34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Removal of Blink Artifacts in Single Channel EEG

Dyana Szibbo, An Luo and Thomas J. Sullivan

Abstract— **Blinks are one of the main sources of distortion in electroencephalographic (EEG) data. Discarding blink-contaminated segments of EEG data would result in considerable information loss when interpreting and analyzing data. This study presents a simple method of blink filtering using a Savitzky-Golay (SG) smoothing filter and compares it to Independent Component Analysis (ICA), a widely accepted method of blink removal. The SG-based blink filtering method arose from the need for blink removal in EEG systems with a low number of channels and limited processing power, specifically reading from the forehead location where the blink disturbance is severe. Real and simulated data were investigated with respect to the method's performance. Using correlation and mutual information to measure performance, the results reveal that the SG-based method can effectively remove blink artifacts and produces results comparable to those obtained using ICA.**

I. Introduction

Blinks are a major problem in the study of brain potentials. The electrical potentials resulting from eye blinks can be much larger than the underlying EEG and can propagate across the scalp, distorting brain signals of interest [1]. A simple solution is to discard any portions of contaminated data; however this could result in substantial data loss and biased data samples [2]. Another approach is to restrict the subjects from blinking, but this limits the experiment's design and can impact the cognitive processes involved [1]. This leaves filtering, removing the artifact yet leaving the underlying EEG, as the most desirable option.

Single channel EEG systems are becoming increasingly popular for applications where real-time processing, usability and low computational cost are of high importance. Some applications of single-channel EEG include brain-computer interfaces [3], sleep scoring programs [4], and Alzheimer's disease recognition [5]. Artifacts appear differently in different channels of EEG; therefore artifact removal in a single-channel of EEG has the advantage that the artifacts can be detected and processed specifically for that channel.

Only a few methods have been proposed for blink removal in single channel EEG, such as Hilbert-Huang reconstruction [6] and Singular Spectrum Analysis [7]. However these techniques can needlessly alter the clean EEG between ocular artifacts and are more computationally intensive than the method proposed.

D. Szibbo is with the Department of Technology Development, NeuroSky, San Jose, CA 95113 USA (phone: 408-806-8921 e-mail: dyana@neurosky.com).

A. Luo is with the Department of Technology Development, NeuroSky, San Jose, CA 95113 USA (e-mail: aluo@neurosky.com).

T. J. Sullivan was with the Department of Technology Development, NeuroSky, San Jose, CA 95113 USA (e-mail: tom@sullivan.to).

The proposed method of blink filtering requires only a single channel of EEG. Once a blink has been detected, a smoothing filter is applied to only the section of EEG with the blink. A smoothing filter lets low frequency components pass and reduces the high frequency components. Since a blink is composed largely of low frequencies, when a smoothing filter is used on EEG data containing a blink, it filters out the higher frequencies to leave us with a blink waveform. This blink waveform can then be subtracted from the original data, removing the estimated blink disturbance yet leaving the higher frequency information intact.

The smoothing filter used in this study is the Savitzky-Golay (SG) filter [8]. It was originally developed for noise reduction in the field of analytical chemistry but has become a widely used method to improve the signal-to-noise ratio of many types of signals [9,10,11] The filter essentially performs a local polynomial regression on a series of values to determine the smoothed value for each point. This method is advantageous as it tends to preserve data features such as peak height and width which are usually affected by other adjacent averaging techniques, and should therefore efficiently model a blink artifact in EEG.

Independent Component Analysis (ICA) is a widely used method of artifact removal in electrophysiological signal processing [12]. ICA requires data from multiple channels of EEG, and is based on the assumption that recorded EEG signals are linear mixtures of unknown independent components within the brain [13]. The goal of ICA is to find a linear transformation of the measured sensor signals such that the resulting source components are as independent as possible. After the source components' calculation, a clean EEG signal can be obtained by eliminating the components which correspond to artifacts.

In our study we used both simulated and real data to evaluate the performance of the SG-based method. We filtered simulated blink-contaminated EEG using the SG-based method as well as ICA and compared the results using the measure of correlation. We then filtered real blink-contaminated EEG and compared the results to ICA using cross-correlation and mutual information with the electrooculogram (EOG).

II. Methods

A. Data Processing

Graz data set 2a was downloaded from the 2008 BCI data competition [14]. There were 9 subjects with approximately 44 minutes of data each, during which the subjects were performing a motor imagery task. Twenty two Ag/AgCl electrodes and 3 EOG channels were recorded

[15]. The left mastoid was used as reference and the right mastoid as ground. The signals were sampled at 250Hz and bandpass-filtered between 0.5 and 100Hz. An additional 50Hz notch filter was used to suppress line noise.

EEGLAB [16] was used to perform our analysis. Portions of data containing artifacts other than blinks were removed by hand. Blink artifacts were identified and their peak marked manually. A script was used to adjust the marked blink latency to ensure it marked the exact peak. The blink events were randomly divided into a training set and a testing set.

The Fz channel was chosen for filtering since this study demonstrates single channel blink removal and the EEG from the frontal locations are severely contaminated by blink artifacts. To remove a blink using the SG-based method, a time period of 0.16 seconds to the left (40 samples) and 0.84 seconds (210 samples) to the right of each blink peak was identified as a 'detected blink'. This time period was chosen based on time for an averaged blink disturbance (330 blinks at the Fz location from the training set averaged together) to return back to baseline on either side of the peak. This could be implemented as a simple detection algorithm by setting an upper threshold for EEG (with any non-EOG artifacts removed) and finding the maximum point for each continuous set of points above this threshold, then stepping left 0.16 seconds and right 0.84 seconds. SG filtering was then performed only on the detected blink segment. The SG filtered portion was subtracted from the original noisy portion to leave a blink-removed EEG signal. The SG filtering parameters were chosen based on filtering an averaged blink artifact: the same average blink from above was added to random portions of uncontaminated EEG from various subjects. SG filtering parameters window length 41 and degree 3 produced filtered data with the highest average correlation to the original uncontaminated EEG.

ICA was performed using the extended infomax ICA algorithm [13] included in the EEGLAB software. Two-dimensional scalp component maps and ICA component spectra were plotted and the component representing the blink was rejected (the component frontally located with a smooth exponentially decreasing spectrum). The result of the ICA filtering on the Fz channel was saved to compare to the SG-based method.

B. Evaluating Performance

Simulated Data: With simulated data, the 'target' data is known and the accuracy of correction can be estimated more precisely than with real data. To create simulated blink-contaminated data, 22 channels of EEG were taken from a random subject (Subject 1) and artifacts, including blinks, were removed manually. An average blink was created by averaging all the blinks from Fz from the test data set. The average blink was varied in height and width, consistent with ranges observed in the real data. Blinks were also inverted to take into account inverted blinks observed in past experience. The latencies of the blinks were determined using a range of blink rates for normal subjects found by [17]. The blink activations were added to the 22 channels using weights for the blink component from ICA that had been run on a separate subject's data (Subject 5). "Fig. 1" shows a 25 second segment of the resulting simulated blink-contaminated data. In total, 23 minutes of simulated data containing 195 blinks were constructed. ICA was then run on all 22 channels of data and the SG-based method was run on the Fz channel. The correlation coefficient for the continuous target data and filtered data for the Fz channel was compared for both methods.

Real Data: With empirical data the true EEG isn't known, therefore an indirect approach must be used to measure the accuracy of blink-removal. In theory, the similarity between the blink contaminated EEG and the EOG should be reduced after blink filtering. Two metrics were used to explore the reduction of similarity with the EOG: a cross-correlation metric and mutual information. Epochs containing the main blink disturbance (0.16 seconds preceding the blink peak and 0.84 seconds following it) were extracted and the metrics were calculated on these epochs. These metrics were calculated for ICA filtered data as well as the SG-based method so that we have a reference point for comparison.

Based on a method developed by [18] during their comparison of EOG artifact rejection techniques the following cross correlation metric measuring the efficiency of blink filtering was used:

$$P = \max\{d\}(C_{CEEG,EOG}(d))\max\{d\}(C_{FEEG,EOG}(d)) \qquad (1)$$

Where $C_{CEEG,EOG}(d)$ is the normalized cross-correlation sequence between the blink-contaminated EEG and EOG at delay d and $C_{FEEG,EOG}(d)$ is the normalized cross-correlation sequence between the blink filtered EEG and the EOG at delay d.

Mutual information, a measure of the dependence between two variables, was the other metric used. While second order statistics, such as covariance and correlation, measure the linear dependence, mutual information is a more general measure for estimating not only linear dependencies, but also dependencies of higher order [19]. Furthermore correlation requires the variables to be Gaussian if their independence is to be tested, whereas mutual information

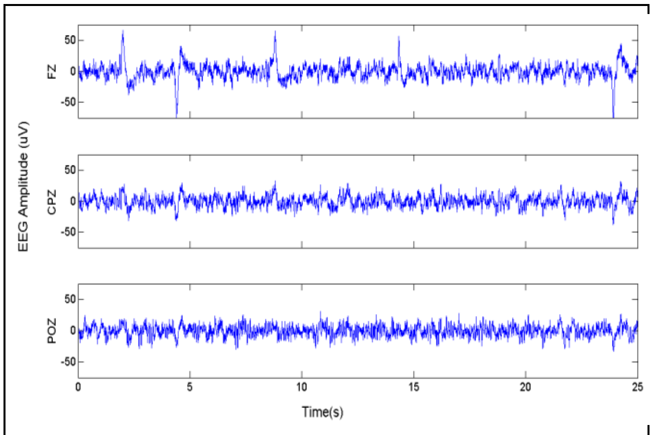

Figure 1. A 25 s segment of simulated blink-contaminated EEG. A subset of three channels out of 22 are displayed.

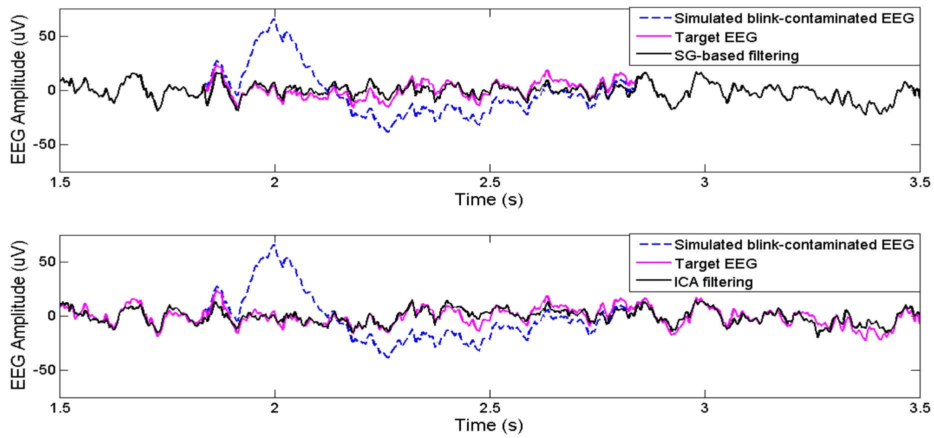

Figure 2. Filtering results on a 2 second segment of simulated blink-contaminated data at the Fz location. Shown is the simulated blink-contaminated EEG, target EEG and filtered EEG (SG-based method and ICA, respectively).

makes no assumption about the distribution. A blink carries most of the mutual information between the EEG and the EOG. Therefore, we assume that after filtering the blink from the EEG there should be less information shared with the EOG than there is between the original raw data and the EOG [20].

III. Results

A. Simulated Data

The results of filtering the simulated data at the Fz location are exemplified in "Fig. 2". By visual inspection, it can be seen that the SG-based method produces filtered EEG very similar to the target EEG. The SG-based method produced filtered data with a correlation coefficient of 0.95 (p<0.001) to the target data and ICA produced filtered data with a correlation coefficient of 0.87 (p<0.001). The root mean square (RMS) value was 9.82uV for the target data at Fz and 7.85uV for the blink activation that was added (a lower RMS value since the data is zero everywhere except where a blink occurs). The signal to noise ratio of the simulated blink-contaminated data at the Fz channel was originally 1.94 dB. After filtering, the SG-based method improved the signal-to-noise ratio (SNR) of the Fz channel to 10.41dB (with a RMS error of 2.96uV from the target data), whereas the ICA method improved it to 6.15dB (with a RMS error of 4.84uV from the target data).

B. Real Data

The performance measure, P, of filtering the real data at Fz (measured by (1)) and standard error is shown in "Fig. 3". A number close to one indicates that the filtered EEG has a largely reduced correlation with the EOG. The SG-based method presented significantly higher performance than ICA, with an average across subjects of 0.72 compared to ICA which had an average performance of 0.52.

The mutual information between raw blink-contaminated EEG at Fz and EOG was compared to the mutual information between the filtered data and EOG. "Fig. 4" shows the average mutual information with the EOG for each subject including standard error bars. The SG-based method shows a slightly larger reduction in information shared with the EOG.

As is usually the case when evaluating filtering on real data, these metrics are not without flaw; similar to how EEG can be contaminated by ocular artifacts, the EOG can be contaminated by brain activity. Therefore having filtered EEG with no correlation to the EOG after filtering should not indicate high performance, but with the performance metric 'P' it would. Similarly, a filtered EEG signal sharing zero mutual information with the EOG would not be properly filtered. However, these metrics are meant to be supplemental evaluations to the preceding section of simulated data analysis in this paper.

IV. Conclusion

Using simulated blink artifacts, the SG-based method can be seen to produce filtered data that is very similar to the target EEG. Compared to ICA, it was shown that the SG-

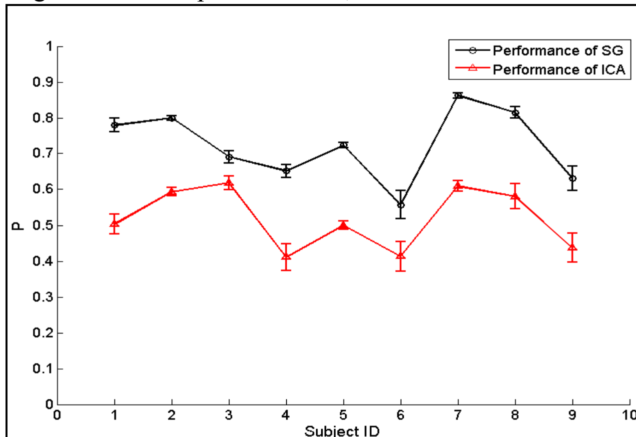

Figure 3. Average performance from (1) for each subject. A number close to one indicates that the filtered EEG has a largely reduced correlation with the EOG. The bars on each data point indicate standard error.

based method produces data that is more highly correlated to the target data, with a higher SNR. When filtering real data, the SG-based method demonstrates similar, if not higher, performance removing EOG compared to ICA.

The SG-based method of filtering is advantageous in a number of ways. The filter can be applied selectively to sections of data where a blink has been detected and leave the rest of the data intact. Since the method uses a simple smoothing filter, it is computationally efficient and can be performed quickly compared to many other methods of noise removal. The method can be easily implemented in real-time and only requires a single channel of data. ICA requires data from multiple channels and is more difficult to learn to perform. It can also be difficult to automate ICA; depending on the method used, an operator may have to examine the components obtained by ICA analysis and decide which component represents the artifact that must be removed. Furthermore, the computing of the components for ICA can take a considerable amount of time.

There are, however, certain drawbacks of the SG-based method relative to ICA. For example, ICA is able to remove more subtle or difficult to detect blink components from channels located further from the forehead. There is also the argument that regressing out EOG activity in the time or frequency domain inevitably involves subtracting a portion of the relevant EEG [21].

In this paper we have only investigated the filter's effect on eye blinks, however further work could be done by expanding the test pool to include other eye movements as well. During pre-processing of eye blink data it was noted that the high-pass filter cut off frequency strongly affects the shape of the blink. Investigation could also be carried out into how pre-processing affects the blink characteristics and the changes necessary in the filtering parameters. Finally, since the SG-based method proposes subtracting data which has been smoothed, further work could be done to investigate any information loss in the lower frequencies.

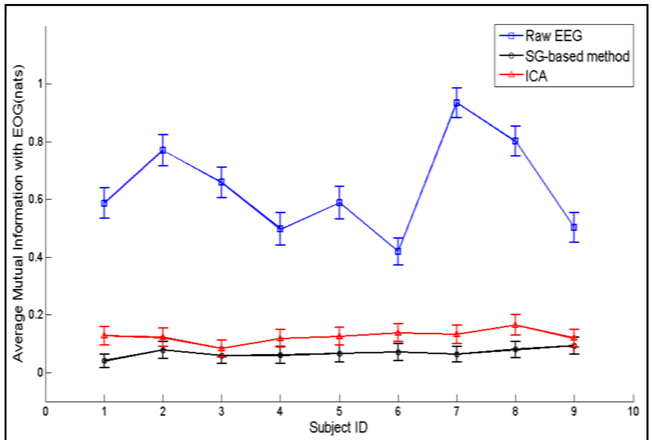

Figure 4. Average mutual information with EOG. The bars on each data point indicate standard error. A number close to zero indicates very little information is shared with the EOG.

REFERENCES

[1] Joyce, C.A., Gorodnitsky, I.F., Kutas, M. (2004). Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology. 41*:313-325.

[2] Gratton, G. (1998). Dealing with artifacts: The EOG contamination of the event-related brain potential. *Behavior Research Methods, Instruments, & Computers*. *30*: 44-53.

[3] Luo, A., Sullivan, T.J. (2010). A user-friendly SSVEP-based brain-computer interface using a time-domain classifier. *Journal of Neural Engineering*. *7*:026010.

[4] Flexer, A., Gruber, G., Dorffner, G. (2005). A reliable probabilistic sleep stager based on a single EEG signal. *Artificial Intelligence in Medicine.33*(3):199-207.

[5] Kim, H.T., Kim, B.Y., Park, E.H., Kim, J.W., Hwang, E.W., Han, S.K., Cho, S. (2005). Computerized recognition of Alzheimer disease-EEG using genetic algorithms and neural work. *Future Generation Computer Systems. 21*(7):1124-1130.

[6] Looney, D., Li, L., Rutkowski, T.M., Mandic, D.P., Cichocki, A. (2008). Ocular Artifacts Removal from EEG Using EMD. *Advances in Cognitive Neurodynamics*. Part IV: 831-835.

[7] Teixeira, A.R., Tome, A.M., Lang, E.W., Gruber, P., Martins da Silva, A. (2006) Automatic removal of high-amplitude artefacts from single-channel electroencephalograms. *Computer Methods and Programs in Biomedicine*. *83*(2):125-138.

[8] Savitzky, A., Golay, M.J.E. (1964). Smoothing and Differentiation of Data by Simplified Least Squared Procedures. *Analytical Chemistry. 36*(8):1627-1639.

[9] Jonsson, P., Eklundh, L. (2004). TIMESAT – a program for analyzing time-series of satellite sensor data. *Computers and Geosciences*. *30*: 833-845.

[10] Chinrungrueng, C., Suvichakorn, A. (2001). Fast edge-preserving noise reduction for ultrasound images. *Nuclear Science*. *48*(3):849-854.

[11] Hoffmann, A., Jager, L., Werhahn, K.J., Jaschke, M., Noachtar, S., Reiser, M. (2000). Electroencephalography during functional echo-planar imaging: detection of epileptic spikes using post-processing methods. *Magnetic Resonance in Medicine. 44*(5):791-798.

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

[13] Lee, T-W., Girolami, M., Sejnowski, T.J. (1999). Independent Component Analysis using an Extended Infomax Algorithm for Mixed Sub-Gaussian and Super-Gaussian Sources. *Neural Computation*. *11*(2): 417-441.

[14] Blankertz, B. (2008). BCI Competition IV. Retrieved from http://www.bbci.de/competition/iv/

[15] Brunner, C., Leeb, R., Muller-Putz, G.R., Schlogl, A., Pfurtscheller, G. (2008). BCI Competition 2008 – Graz data set A. Retrieved from http://www.bbci.de/competition/iv/desc_2a.pdf.

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

[17] Bentivoglio, A.R., Bressman, S.B., Casetta, E., Carretta, D., Tonali, P., Albanese, A. (1997). Analysis of blink rate patterns in normal subjects. *Movement Disorders*. *12*(6): 1028-1034. Berg, P., Scherg, M. (1991). Dipole models of eye movements and blinks. *Electroencephalography and Clinical Neurophysiology*. *79*: 36-44.

[18] Klados, M.A., Papadelis, C., Lithari, C.D., Bamidis, P.D. (2009). The Removal of Ocular Artifacts from EEG Signals: A Comparison of Performances for Different Methods. *IFMBE Proceedings*. *22*(10): 1259-1263.

[19] Li, W. (1990). Mutual Information Functions versus Correlation Functions. *Journal of Statisical Physics. 60*(5/6):823-837.

[20] Hoffmann, S., Falkenstein, M. (2008). The correction of eye blink artefacts in the EEG: a comparison of two prominent methods. *PLoS One. 3*(8):e3004.

[21] Oster, P.J., Stern, J.A. (1980). Measurement of eye movement electrooculography. *Techniques in Psychophysiology*. Wiley, Chichester, 275-309.