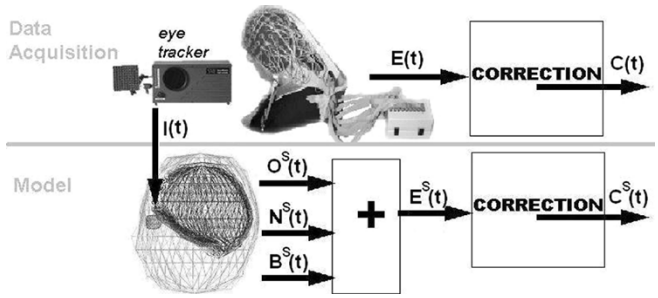


Fig. 2. Diagram illustrating the nomenclature of different data types. Part of the figure is taken from www.biosemi.com, with permission of BioSemi.

equal. Other performance measures, as in [7], can stress on specific properties of the EEG by emphasizing e.g., evoked potentials or alpha rhythms. This can be done by removing all frequencies not belonging to the alpha band for both $\mathrm{B^S(t)}$ and $\mathrm{C^S(t)}$, and compare the resulting band-limited signals. There are important factors like the number of electrodes and the configuration of electrodes that can have a significant effect on an algorithms performance. In this paper, we determine which algorithm performs best for a specific electrode configuration.

For the performance measure we use the SNR of both $\mathrm{E^S(t)}$ and $\mathrm{C^S(t)}$. Prior to SNR calculation the mean of $\mathrm{B^S(t)}$ and $\mathrm{E^S(t)}$ is subtracted, because in EEG recordings the DC component is also removed. We define $\mathrm{SNR_E}$ as

$$\mathrm{SNR_E} = \frac{\frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}} \mathrm{B^S(t)^2}}{\frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}} (\mathrm{E^S(t)} - \mathrm{B^S(t)})^2}. \tag{1}$$

The noise in $\mathrm{E^S(t)}$ is given by $\mathrm{E^S(t)} - \mathrm{B^S(t)}$ and, thus, contains both $\mathrm{O^S(t)}$ and $\mathrm{N^S(t)}$.

For $\mathrm{C^S(t)}$ we also define $\mathrm{SNR_C}$ as

$$\mathrm{SNR_C} = \frac{\frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}} \mathrm{B^S(t)^2}}{\frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}} (\mathrm{C^S(t)} - \mathrm{B^S(t)})^2}. \tag{2}$$

We have to assume here that the signal part in the SNR, the numerator, does not change due to correction. After correction this numerator still has the same mean amplitude and shape as before and, thus, still is equal to $\mathrm{B^S(t)}$. None of the correction algorithms that are used in this paper (intentionally) change the amplitude of $\mathrm{B^S(t)}$, therefore this assumption appears valid. An effective correction algorithm will reduce the noise and will have $\mathrm{SNR_C} > \mathrm{SNR_E}$ and have a high value for $\mathrm{SNR_C}$. The gain in SNR, $\gamma$, is a good indicator of the algorithm's performance

$$\gamma = \frac{\mathrm{SNR_C}}{\mathrm{SNR_E}}. \tag{3}$$

This $\gamma$-value can be calculated for each electrode. With $\mathrm{M}$ electrodes, this results in $\mathrm{M}$ $\gamma$-values. $\mathrm{SNR_C}$, $\mathrm{SNR_E}$ and $\gamma$ are not evenly distributed over the $\mathrm{M}$ electrode positions on the

scalp. To obtain an overall score for each correction algorithm the $\gamma$-values are averaged and converted to a decibel scale

$$\mathrm{G} = 20 * \log_{10}\left[\frac{1}{\mathrm{M}}\sum_{\mathrm{m}=1}^{\mathrm{M}}\gamma_{\mathrm{m}}\right]. \quad (4)$$

The resulting score G is an indicator for the performance of a correction algorithm. G (in decibels) is positive if the SNR has improved. It has a negative value if the SNR has decreased and is zero only if the SNR has not changed. An alternative definition for G could use a weighting function to give more weight to those electrodes that contain the largest artifacts. In this alternative situation, electrodes with only small eye movement artifacts, e.g., the occipital electrodes, have only a very small contribution in determining G. This alternative has some advantages over the definition we use. The alternative G would enable us to determine a G that is mainly based on those electrodes where correction is most needed. The main disadvantage is that if a correction algorithm would increase SNR at frontal sites and decrease SNR at occipital sites this will stay unnoticed in the alternative G score, and the corrected EEG will contain artifacts introduced by the correction algorithm. This is the main reason why we did not use the alternative G.

## V. Evaluated Algorithms

In principle, any correction algorithm can be evaluated by the score outlined in this paper. Over the last decades many algorithms have been proposed. ICA algorithms appear in recent articles on artifact rejection [7], [18]. FastICA [19] is an efficient implementation of ICA, joint approximate diagonalization of eigen matrices (JADE) [20] was among the first ICA algorithms to be developed. The main difference between regression-based algorithms on the one hand, and PCA and ICA algorithms on the other is that the regression-based methods do not use the data of EEG electrodes simultaneously but perform correction for 1 EEG electrode at a time, whereas the PCA and ICA methods use data from all EEG and EOG electrodes simultaneously to perform the correction. This might be an advantage of the PCA and ICA methods.

In this paper, a selection of six algorithms, shown in Table III, is evaluated.

We use an automated detection of components for the PCA, SOBI, JADE, and FastICA algorithms. If the cross correlation between an extracted component and one of the recorded EOG signals exceeds a threshold value, the component is marked as an ocular component and is removed. The threshold value was optimized in a pilot measurement and set to 0.7. The R-ARE and the SOBI algorithm require a choice of parameters. For R-ARE this is the order of the AR-model used to model the EEG, set to 4 following [22]. For the SOBI algorithm a number of time lags need to be chosen. The components retrieved by the SOBI algorithm are not correlated with each other. The autocorrelation of each component is calculated at the time lags chosen, the SOBI algorithm maximizes the sum of these autocorrelations. In this paper, we use lags of 0, 1, 2, 3, 5, 10, and 20 samples at 256 Hz. These lags were found to yield good corrections in a pilot study. It should be noted that the component-based algorithms probably have a different performance when two EOG electrodes are used with one positioned above the eye and one beside the eye. In addition, odd numbers of EOG electrodes could be used with the component-based algorithms. ICA algorithms require more electrodes than there are expected sources (dipoles) and also require more time points than the square of the number of electrodes. In order to comply with the latter requirement, the time window of 10 s. used in this study is insufficient. Either this window should be increased, or the sample frequency of the EEG should be increased. This would lead to very large datasets containing over $10^5$ samples for every EEG electrode and the use of such large sets is computationally undesirable for most of the algorithms. It should therefore be noted that for the dataset simulated in this study, with only 2560 time samples, the ICA algorithms do not perform optimally for the two largest EEG electrode configurations. Increasing the window size will probably increase their performance, but at the expense of larger computational effort. Even though the simulated dataset was too small for the ICA algorithms, they still succeeded in removing part of the artifact, resulting in a positive G. Therefore, the results of applying ICA for correction of the larger electrode configurations will still be mentioned in the results of this study.

TABLE III
The Six Algorithms Evaluated in This Paper.

| Abbreviation | Description | Reference |
|---|---|---|
| MLR | Multiple linear regression | [21] |
| R-ARE | Regression-based algorithm using regression with an autoregressive structure to model the EEG | [22] |
| PCA | Principal component analysis | [23] |
| SOBI | Second Order Blind Identification, a component analysis technique that exploits autocorrelation the components. | [24] |
| JADE | Joint approximate decomposition of eigenmatrices. One of the first ICA algorithms. | [20] |
| FastICA | An efficient ICA algorithm. | [19] |

## VI. Results

In Fig. 3, an example of the measured $\mathrm{E(t)}$ is shown. To illustrate that this data resembles simulated EEG data, $\mathrm{E^S(t)}$, $\mathrm{B^S(t)}$, and $\mathrm{O^S(t)}$ are also displayed.

The left column shows a 10 s window ($\mathrm{window\ length} = 2560\ \mathrm{samples}$) of the data, the right column shows the amplitude spectrum after applying a Hanning window. The difference in variability between $\mathrm{E(t)}$ and $\mathrm{E^S(t)}$ is caused by the error in the ratio $\mathrm{B^S(t)/O^S(t)}$. This ratio was set to 1:3 for the Fpz position, as discussed in Section II. For other positions the simulated ratio can deviate from the ratio for recorded data because the simulation uses only a limited number of dipoles.

Data segments of 10 s (2560 samples) are corrected by all 6 methods and for all 5 electrode configurations. G scores are calculated for each participant for all EEG and EOG configurations and then averaged over the 9 participants. The results are shown in Fig. 4.

This figure shows the following results.

- All performance values are positive or 0. This implies that all correction algorithms improve the SNR for all configurations studied.

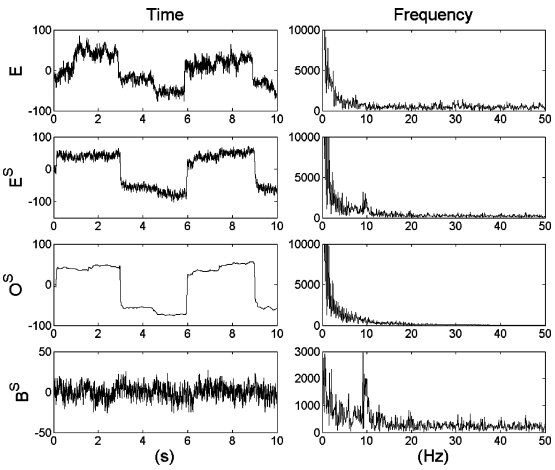


Fig. 3. From top to bottom: measured EEG, simulated measured EEG, simulated artifact, and simulated true EEG for the F7 position. Left column shows 10 s of data (2560 samples) with y axis in microvolts, right column shows the amplitude spectrum.

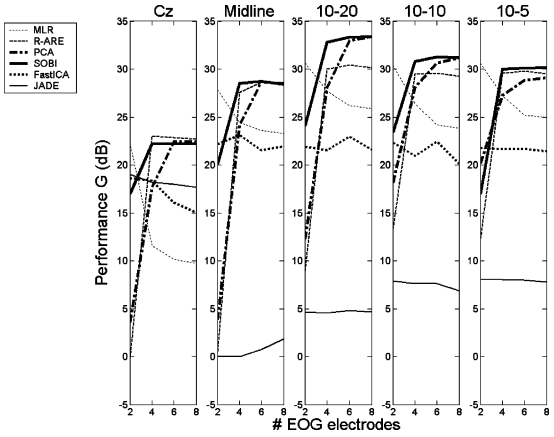


Fig. 4. Performances. The 5 graphs each show the score for one of the EEG configurations. The x axis of these graphs indicate the number of EOG electrodes that was combined with the EEG configuration. For all 20 possible electrode configurations ($5 * 4$) the performance grade G in decibels is plotted for the 6 algorithms.

- The best performing algorithms increase the SNR by approximately 25–35 dB.
- The differences between the various algorithms are large. FastICA and JADE deviate most from the other algorithms.
- When four EOG electrodes are applied, PCA and SOBI have much better correction performance than when using only two EOG electrodes.
- Adding more electrodes (EEG or EOG) in general improves the correction of the component-based algorithms.
- For increasing number of EOG electrodes, MLR has a decreasing performance, whereas the other algorithms have an increasing performance.
- For almost every EEG configuration the SOBI algorithm shows the best correction.
- The highest level of performance for 4 out of 5 EEG configurations is obtained by applying the SOBI algorithm using 6 EOG electrodes. Averaged over the 5 EEG configurations the SNR improvement is 29.0 dB, with a standard deviation of 4.3 dB.

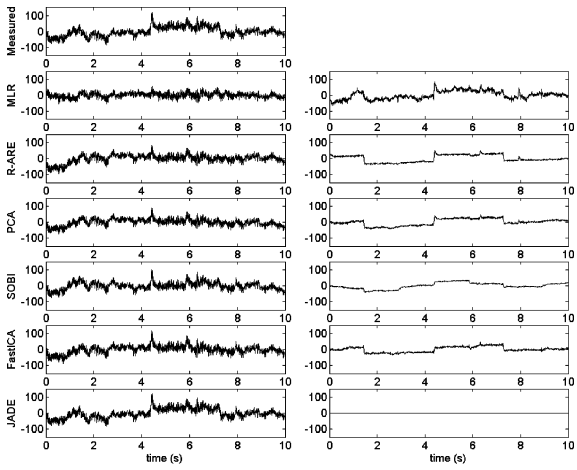


Fig. 5. Results of the algorithms applied to recorded data of one participant. On the upper left is the measured EEG at Fp2 position. The other graphs in the left column show the EEG after applying the six correction algorithms. In the right column, the rejected part of the EEG is shown for each corresponding method. All y axis show potential (microvolts).

- In most EEG configurations, however, the results of applying MLR with only 2 EOG electrodes are close to SOBIs performance, with an average SNR improvement of 28.1 dB and with a standard deviation of 3.6 dB.

To illustrate the results, we also apply the correction algorithms to the EEG data that was recorded during an experiment. The corrected EEG is shown in Fig. 5.

## VII. Discussion

First of all, it should be mentioned that the use of an eye tracker in this performance evaluation does not imply that all future EEG studies require eye tracker recordings. Once an algorithm is validated with the use of an eye tracker and appears to perform adequately, it can be applied to correct EEG data, without using eye tracker recordings. The recordings of the eye tracker were only used in this study to provide "clean" input data for the model simulations. Compared to other evaluations that use EOG or extracted components to model EOG artifacts, the eye tracker approach has the important advantage that there is no brain related disturbance in the simulated eye movement. Using the "boundary element method" we are able to simulate recordings for different EEG electrodes configurations that are often used.

The results for the ICA algorithms, JADE and FastICA, are rather disappointing, which is in agreement with a recent study in which the FastICA algorithm was used [7]. The most probable reason for this is that components were extracted that were not (entirely) of ocular origin. These components have a correlation of more than 0.7 with one of the EOG signals, but were very noisy. Removing them probably attenuates the EOG artifact, but introduces strong noise into $\mathrm{C^S(t)}$. A manual component selection, as in [7], did not yield better results. However, results might improve by using a larger window and, thus, more samples for correction. When performing an MLR, the first EOG signals will remove most of the eye movement artifact, which explains the good performance of the MLR algorithm with only two EOG electrodes. If more EOG signals are used in MLR, they will effectively start to remove actual EEG information.

Remarkably, the performance for four EOG electrodes is worse than when using only two EOG electrodes. This is not intuitive since the first two EOG electrodes do not have information on vertical eye movements, present in $\mathrm{E^S(t)}$. A possible explanation for this result is that the EOG electrodes in the model are not positioned exactly in line with the eyes. This might be due to the mesh used in this study. For the PCA and SOBI algorithms the increase in performance with increasing number of electrodes was expected because these algorithms make use of the information in all electrodes. Applying more than 6 EOG electrodes does not increase the performance any further. Our data acquisition is performed using the 10–20 system with 6 EOG electrodes, thus corresponding to the middle graph in Fig. 4. In Fig. 5, we can visually verify that the part of the signal removed by the SOBI algorithm shows most resemblance with the eye movement artifact in the recorded data. SOBI is the only algorithm that clearly removes a signal that changes every 1.33 s. This is the time the dot in our experiment stayed at the same position.

In Fig. 5, JADE did not remove any artifact. This would assume a G value of zero. However, this example of JADE correction is based only on results for one subject, and on real data. In Fig. 4, results for nine simulated signals are averaged. In some of these nine signals, JADE also did not remove any artifact, in others some artifact was removed. Averaged result are shown in Fig. 4. The PCA, MLR, R-ARE, and FastICA algorithms remove most of the artifact, but not as much as SOBI.

The results shown are limited to eye movements. For blinks, the model should be more detailed around the eyes. The movement of the eyelid could have a very large influence on the conductivity properties of the model near the eyes because tissues here have very different conductivities. Movement of the eyelid, thus, would not directly influence the eye-dipole itself but rather the model surrounding the dipole. Other studies [1] state that the location of the dipole causing the eye-blink artifacts is different from the location of the eye-movement dipole. However, since here a 3-shell head model was used, the exact tissue structure around the eyes was not modeled and no sliding eyelid could be simulated. The absence of these characteristics in the model might cause a shifted dipole location when performing a source localization for eye blinks.

Some of the parameters that are described in Section II, and that are used in the model to simulate the data, are obtained from experiments. These include the number of dipoles, the EEG spectrum shape, the amplitude ratio between $\mathrm{B(t)}$ and $\mathrm{O(t)}$ of 1:3, and the amplitude of the electrode-noise $\mathrm{N(t)}$. Small fluctuations in the number of dipoles and the spectral shape will have almost no effect on the calculated SNR improvement. The potential distribution over the scalp was already smoothed, and adding more brain dipoles will not change this distribution significantly. The EEG spectrum shape is averaged over 25 measurements. However, the ratio between $\mathrm{B(t)}$ and $\mathrm{O(t)}$ and the amplitude of $\mathrm{N(t)}$ can have impact on the calculated SNR improvement. Changing the ratio to 1:1 would reduce the need for a correction algorithm because it reduces the artifact. This implies that detecting the artifact is more difficult and the SNR improvement will probably decrease. On the other hand, changing the ratio to 1:5 will increase the need for a correction algorithm. Artifact detection will be easier now and the SNR improvement might be higher. But it is not likely that the SNR of the corrected signal is higher because that would imply that very strong artifacts would yield the cleanest results after correction. If there is more electrode-noise, $\mathrm{N(t)}$, in the measurement, both the SNR prior to correction and after correction will decrease. The magnitude of $\mathrm{N(t)}$ is determined by e.g., the placement of the electrode.

The score G we define in this paper is used to determine which correction algorithms perform well. It does not indicate whether the correction algorithms perform well enough. The definition on what is well enough is very subjective and depends on the specific application of the EEG recording. Moreover, G indicates an increase in SNR whereas the statement of what is well enough should be based on the SNR after correction. The $\mathrm{SNR_C}$ would be more suitable for this. The aim of this paper is to evaluate and compare correction algorithms based on their ability to remove eye movement artifacts and not to determine the (subjective) threshold of what $\mathrm{SNR_C}$ is well enough. Therefore, G is defined not only by the $\mathrm{SNR_C}$ after correction, but also by the $\mathrm{SNR_E}$ prior to correction.

Our simulations might be improved by using a more sophisticated BEM implementation to reduce simulation error and give a better resemblance between $\mathrm{E(t)}$ and $\mathrm{E^S(t)}$. Another improvement to the simulations might be the use of a more detailed mesh of the head, including e.g., details about the eye sockets in the skull [25]. Both improvements will increase complexity. Our main goal was to evaluate the algorithms, not to simulate EEGs as realistic as possible and therefore these improvements were not made. Without the improvements, the simulated EEG might deviate from recorded EEG on some properties. However, the properties that are of most importance for our evaluation are simulated, like the instantaneous way in which the conduction through the head takes place and the different origin of ocular and cerebral signals. The simulated signals therefore have those characteristics that make correction a challenging task.

## VIII. Conclusion

We present a method to evaluate algorithms that correct eye movement artifacts in simulated EEG recordings. The method is based on eye movement data recorded by an eye tracker, ensuring that the eye movement recordings are uncontaminated by brain activity. The data for the evaluation are simulated using a realistic model of the head, based on the boundary element method. Correction results from six algorithms are evaluated. As a criterion of how well a correction performs, we compare the signal to noise ratio before and after applying each correction algorithm. All algorithms reduce the artifact for all electrode configurations. MLR performs best when only two EOG electrodes are used. We recommend using 6 EOG electrodes and performing a SOBI artifact correction. In practice, however, calculations with the SOBI algorithm are computationally demanding, especially with the 10–5 system. If faster solutions are required, the use of the MLR algorithm with 1 H-EOG is recommended. Comparing our results with [7] we see that in both studies MLR and PCA perform better than FastICA. However, the SOBI method, not included in [7], performs in our study even better. This indicates that PCA with autocorrelated components, as applied in SOBI, is also a good candidate for correction.

Other correction algorithms, not mentioned in this paper, can also be evaluated by the method presented here. Given the assumptions stated in this paper, we have shown that the proposed method can be used to determine how well an algorithm performs and which correction algorithm can be used best for any specific EEG/EOG electrode configuration.

## REFERENCES


[1] P. Berg and M. Scherg, "Dipole models of eye movements and blinks," *Electroencephalogr. Clin. Neurophysiol.*, vol. 79, pp. 36–44, 1991.

[2] M. S. Hamalainen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, "Magnetoencephalography – Theory, instrumentation, and applications to noninvasive studies of the working human brain," *Rev. Modern Phys.*, vol. 65, no. 2, pp. 413–497, 1993.

[3] C. H. M. Brunia, J. Mocks, M. M. C. van den Berg-Lenssen, M. Coelho, M. G. H. Coles, T. Elbert, T. Gasser, G. Gratton, E. C. Ifeachor, B. W. Jervis, W. Lutzenberger, L. Sroka, A. W. van Blokland Vogelesang, G. Van Driel, J. C. Woestenburg, P. Berg, W. C. McCallum, P. D. Tuan, P. V. Popock, and W. T. Roth, "Correcting ocular artifacts in the EEG: A comparison of several methods," *J. Psychophysiol.*, vol. 3, pp. 1–50, 1989.

[4] R. J. Croft and R. J. Barry, "Removal of ocular artifact from the EEG: A review," *Neurophysiologie Clinique*, vol. 30, no. 1, pp. 5–19, 2000.

[5] T. P. Jung, C. Humphries, T. W. Lee, S. Makeig, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts: Comparison between ICA and PCA," *Neural Netw. Signal Process.*, pp. 63–72, 1998.

[6] R. Verleger, "Should we really use different estimates for correcting EEG artefacts produced by blinks and by saccades?," *J. Psychophysiol.*, vol. 14, no. 4, pp. 204–206, 2000.

[7] G. L. Wallstrom, R. E. Kass, A. Miller, J. F. Cohn, and N. A. Fox, "Automatic correction of ocular artifacts in the EEG: A comparison of regression-based and component-based methods," *Int. J. Psychophysiol.*, vol. 53, no. 2, pp. 105–119, 2004.

[8] T. F. Oostendorp and A. van Oosterom, "Source parameter estimation in inhomogeneous volume conductors of arbitrary shape," *IEEE Trans. Biomed. Eng.*, vol. 36, no. 3, pp. 382–391, Mar. 1989.

[9] C. Bradley and A. Pullan, "Application of BEM in biopotential problems," *Eng. Anal. Boundary Elements*, vol. 26, pp. 391–403, 2002.

[10] R. Hoekema, G. H. Wieneke, F. S. S. Leijten, C. W. M. van Veelen, P. C. van Rijen, G. J. M. Huiskamp, J. Ansems, and A. C. van Huffelen, "Measurement of the conductivity of skull, temporarily removed during epilepsy surgery," *Brain Topogr.*, vol. 16, no. 1, pp. 29–38, 2003.

[11] J. B. Nixon, P. E. Rasser, M. D. Teubner, C. R. Clark, and M. J. Bottema, "Numerical model of electrical potential within the human head," *Int. J. Numerical Meth. Eng.*, vol. 56, pp. 2353–2366, 2003.

[12] Advanced Source Analysis Software. ANT-software. [Online]. Available: http://www.ant-software.nl

[13] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clin. Neurophysiol.*, vol. 112, no. 4, pp. 713–719, 2001.

[14] M. F. Marmor and E. Zrenner, "Standard for clinical electro-oculography," *Documenta Opthalmologica*, vol. 85, pp. 115–124, 1993.

[15] C. W. Hesse, E. Seiss, R. M. Bracewell, and P. Praamstra, "Absence of gaze direction effects on EEG measures of sensorimotor function," *Clin. Neurophysiol.*, vol. 115, no. 1, pp. 29–38, 2004.

[16] H. Jasper, "The ten twenty electrode system of the international federation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 10, pp. 371–375, 1958.

[17] G. E. Chatrian, E. Lettich, and P. L. Nelson., "Ten percent electrode system for topographic studies of spontaneous and evoked EEG activity," *Am. J. EEG Technol.*, vol. 25, pp. 83–92, 1985.

[18] C. A. Joyce, I. F. Gorodnitsky, and M. Kutas, "Automatic removal of eye movement and blink artifacts from EEG data using blind component separation," *Psychophysiology*, vol. 41, no. 2, pp. 313–325, 2004.

[19] A. Hyvarinen, "Survey on independent component analysis," *Neural Computing Surveys*, vol. 2, pp. 94–128, 1999.

[20] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Inst. Elect. Eng Proc. F Radar and Signal Processing*, vol. 140, no. 6, pp. 362–370, 1993.

[21] S. Chatterjee and A. S. Hadi, "Influential observations, high leverage points, and outliers in linear regression," *Statist. Sci.*, pp. 379–416, 1986.

[22] M. M. C. van den Berg-Lenssen, C. H. M. Brunia, and J. A. Blom, "Correction of ocular artifacts in EEGs using an autoregressive model to describe the EEG; a pilot study," *Electroencephalogr. Clin. Neurophysiol.*, vol. 73, pp. 72–83, 1989.

[23] G. J. M. van Boxtel, "Computational and statistical methods for analyzing event-related potential data," *Behavior Res. Meth., Instrum. Comput.*, vol. 30, no. 1, pp. 87–102, 1998.

[24] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[25] Z. Akalin-Acar and N. G. Gencer, "An advanced boundary element method (BEM) implementation for the forward problem of electromagnetic source imaging," *Phys. Med. Biol.*, vol. 49, pp. 5011–5028, 2004.


**Joep J. M. Kierkels** was born in Baexem, the Netherlands, in 1979. He received the M.Sc. degree in biomedical engineering from the Eindhoven University of Technology, Eindhoven, the Netherlands, in 2002. He is currently working toward the Ph.D. degree at the department of electrical engineering, Eindhoven University of Technology.

His research interests include electroencephalography, signal separation, model validation and artifact reduction.

**Geert J. M. van Boxtel** was born in Tilburg, The Netherlands, in July 1958. He received the M.Sc. degree in psychology in 1989 from Tilburg University. In 1994, he received the Ph.D. degree from the same university.

He was a Postdoc at the University of Amsterdam, Amsterdam, The Netherlands, and at the University of Dortmund, Dortmund, Germany. He is currently an Assistant Professor of Psychology at Tilburg University, Tilburg, The Netherlands. His current research interests include cognitive neuroscience, especially event-related brain potentials and its applications.

Dr. Van Boxtel has been a member of the Society for Psychophysiological Research and the Federation of European Psychophysiological Societies since 1989.

**Leo L. M. Vogten** was born in Kerkrade, the Netherlands, in 1945. He received the M.Sc. degree in electrical engineering in 1971 and the Ph.D. degree in 1983, both from the Eindhoven University of Technology (TU/e), Eindhoven, The Netherlands.

He joined the Institute for Perception Reseach of the TU/e in 1972 and worked on auditory masking, speech analysis, speech synthesis, and text to speech conversion and the application of speech in user interfaces.

Currently, he is assistant professor at the Signal Processing Systems group of the department of Electrical Engineering of the TU/e and his research interests are in the field of medical signal processing.