35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

An Automatic Electroencephalography Blinking Artefact Detection and Removal Method Based on Template Matching and Ensemble Empirical Mode Decomposition

Paschalis A. Bizopoulos, Tarik Al-Ani, Dimitrios G. Tsalikakis, Alexandros T. Tzallas, Dimitrios D. Koutsouris, *Senior Member, IEEE* and Dimitrios I. Fotiadis, *Senior Member, IEEE*

Abstract— **Electrooculographic (EOG) artefacts are one of the most common causes of Electroencephalogram (EEG) distortion. In this paper, we propose a method for EOG Blinking Artefacts (BAs) detection and removal from EEG. Normalized Correlation Coefficient (NCC), based on a predetermined BA template library was used for detecting the BA. Ensemble Empirical Mode Decomposition (EEMD) was applied to the contaminated region and a statistical algorithm determined which Intrinsic Mode Functions (IMFs) correspond to the BA. The proposed method was applied in simulated EEG signals, which were contaminated with artificially created EOG BAs, increasing the Signal-to-Error Ratio (SER) of the EEG Contaminated Region (CR) by 35dB on average.**

I. Introduction

Electroencephalogram (EEG) is a noninvasive measurement of the brain's electrical activity obtained from electrodes located on the scalp over multiple areas of the head. Over the last decades and under an increasing medical need, EEG became an important diagnostic tool for monitoring and managing dysfunctions and various neurological disorders of the human brain.

Electrical sources other than the brain, which interfere during EEG acquisition, are called artefacts. Artefacts usually have higher amplitude than the EEG and their frequency content overlaps with the cortical signal's content. This results in the distortion of the basics characteristics of EEG affecting the results of its analysis. Therefore preprocess of Artefact Detection and Removal (ADR) is necessary before using EEG.

A main problem of the analysis of EEG is the contamination of the recordings from artefacts created from eye movements and Blinking Artefacts (BA). This kind of noise is generally called Electrooculographic (EOG) artefact. The difficulty to immobilize eye movements for a certain period of time makes EOG removal a prime necessity. The filtering procedure of that noise is extremely difficult because EOG has a frequency spread that overlaps with the frequency of EEG.

In order to detect and remove the presence of EOG and general muscle movement artefacts a variety of methods have been proposed. Among these methods Independent Component Analysis (ICA) [1] and Autoregressive Models (AR-models) [2] promised EOG reduction but quantitative comparison on a representative dataset was unavailable.

Ensemble Empirical Mode Decomposition (EEMD), which was proposed from Huang [3, 4], has been applied for analyzing non-stationary and non-linear biosignals. More specifically, in EEG signals EEMD was applied in several problems as in seizure classification [5], and alpha rhythms analysis [6], while a variation of EMD that combines Phase Locking (EMDPL) was used for synchronization detection [7].

Several studies on EOG reduction suggested that EEMD in comparison with other methods like ICA [8] and wavelet transform [9] performed sufficiently reliable EOG correction. EEMD has also been combined with other traditional signal analysis methods in removing various types of muscle and EOG artefacts with satisfactory results [11].

The aim of this study is to present a method for BAs detection and removal. The general idea at the basis of the present study is the combination of template matching (TM), using Normalized Correlation Coefficient (NCC) and EEMD as a tool for detecting and eliminating EOG BAs. The automatic nature of the method relies on the number of predefined Templates stored in the Artifact Template Library.

II. Methods

A. *Template Matching using Normalized Correlation Coefficient*

The first step of ADR is the detection of the Contaminated Regions (CRs) that contain BAs. We used NCC based on a predefined template since it is a measure of

P.A. Bizopoulos is with the Départment Informatique et Télécommunication, ESIEE-Paris, Cité Descartes-BP 99, 93162, Noisy-Le-Grand, France (e-mail: bizopoulos.paschalis@gmail.com).

T. Al-Ani is with LISV-Université de Versailles Saint-Quentin-en-Yvelines, 10-12 5 Av.de l'Europe, 78140, Vélizy, France and also with the Départment Informatique et Télécommunication, ESIEE-Paris, Cité Descartes-BP 99, 93162, Noisy-Le-Grand, France.

D.G. Tsalikakis is with the Department of Informatics and Telecommunications Engineering, University of Western Macedonia, Kozani GR 50 100, Greece (e-mail: dtsalikakis@uowm.gr).

A.T. Tzallas is with the Unit of Medical Technology and Intelligent Information Systems, Ioannina GR 45 110, Greece (e-mail: atzallas@cc.uoi.gr).

D.D. Koutsouris is with the Biomedical Engineering Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), GR-157 73 Zografou, Athens, Greece (e-mail: dkoutsou@biomed.ntua.gr).

D.I. Fotiadis is with the Unit of Medical Technology and Intelligent Information Systems, Faculty of Material Science and Engineering, University of Ioannina, Ioannina GR 45 110, Greece and also with the Biomedical Research Institute, FORTH, Ioannina GR 45 110, Greece (corresponding author: +30-26510-98803; fax: +30-26510-97092; e-mail: fotiadis@cs.uoi.gr).

the quantity of similarity of two signals and BAs, in most cases, share the same characteristic waveform. Therefore NCC can be seen as a quantitative comparison measure of the covariance degree between two signals. The NCC is given from Eq.(1), where $y(t)$ is a signal of N_y samples length, $\bar{y}$ is the mean value of the signal, $x(t)$ is an template of $N_x \leq N_y$ sample length, $\bar{x}$ is the mean value of the template and $t_d \leq N_y$ is a variable indicating the position of $x(t)$ inside $y(t)$:

$$NCC(t_d) = \frac{\sum_{t=1}^{N_x}\left(\left(x(t) - \bar{x}\right)\left(y(t - t_d) - \bar{y}\right)\right)}{\sqrt{\left(\sum_{t=1}^{N_x}(x(t) - \bar{x})^2\right)\left(\sum_{t=1}^{N_y}(y(t - t_d) - \bar{y})^2\right)}}. \quad (1)$$

The choice of a position t_d as a possible start of the CR depends on whether the following criterion described in Eq.(2) is satisfied. NCC_{thr} is the NCC threshold for which holds $NCC_{thr} \leq 1$:

$$NCC_{thr} \leq NCC(t_d). \quad (2)$$

B. Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) is a locally adaptive method of decomposing non-stationary multicomponent signals. As its name signifies, it is an empirical method since it is not mathematically formulated, but it is quite effective in analyzing non-stationary and non-linear signals. The EMD's basic principle is relatively simple and its advantages make it a powerful advanced signal processing technique.

EMD generates a set of approximately orthogonal time oscillations of decreasing frequency (with respect to the inner product), that coexist in the signal, called Intrinsic Mode Functions (IMFs) [3, 4]. IMFs are monocomponent signals having a well-behaving Hilbert Transform and so their instantaneous frequencies can be calculated [12].

$N_e, N_z \in \mathbb{N}$ are defined as the extrema and the zero crossings of a function respectively. $h_{\max}(t)$ and $h_{\min}(t)$ are defined as the upper and lower envelopes of that function. An IMF is a function which satisfies [4]:

$$N_z - 1 \leq N_e \leq N_z + 1, \quad (3)$$

$$\frac{\left|h_{\max}(t) + h_{\min}(t)\right|}{2} = 0 \text{ , } \forall t \in [0,T]. \quad (4)$$

Given a discretely sampled signal $x(t)$, of a finite cardinality, on a finite interval $[0,T] \subset \mathbb{R}$, the EMD algorithm can be applied [3, 4] and $x(t)$ can be represented as:

$$x(t) = \sum_{k=1}^{n} imf_k(t) + r(t), \quad \forall t \in [0,T], \quad (5)$$

where $r(t)$ is the monotonous function that results from the EMD algorithm and is considered as the residue.

C. Ensemble Empirical Mode Decomposition

The second step of ADR is the removal of the BAs from the CRs and for that purpose we used EEMD. Ensemble EMD (EEMD) is a Noise Assisted Data Analysis (NADA) method that was proposed in order to overcome the difficulties that EMD faces [3].

Mode mixing is the simultaneous presence of widely different rhythms in the same IMF. This problem often appears in EMD and results in disabling the physical interpretation of the IMFs.

The problems of mode mixing and unstable IMFs are significantly reduced with EEMD. EEMD averages the IMFs, obtained by EMD, by applying independent series of Gaussian white noise, with a prescribed standard deviation N_{std}, to the signal.

The EEMD algorithm can be applied [3] and $x(t)$ can be represented as the average over the ensemble:

$$imf_k(t) = \frac{1}{NE}\sum_{m=1}^{NE} imf_{k,m}(t), \quad k \in [1,n], \ \forall t \in [0,T] \quad (6)$$

where NE is the Number of Ensemble, $imf_{k,m}(t)$ is the set of the IMFs of the m[th] iteration of EEMD.

III. PROPOSED METHOD

$EEG_r(t)$ is defined as the recorded EEG signal of length N. $EEG_{r_s}(t)$ is defined as the s^{th} segment of $EEG_r(t)$ of length N_s.

A. Contaminated Region Detection

NCC_{thr} is defined as the user-defined threshold and $Temp_u(t) \subset EEG_{r_s}(t)$ is the user-defined template of the u^{th} BA of N_{Temp_u} length sample. So $I_{t_{d_w}}^{N_{Temp_u}}$ are defined as the possible intervals of the CRs, with origin t_{d_w} and length N_{Temp_u}, which using Eq.(1) and (2), satisfy:

$$NCC_{thr} \leq NCC(t_{d_w}) \Rightarrow$$

$$NCC_{thr} \leq \frac{\sum_{t=1}^{N_{Temp_u}}\left(\left(Temp_u(t) - \overline{Temp_u}\right)\left(EEG_{r_s}(t - t_{d_w}) - \overline{EEG_{r_s}}\right)\right)}{\sqrt{\left(\sum_{t=1}^{N_{Temp_u}}(Temp_u(t) - \overline{Temp_u})^2\right)\left(\sum_{t=1}^{N_s}(EEG_{r_s}(t - t_{d_w}) - \overline{EEG_{r_s}})^2\right)}}. \quad (7)$$

B. Contaminated Region Decomposition

$EEG_{reg_s}(t) \subset EEG_{r_s}(t)$ are defined as the regions at the intervals $I_{t_{d_w}}^{N_{Temp_u}}$ of $EEG_{r_s}(t)$. Therefore by applying EEMD on the CRs we obtain:

$$EEG_{reg_s}(t) = \sum_{k=1}^{n} imf_k(t) + r(t). \quad (8)$$

The NE and the amplitude of the white noise (Nstd) added during EEMD are chosen according to [3].

C. Blinking Artefact Rejection

BAs could be considered as the trend that the EEMD detects in the detected CRs:

$$EEG_{reg_s}(t) = EEG_c(t) + Arti_{mu}(t). \tag{9}$$

N_{imf_k} is defined as the length and $\overline{imf_k}$ the mean of the k^{th} IMF. The Standard Deviation (SD) σ_{imf_k} of the kth IMF is then calculated:

$$\sigma_{imf_k} = \sqrt{\frac{1}{N_{imf_k}} \sum_{i=1}^{N_{imf_k}} \left(imf_k(i) - \overline{imf_k} \right)^2}. \tag{10}$$

The center of the BA Rejection algorithm relies on prescribing a value for $0 < p \leq 1$ so that the relative difference of the SDs of two successive IMFs is satisfied:

$$\frac{\sigma_{imf_D} - \sigma_{imf_{D-1}}}{\sigma_{imf_D}} > p. \tag{11}$$

Thus the problem is reduced to: $\exists D$, $2 \leq D \leq n+1$, where D is the number of IMFs prior contamination by the trend of the CR. So we obtain:

$$EEG_c(t) = \sum_{i=1}^{D} imf_i(t), \tag{12}$$

$$Arti_{mu}(t) = \sum_{i=D+1}^{n} imf_i(t) + r(t). \tag{13}$$

This method through one full circle is able to detect all the b^{th} kind of BAs. It has to be mentioned that b, which is the kind of the BA, can be defined manually by the user (Manual BA Detection) or automatically using the BA Template Library (Automatic BA Detection). $Art_{mb}(t)$ is defined as the m^{th} instance of the b^{th} BA kind, $EEG'_{c_s}(t)$ is the contaminated EEG from other kind of BAs and L_b is the number of the b^{th} kind of BAs. $EEG'_{c_s}(t)$ is free from the b^{th} kind of BA. This process can be modeled as follows:

$$EEG_{r_s}(t) = EEG'_{c_s}(t) + \sum_{m=1}^{L_b} Arti_{mb}(t). \tag{14}$$

D. Final Output

CRs are handled as the linear combination of the EEG from cerebral activity and the sum of all BA activity. $EEG_{c_s}(t)$ is defined as the clean EEG and K is the number of different kind of BAs. Therefore the $EEG_{r_s}(t)$ can be modeled as:

$$EEG_{r_s}(t) = EEG_{c_s}(t) + \sum_{b=1}^{K} \sum_{m=1}^{L_b} Art_{mb}(t). \tag{15}$$

The execution of the method is finished when all the segments of $EEG_r(t)$ are processed. The flowchart of the purposed method is shown in “Fig. 1”.

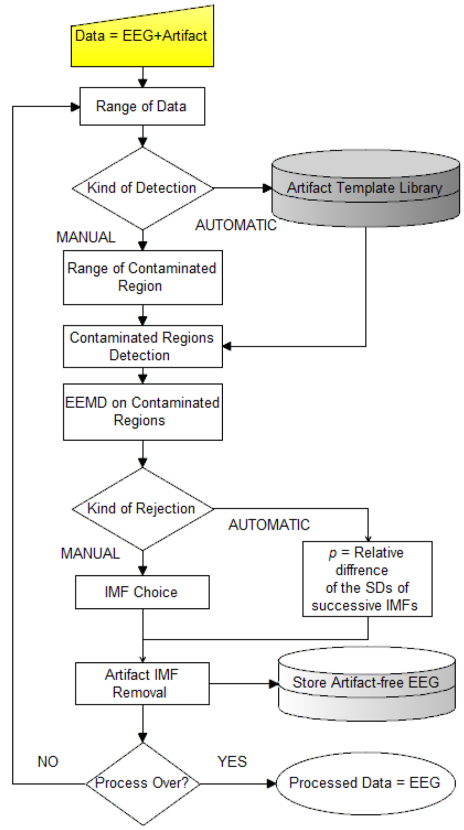

Figure 1. Flowchart of the purposed method

IV. Results

Simulated EEG Data [13] was used with 10 trials. Each trial contains 1000 samples with a sampling frequency 250Hz. The artificially generated BAs had amplitude of $A_{Art} = -56mV$ and a normalized $A_{Art_{Norm}} = -0.747$ (with respect to the maximum amplitude $A_{\max} = -74.9657mV$ that appeared in the segment). Their frequency was 2 Hz and their peaks appeared every 1000 samples. This resulted into 10 BAs in the generated signal.

For detecting the BAs, the NCC threshold was set to $NCC_{thr} = 0.9$ in order to decrease the possibility of false positive CR detection. The Manual method was chosen with a BA template from the 1050th to the 1160th sample. For decomposing the CRs, EEMD was used by setting $N_{std} = 0.55$ and $NE = 500$.

The selections of the IMFs that correspond to the BAs were applied by setting p in a range of values $0.35 < p < 0.55$ and were subtracted from the original data. The characteristic of this subtraction was that only the first 3 IMFs out of 6, including the residue, were kept as useful EEG data. The same method was tested for real EEG contaminated with BAs and for values $p \approx 0.74$ the rejection was satisfactory with variable number of IMFs kept for each CR.

A sub-segment of the EEG signal and the respective outputs from CR detection and BA rejection are shown in “Fig. 2”.

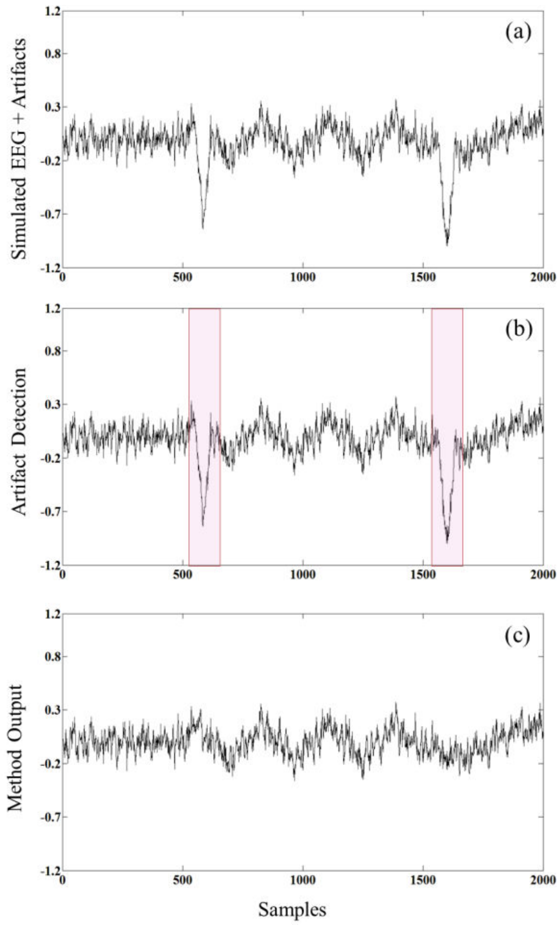

Figure 2. (a) The EEG segment with 2 out of 10 BAs that were detected, (b) The contaminated regions as detected by the NCC (c) The clear EEG segment as the output of the proposed method.

The average quantitative measures, for measuring the error of the CRs of the processed segment in respect with the clean CRs, that were used to validate the process on the contaminated regions were: the Absolute Mean of Error (AME), the Standard Deviation of Error (SDE), the Normalized Root Mean Squared Error (NRMSE), the Pearson Product-Moment Correlation Coefficient (R) and the Signal to Error Ratio (SER).

As it is observed by the results in "Table. I", a decrease in AME, SDE, NRMSE and an increase in R and SER indicate that BA rejection from the CR was satisfactory.

V. CONCLUSIONS

In this paper we have presented a method for detecting and removing EOG BAs. The main idea is based on the combination of template matching, EEMD and the choice of a value for the relative difference of the SDs of two successive IMFs.

The purposed method will be expanded for detecting other EEG events in simulated and real EEGs and its performance will be compared with other standard methods, like Joint Approximation Diagonalisation of Eigenmatrices (JADE). The expansion of the BA template library and the consideration of artefacts characteristics in order to detect variant BAs are also necessary.

TABLE I. PERFORMANCE MEASURES

Measures	**Error comparison of the contaminated regions on average**		
	Before BAs Rejection	***After BAs Rejection***	***Difference obtained***
AME	0.268	0.035	-0.233 (-86.94%)
SDE	0.294	0.055	-0.239 (-81.29%)
NRMSE	0.371	0.215	-0.156 (-42.05%)
R	0.247	0.767	+0.520
SER	-27.941dB	7.649dB	+35.590dB

REFERENCES

[1] D. Safieddine, A. Kachenoura, L. Albera, G. Birot, A. Karfoul, A. Pasnicu, A. Biraben, F. Wendling, L. Senhadji, I. Merlet, "Removal of muscle artifact from EEG data: comparison between stochastic (ICA and CCA) and deterministic (EMD and wavelet-based) approaches," EURASIP Journal on Advances in Signal Processing, vol. 127, 2012.

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

[3] Z. Wu and N.E. Huang, "Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method," *Advances in Adaptive Data Analysis.* vol.1, No.1, 2009, pp. 1-41.

[4] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung and H.H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society London A*, vol.454 no.1971, 8 March 1998, pp. 903-995.

[5] R. J. Oweis and E. W. Abdulhay, "Seizure classification in EEG signals utilizing Hilbert-Huang transform," *Biomed Eng Online* vol. 10, No. 38, 2011.

[6] M.T. Lo, P.H. Tsai, P.F. Lin, C. Lin and Y.L. Hsin, "The Nonlinear and Nonstationary Properties in EEG Signals: Probing the complex fluctuations by Hilbert-Huang Transform," *Advances in Adaptive Data Analysis*, vol. 1, No. 3, 2009, pp. 461-482.

[7] S. Reed, C. M. and S. J. Nasuto, "A novel approach to the detection of synchronization in EEG based on empirical mode decomposition," *J Comput Neurosci,* vol. 23, No. 1, 2007, pp. 79-111.

[8] D. Safieddine, A. Kachenoura, L. Albera, G. Birot, A. Karfoul, A. Pasnicu, A. Biraben, F. Wendling, L. Senhadji, I. Merlet: "Removal of muscle artifact from EEG data: comparison between stochastic (ICA and CCA) and deterministic (EMD and wavelet-based) approaches," *EURASIP Journal on Advances in Signal Processing,* vol. 127, 2012.

[9] B. Mijovic, M. Vos, I. Gligorijevic, J. Taelman, S. Huffel, "Source Separation from Single-Channel Recordings by Combining Empirical-Mode Decomposition and Independent Component Analysis," *IEEE Transactions on Biomedical Engineering,* vol. 57, No. 9, 2010, pp. 2188-2196.

[10] J.P. Lindsen and J. Bhattacharya, "Correction of blink artifacts using independent component analysis and empirical mode decomposition," *Psychophysiology,*vol. 47, No. 5, 2010, pp. 955-960.

[11] N.E. Huang, Z. Wu, S.R. Long, K.C. Arnold, X. Chen, K. Blank, "On instantaneous Frequency," *Advances in Adaptive Data Analysis*. vol.1, No.2, 2009, pp. 177-229.

[12] N. Yeung, R. Bogacz, C.B. Holroyd and J.D. Cohen, "Detection of synchronized oscillations in the electroencephalogram: An evaluation of methods," *Psychophysiology*, vol. 41, 2004, pp. 822-832.