30th Annual International IEEE EMBS Conference
Vancouver, British Columbia, Canada, August 20-24, 2008

Quantitative Evaluation of Ocular Artifact Removal Methods Based on Real and Estimated EOG Signals

Borna Noureddin, Peter D. Lawrence, *Senior Member, IEEE*, Gary E. Birch, *Member, IEEE*

Abstract— **We propose a novel metric for quantitatively evaluating ocular artifact (OA) removal methods on real electroencephalogram (EEG) data. For real EEG, existing metrics measure the amount of artifact removed. Our metric measures how much a given method is likely to distort the underlying EEG. The new metric was used to evaluate two existing OA removal algorithms that use the electro-oculogram (EOG) as a reference signal. The combination of a previous metric and our new metric showed there is a trade-off between how well an algorithm removes OAs and how likely it is to distort the underlying EEG. These algorithms require a reference EOG signal, yet for certain applications (e.g., a brain computer interface or BCI) it is preferable or necessary to avoid attaching electrodes around the eyes. We thus also used various combinations of up to 55 channels of EEG to estimate the EOG reference. The metric was again used to compare the use of estimated vs. measured EOG. Our initial results showed that using EOG estimated from as few as 4 EEG electrodes increased the likelihood of distorting the EEG from 14% to 19% and from 21% to 23% for the two algorithms. For some applications (e.g., BCI), the slight reduction in performance may be acceptable in order to avoid using EOG electrodes.**

I. Introduction

Eye movements and blinks pose a serious problem for electroencephalogram (EEG) measurements, and their removal from measured EEG is an ongoing research problem [1][2][3][4]. Almost all current approaches to ocular artifact (OA) detection and removal [4][5][6][7][8][9] use one or more electro-oculogram (EOG) signals either directly or indirectly as a reference. A few remove OAs without the need for a reference EOG signal. These approaches, however, are often not fully automated (e.g., require manual selection of either a threshold [10] or of components to reject [11]), are not able to handle all sources of OA (e.g., only blinks [10]) or are not suitable for real-time use (e.g., require large amounts of data [1]).

In all cases, the performance of the OA removal (OAR) method is normally evaluated using simulated data. Artifact-free EEG data and an artifact signal are combined artificially and processed using the OAR algorithm. Some feature of the output of the algorithm (e.g., signal-to-noise ratio, or SNR) is compared to the original artifact-free EEG. For real EEG, the artifact-free ("true") EEG is not known, so the performance of the algorithm on real data is usually reported subjectively [2][3][12], often based on visual inspection of the resulting waveforms. Puthusserypady et. al. [13] measured the ratio of the power of the artifact signal removed to the EEG signal remaining as a metric for real data, proposing that the higher the ratio, the better the performance of the algorithm. They assumed, however, that the algorithm is only being applied to data with significant OA. For data that does not contain OA, a higher ratio is not necessarily indicative of better performance. Therefore, a metric that evaluates the performance of an OAR algorithm consistently on data that has periods both with and without OA is needed.

In this paper, we propose a new metric that indicates how much an OAR algorithm may distort the underlying EEG. When combined with the power ratio of Puthusserypady et. al. [13], the metric can be used to effectively evaluate OAR algorithms on real EEG data. We applied the metric to evaluate two OAR algorithms ([5] and [7]) that are online, fully automated, and shown to perform well during OAs. We found a trade-off between how well each algorithm removes OAs and how likely it is to distort the EEG.

In addition, like most OAR methods, the algorithms used require a reference EOG signal. However, for real-world, real-time, online applications (e.g., a brain computer interface or BCI), it is highly desirable (and sometimes even a requirement) not to place electrodes on the subject's face. Although methods exist to remove OAs, in principle, without the EOG signal [3][11], they are only suitable for offline analysis. We therefore used a linear combination of available EEG channels to estimate the EOG signal required by many OAR algorithms, thus providing an alternative to directly measuring the EOG. We explored various combinations of EEG channels, and report in this paper the effect of using estimated (as opposed to measured) EOG on the performance of the above two OAR algorithms ([5] and [7]).

Manuscript received April 7, 2008. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grant 90278-06 and Discovery Grant 4924-05, in part by Canadian Institutes of Health Research under Grant MOP-72711, and in part by Precarn IRIS NCE.

Borna Noureddin is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada V6T 1Z4 (corresponding author; ph: 604-822-9215; fax: 604-822-5949; e-mail: bornan@ece.ubc.ca).

Peter D. Lawrence is with the Department of Electrical and Computer Engineering and Institute for Computing, Information and Cognitive Systems, University of British Columbia, Vancouver, BC, Canada V6T 1Z4 (e-mail: PeterL@ece.ubc.ca).

Gary E. Birch is with the Neil Squire Society, Burnaby, BC, Canada V5M 3Z3 and the Department of Electrical and Computer Engineering and Institute for Computing, Information and Cognitive Systems, University of British Columbia, Vancouver, BC, Canada V6T 1Z4 (e-mail: garyb@neilsquire.ca).

II. METHODS

A. Performance metric

Typically, OAR methods are evaluated using simulated data. The top half of Figure 1 shows the approach generally taken for OAR. **A** (ocular artifact), **N** (noise) and **B** (true EEG) are generated, and the resulting **Y** (EEG with artifact removed) is compared to **B**. For example, in [2], $\mathbf{E_e}$ (EEG electrode signal), **Y**, and **B** are used to measure how well an algorithm removes OAs from measured EEG in simulations. For real data, however, **A**, **N** and **B** are unknown, making such a metric unsuitable.

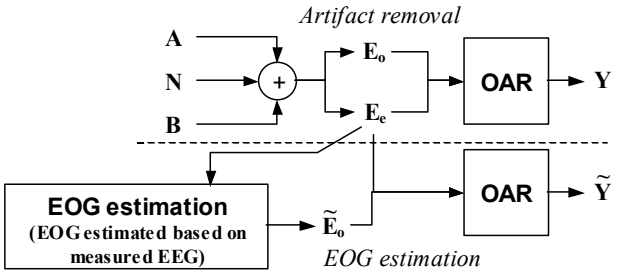

Figure 1 OA removal & EOG estimation: **A** is the ocular artifact signal, **B** is the underlying brain signal, **N** is measurement noise, $\mathbf{E_o}$ is the signal measured at EOG electrode sites, $\mathbf{E_e}$ is the signal measured at EEG electrode sites, "OAR" is the OA removal algorithm, and **Y** and $\tilde{\mathbf{Y}}$ are the estimations of the true EEG at the EEG electrode sites. Ideally, $\mathbf{Y}=\tilde{\mathbf{Y}}=\mathbf{B}$.

Puthusserypady et. al. [13] proposed the following metric for real EEG:

$$R = \frac{\sum_{1}^{C}\sum_{1}^{N}(\mathbf{E_e}-\mathbf{Y})^2}{\sum_{1}^{C}\sum_{1}^{N}\mathbf{Y}^2} \quad (1)$$

where N is the number of samples and C the number of channels recorded. The numerator represents the power of the artifact signal removed from all EEG electrodes, and the denominator the power of the remaining EEG. During an artifact, higher R values are better; otherwise, lower values of R are desirable. To use an extreme example, if the OAR algorithm considered the entire signal to be artifact, $Y \to 0$ and $R \to \infty$: R is maximal, but clearly the algorithm has poor performance. The evaluation needs to measure how much the algorithm distorts the EEG. If the denominator measures the power of the measured EEG as follows:

$$R' = \frac{\sum_{1}^{C}(\mathbf{E_e}-\mathbf{Y})^2}{\sum_{1}^{C}\mathbf{E_e}^2} \quad (2)$$

then whenever $R'>1$, the power in the removed artifact signal is higher than that of the original EEG, indicating that the algorithm has likely removed too much signal or introduced new artifacts. The more often this occurs, the worse the performance of the algorithm. Therefore, we measured both R (how well an algorithm removes artifacts) and the percentage of samples ε in which $R'>1$ (how often it removes too much signal).

The metrics were calculated for two OAR algorithms. The RLS algorithm described in [5] has fast convergence and is suitable for online use. The time-varying H^∞ algorithm of [7] converges more slowly but was previously shown to perform slightly better than the RLS algorithm. The implementation found in [14] was used for both algorithms.

B. EOG approximation

As shown in the bottom half of Figure 1, the measured EEG can be used to estimate the signals measured simultaneously at the EOG sites as follows:

$$\mathbf{w} = \mathbf{E_o} \cdot \mathbf{E_e}^{\#}$$
$$\tilde{\mathbf{E}}_\mathbf{o} = \mathbf{w} \cdot \mathbf{E_e} \quad (3)$$

where $\mathbf{E_e}^{\#}$ is the pseudo-inverse of $\mathbf{E_e}$ and **w** is calculated using training data. The calculated **w** is then used on the remaining test data to calculate $\tilde{\mathbf{E}}_\mathbf{o}$. The choice of training and test data (both of which contain both contaminated and artifact-free EEG) is described below.

C. Data Collection

In order to test the new metric ε and the EOG estimation, 57 channels of EEG (including 2 linked mastoids – see Figure 2) and 7 channels of EOG (one above and below each eye, one on each outer canthus and one on the nasion) were collected from one subject, sitting approximately 50cm in front of a computer monitor. All signals were sampled by the same amplifier at 200Hz, with a low-pass filter of 70Hz. The signals were further filtered digitally at 30Hz.

Two sessions of data were collected, and the subject was instructed to perform 9 tasks in each session. For the first task, the subject was shown a picture of clouds and asked to relax. During the second task, the subject was shown a picture with hidden faces, and asked to count the number of faces (mental task). Three sets of 45 seconds of each of the first 2 tasks were collected for each session. During the third task, a dot was displayed in the centre of the computer monitor, and the subject was asked to stare at the dot for 45 seconds. This was repeated six times for a total of 270 seconds per session. During the fourth task, a 4x4 grid of numbers was displayed on the monitor. Each number was sequentially highlighted for 2 seconds, and the subject was instructed to follow the highlighted number. During the fifth task, the same grid was shown, but the subject was instructed to blink at least once during each 2-sec interval. The sixth and seventh tasks were the same as the fourth and fifth, respectively, but the subject was instructed to also perform a hand extension during each 2-sec interval. A total of 16 trials per session of 32 seconds each were collected from the subject for this set of tasks. During the eighth task, a small red dot was shown sequentially for 2 seconds on each corner

of the computer monitor, and the subject was instructed to follow the dot. The ninth task was the same as the eighth task, except the subject was instructed to blink at least once during each 2-sec interval. A total of 16 trials per session of 24 seconds each were collected from the subject for this set of tasks. The resulting data consisted of EEG and EOG measured during a range of mental states, including periods with and without blinks and eye movements, taken over two sessions collected on different days. The first trial of each task (training data) was used to calculate $\mathbf{w}$, which was then used on the remaining trials (test data) to calculate $\tilde{\mathbf{E}}_o$.

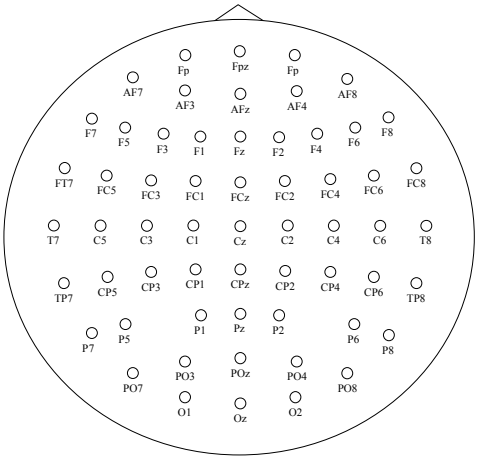

Figure 2 EEG electrodes

III. Results

A. Performance metric

TABLE 1 shows average values for R and ε for each of the two algorithms for each of the two sessions separately, as well as the overall average over all sessions. In each case, the results are shown for all the data collected, for only those segments known to contain ocular artifacts, and for only those segments not containing artifacts. Each EOG channel was band-pass filtered between 1-2Hz. Segments where the magnitude of the vertical EOG (difference between the electrodes above and below each eye) was greater than 10μV or the magnitude of the horizontal EOG (difference between each outer canthus and nasion electrodes) was greater than 15μV were marked as containing artifacts. All other segments were considered artifact-free. The results consistently show that for higher R values (better performance according to [13]), the ε value is higher, demonstrating that there is a tradeoff between how much artifact an algorithm removes and how much it risks introducing new artifacts.

B. EOG estimation

The first trial of each of the 9 tasks was used to calculate $\mathbf{w}$ in (3) using all 55 channels of EEG. The resulting $\mathbf{w}$ was then used to calculate $\tilde{\mathbf{E}}_o$ and ultimately $\tilde{\mathbf{Y}}$. The values for R and ε were then calculated using $\tilde{\mathbf{Y}}$ (see TABLE 2). Here, "all sessions" means that $\mathbf{w}$ was calculated using data from the first trials of session 1 only, but applied to both session 1 and session 2 data for EOG estimation and OA removal. This shows the result of using a subject-specific (as opposed to session-specific) mapping of EEG to EOG. Interestingly, the value for R is consistently *higher* when using estimated EOG, which might lead one to believe that using the estimated EOG is actually better than using measured EOG. However, the higher value for ε in each case shows that such a conclusion would be erroneous, since the estimation of the EOG causes both algorithms to introduce more artifacts.

	All data		OA segments only		Non-OA segments only	
	R	**ε (%)**	**R**	**ε (%)**	**R**	**ε (%)**
RLS session 1	2.4	16	3.3	23	0.7	10
RLS session 2	1.6	12	2.6	20	0.3	6
RLS all sessions	2.0	14	2.9	22	0.5	8
H^∞ session 1	12.0	23	19.0	29	2.1	17
H^∞ session 2	5.1	20	9.9	27	1.3	14
H^∞ all sessions	8.7	21	14.0	28	1.7	15

TABLE 1 COMPARISON OF METRICS, USING ACTUAL MEASURED EOG.

	All data		OA segments only		Non-OA segments only	
	R	**ε (%)**	**R**	**ε (%)**	**R**	**ε (%)**
RLS session 1	3.3	22	4.3	26	1.0	18
RLS session 2	2.9	21	4.3	26	1.1	17
RLS all sessions	2.8	20	3.8	25	0.9	16
H^∞ session 1	18.0	26	27.0	30	4.0	22
H^∞ session 2	8.4	24	15.0	28	2.9	20
H^∞ all sessions	14.0	25	22.0	29	3.7	21

TABLE 2 RESULTS USING EOG APPROXIMATED FROM 55 EEG CHANNELS.

The choice of EEG electrodes used to estimate the EOG affects the quality of the estimation. Therefore, R and ε were determined for $\mathbf{w}$ and $\tilde{\mathbf{Y}}$ calculated using a variety of combinations of EEG electrodes: all 55 electrodes (full head); Fp, AF, F, FC and C electrodes (anterior); Fp, AF and F electrodes (frontal); Fp and AF electrodes; and Fpz, Cz, AF7 and AF8 electrodes. The results are shown in TABLE 3 and TABLE 4 for the RLS and H^∞ algorithms, respectively, for "all sessions" as described above.

IV. Discussion

The results show that the H^∞ algorithm removes more OA (R is consistently higher in TABLE 1) than the RLS algorithm, as reported in [7], but that it may also distort the EEG more (ε is consistently higher in TABLE 1). The choice of algorithm will depend on whether the preference is to maximize OA removal or minimize EEG distortion. In addition, for both algorithms, R is consistently lower when there is no OA and higher when there is OA, which confirms that the algorithms remove more when an OA exists. However, during OA removal, both algorithms also potentially remove more EEG signal on average during periods with OA (ε is higher).

A comparison of R and ε for the algorithms using both measured and estimated EOG signals shows that, when

estimating EOG, more OA is removed (R is higher for estimated EOG than for measured EOG) and the likelihood of distorting the EEG is increased from 14% and 21% using measured EOG to anywhere between 18-21% and 23-26% (see "All data" in TABLE 3 and TABLE 4) for the RLS and H^{∞} algorithms, respectively. In both cases, using estimated EOG affects the performance when there are no artifacts (see "Non-OA segments only" in TABLE 3 and TABLE 4). Ultimately, whether the effect is sufficiently small depends on the application.

	All data		OA segments only		Non-OA segments only	
	R	ε (%)	R	ε (%)	R	ε (%)
EOG	2.0	14	2.9	22	0.5	8
Full head	2.8	20	3.8	25	0.9	16
Anterior only	2.9	21	4.0	25	0.9	17
Frontal only	3.3	21	4.5	24	1.1	17
Fp/AF only	3.2	18	4.3	22	1.1	14
Fpz+Cz+AF7+AF8	5.3	19	6.9	23	2.3	16

TABLE 3 RESULTS OF USING MEASURED EOG AND EOG APPROXIMATED FROM DIFFERENT CONFIGURATIONS OF EEG USING THE RLS ALGORITHM.

	All data		OA segments only		Non-OA segments only	
	R	ε (%)	R	ε (%)	R	ε (%)
EOG	8.7	21	14.0	28	1.7	15
Full head	14.0	25	22.0	29	3.7	21
Anterior only	14.0	26	22.0	30	3.7	22
Frontal only	13.0	25	21.0	29	3.5	21
Fp/AF only	12.0	23	18.0	26	3.2	19
Fpz+Cz+AF7+AF8	15.0	23	22.0	27	4.7	20

TABLE 4 RESULTS OF USING MEASURED EOG AND EOG APPROXIMATED FROM DIFFERENT CONFIGURATIONS OF EEG USING THE H^{∞} ALGORITHM.

Overall, the H^{∞} algorithm is less affected by using estimated EOG, regardless of the combination of EEG electrodes used. For both algorithms, using the Fpz, Cz, AF7 and AF8 electrodes seems to give the best estimation results.

V. CONCLUSIONS

The preliminary results reported in this paper demonstrate that our new metric, which measures the likelihood of an OAR algorithm to distort the EEG, effectively complements a previous metric, which measures how much artifact is removed. Both metrics are suitable for use with real EEG data, and both are needed to properly evaluate the performance of OAR algorithms. Using the metrics with two existing OAR algorithms showed that the more OA each algorithm removes, the more likely it is to distort the EEG.

The two algorithms evaluated, like most other OAR algorithms, require a reference EOG signal, yet for some applications (e.g., BCI), it is preferable or necessary to avoid attaching electrodes around the eyes. The new metric was therefore also used to assess how much estimating the reference EOG signal from EEG electrodes (instead of measuring the EOG directly) would affect the performance of the selected algorithms. The results show that using four specific EEG electrodes (Fpz, Cz, AF7 and AF8) to estimate the reference EOG signal increased the likelihood of distorting the EEG from 14% to 19% for one algorithm, and from 21% to 23% in the other, which may be sufficiently low to be used in applications where it is not desirable or possible to use EOG electrodes.

The new metric can be used to analyze data from more subjects to verify the results above. Also, since OA removal is a pre-processing step for other applications, the ultimate performance of OAR algorithms will need to be measured in the context of specific applications. For example, analyzing the error rate or false positive rate of a BCI would provide an additional, practical measure of the effectiveness of OAR algorithms on real data.

REFERENCES

[1] G. Gomez-Herrero, W. De Clercq, H. Anwar, O. Kara, K. Egiazarian, S. Van Huffel, W. Van Paesschen, "Automatic Removal of Ocular Artifacts in the EEG without an EOG Reference Channel," *Proceedings of the 7th Nordic Signal Processing Symposium, NORSIG 2006,* pp. 130-133, 2006.

[2] J. J. M. Kierkels, G. J. M. van Boxtel and L. L. M. Vogten, "A model-based objective evaluation of eye movement correction in EEG recordings," *Biomedical Engineering, IEEE Transactions on,* vol. 53, pp. 246-253, 2006.

[3] K. Ting, P. Fung, C. Chang, F. Chan, "Automatic correction of artifact from single-trial event-related potentials by blind source separation using second order statistics only," *Medical engineering physics 28(8),* vol. 28, pp. 780-794, 2006.

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

[5] P. He, G. Wilson and C. Russell, "Removal of ocular artifacts from electro-encephalogram by adaptive filtering," *Med. Biol. Eng. Comput.,* vol. 42, pp. 407-412, 2004.

[6] T.J. Liu, D.Z. Yao, "Removal of the ocular artifacts from EEG data using a cascaded spatio-temporal processing," *Computer methods and programs in biomedicine 83(2),* vol. 83, pp. 95-103, 2006.

[7] S. Puthusserypady, T. Ratnarajah, "H^{∞} adaptive filters for eye blink artifact minimization from electroencephalogram," *Signal Processing Letters, IEEE,* vol. 12, pp. 816-819, 2005.

[8] A. Schlogl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," *Clinical Neurophysiology,* vol. 118, pp. 98-104, 2007.

[9] D. Talsma, "Auto-adaptive averaging: Detecting artifacts in event-related potential data using a fully automated procedure," *Psychophysiology,* vol. 45, pp. 216-228, 2008.

[10] Y. Li, Z. Ma, W. Lu and Y. Li, "Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach," *Physiol. Meas.,* vol. 27, pp. 425-436, 2006.

[11] C.A. Joyce, I.F. Gorodnitsky, M. Kutas, "Automatic removal of eye movement and blink artifacts from EEG data using blind component separation," *Psychophysiology 41(2),* vol. 41, pp. 313-325, 2004.

[12] N. Ille, P. Berg, M. Scherg, "Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies," *Journal of clinical neurophysiology 19(2),* vol. 19, pp. 113-124, 2002.

[13] S. Puthusserypady, T. Ratnarajah, "Robust adaptive techniques for minimization of EOG artefacts from EEG signals," *Signal processing 86(9),* vol. 86, pp. 2351-2363, 2006.

[14] A. Delorme, S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods 134(1),* vol. 13, pp. 9-21, 2004.