A Kurtosis-Based Automatic System Using Naïve Bayesian Classifier to Identify ICA Components Contaminated by EOG or ECG Artifacts

M.A. Klados[1], C. Bratsas[1], C. Frantzidis[1], C.L. Papadelis[2], and P.D. Bamidis[1]

[1] Aristotle University of Thessaloniki, School Of Medicine, Laboratory of Medical Informatics, P.O. Box 323 54124, Thessaloniki, Greece
[2] Center for Brain/Mind Sciences (CIMEC), University of Trento, Mattarello, Trentino, Italy

Abstract— **Electrical signals detected along the scalp by an Electroencephalogram (EEG), but that originate from non-cerebral origin are called artifacts. Especially when these artifacts are produced by the human body we talk about biological artifacts. The most common biological artifacts are the electrical signals produced by ocular and heart activity. EEG data is almost always contaminated by such artifacts. The last decade Independent Component Analysis (ICA) has a crucial role in neuroscience and it takes great attention for artifact rejection purposes. According to ICA's methodology, EEG signals are decomposed to statistical Independent Components (IC) and then an EEG specialist is called to recognize the artifactual ICs. Some of the major limitations of the current approach are that the aforementioned selection is subjective, it demands a high skill EEG operator, it is time consuming and it cannot be applied in online processing. Our study employs machine learning techniques in order to recognize the contaminated ICs with ocular or heart artifacts. More specific 19-channel EEG datasets from 86 normal subjects were decomposed using ICA (19x86=1634 ICs in total). Then three independent observers marked an IC as artifactual if it includes ocular or heart artifacts, otherwise it was marked as normal. Then kurtosis was computed in short segments with 1250 sample points fixed length without overlap for each IC. The mean kurtosis value was computed for each IC and the Naïve Bayes Classifier (NBC) classifier was adopted in order to classify the ICs as artifactual or normal. The results suggest that the NBC has correctly classified 1611/1634 ICs (98.5924 %) so it can be suggested that kurtosis seems to be convenient for the classification of contaminated ICs by ocular or heart artifacts.**

Keywords— **ICA, Naïve Bayes Classifier, EOG, ECG, Artifacts.**

I. Introduction

Nowadays electroencephalography (EEG) is commonly used for understanding the cerebral functions as well as for evaluating the neuronal abnormalities, brain injuries and disorders. Electric potentials originated from brain tissues are low voltage signals, thus they are vulnerable not only to the different kinds of external noise but also to the physiological artifacts derived from internal sources of our body. The absence of artifacts is crucial for the accurate evaluation of EEG signals, so artifact rejection (AR) is a key step at the preprocessing level in both real time applications [1] and offline analyses. The most frequently seen biological artifacts are due to ocular, heart or muscular activity while the most troublesome of them are the Electrooculographic (EOG) and the Electrocardiographic (ECG) artifacts [2].

This piece of paper focuses in physiological artifacts derived by ocular and heart activity. Ocular artifacts are high voltage patterns in the cerebral signals caused by eye – blinking, or low – frequency patterns produced by eye movements [3]. It is known that cornea is positively charged with reference to retina. This retinocorneal potential difference generates a dipole within the eye – ball, thus ocular artifacts are in due to the reorientation of the aforementioned dipole [4,5]. Ocular activity introduce substantial artifacts into EEG signals (backward propagation), and they are most prominent at anterior sites [6]. On the other hand ECG artifacts are related to the field of the heart potentials over the surface of the scalp. Generally, people with short and wide necks have the largest ECG artifacts on their EEGs. ECG artifacts appear as sharp waves which are strongly correlated with the QRS complex and they can be easily recognized in EEG by their rhythmic (Fig 1).

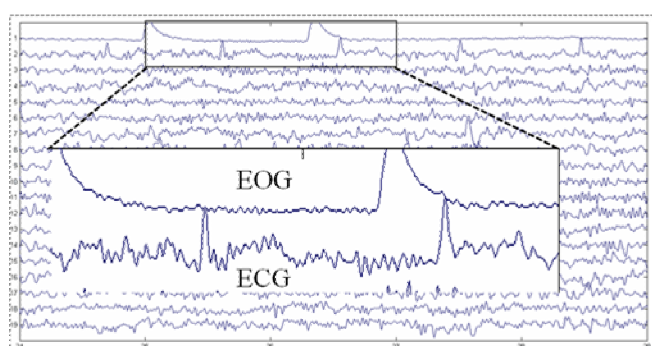

Fig. 1 **EOG and ECG Artifactual Components**. As it can be seen, in the above figure, the first IC is related with EOG activity where two eye-blinks are clearly observable. On the other hand the second IC is contaminated by ECG artifacts where the QRS complex is obvious

P.D. Bamidis and N. Pallikarakis (Eds.): MEDICON 2010, IFMBE Proceedings 29, pp. 49–52, 2010.
www.springerlink.com

It is widely accepted that the artifactual signals are independent from the ongoing cerebral activity, so artifactual signals should be extracted by the Independent Component Analysis (ICA) method. The last decade ICA has a crucial role in neuroscience research and it takes great attention for artifact rejection purposes [7, 8]. In order to make an automatic artifact rejection system there is an urgent need of some features capable to compute the probability of an IC to be artifactual or not. Many studies proposed the joint use of kurtosis with different entropies [9, 10]. As it can be noticed kurtosis is a common feature in both approaches. According to the literature [11] kurtosis is positive for "spiky" activity distributions, typical feature of artifactual components containing ECG and EOG activity; in contrast to this, kurtosis is negative for "flat" activity distributions [12]. So kurtosis should be the proper criterion for the recognition of the contaminated ICs by EOG and ECG activity. It is understandable that in a feature-based automatic systems it is preferable to use as less features as is possible without harming its performance. The current study proposes an automatic system for the recognition of the contaminated ICs by EOG and ECG artifacts.

So the remainder of this paper is structured as follows. In section II methodological background is provided alongside a detailed description of the proposed herein approach. Section III consists of tables and illustrations of the results; the latter are finally discussed in last two sections of the paper.

II. Materials and Methods

Independent Component Analysis

Independent Components Analysis (ICA) tries to recover independent source signals

$$s = \{s_1(t), s_2(t), \dots, s_n(t)\}$$

once they are linearly mixed by an unknown matrix $\mathbf{A}$ without a priori knowledge about the sources or the mixing process, only n recorded mixtures of above sources are known

$$x = \{x_1(t), x_2(t), \dots, x_n(t)\}.$$

The major problem is to find a square matrix $\mathbf{W}$ which will recover a version $u = \mathbf{W}x$ very close to the original sources. Bell and Sejnowski [13] proposed a simple neural network algorithm for BSS of n recorded signals x, to n independent sources s, using the information maximization principle (INFOMAX). They proved that maximizing the joint entropy $H(y)$, of the output of a neural processor, the mutual information among the output components is minimized. Ext-ICA extends the ability of the INFOMAX algorithm to perform BSS on n recorded signals x, having either sub – Gaussian or super – Gaussian distributions [14].

Naïve Bayesian Classifier

A Naïve Bayesian Classifier (NBC) is a simple classifier based on Bayes theorem with enhanced independence assumptions. In other words a NBC assumes that the occurrence of a particular event is uncorrelated with the occurrence of any other event. Abstractly, the probability model for the NBC is a conditional model:

$$p(C \mid F_1, \dots, F_n) = \frac{1}{Z} p(C) \prod_{i=1}^{n} p(F_i \mid C)$$

where Z is a scaling factor dependent on feature variables $\{F_i\}_{i=1,\dots,n}$ and C is the dependent class variable on $\{F_i\}_{i=1,\dots,n}$. The NBC uses the aforementioned model and combines it with a decision rule. The most common rule is to choose the most probable hypothesis, which is already known as the maximum a posteriori decision rule. The NBC is then described by the following function:

$$cl(f_1, \dots, f_n) = \arg\max_{c} p(C = c) \prod_{i=1}^{n} p(F_i = f_i \mid C = c)$$

Kurtosis

Kurtosis was used in order to detect "abnormal" peaked distributions as far as kurtosis is a measure of peakedeness. So we are expecting to have higher kurtosis values for artifactual ICs rather than for normal (Fig. 2).

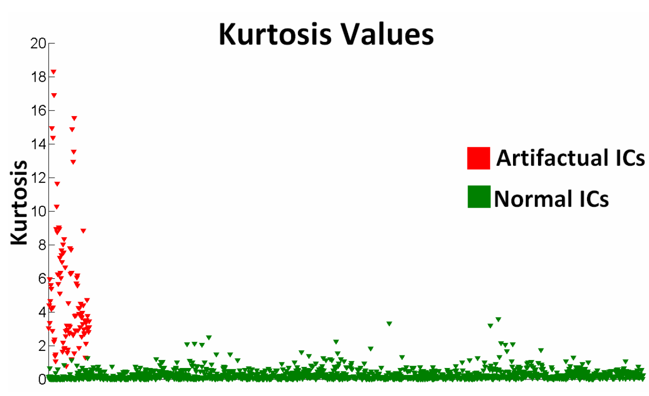

Fig. 2 **Kurtosis values for normal and artifactual ICs**. As it can be seen, in the above figure, the kurtosis values for artifactual ICS are much higher compared to kurtosis values for normal ICs

Kurtosis was computed by the next formula:

$$K = m_4 - 3m_2^2$$

where m_n is the n^{th} central moment:

$$m_n = E\left\{ (x-\tilde{x})^n \right\}.$$

Real EEG Data

Real EEG data have been obtained from twenty seven healthy subjects [14 males (mean age: 28.2±7.5) and 13 females (mean age: 27.1±5.2)] during a visual evoked potential (VEP) experiment. During this experiment, subjects were exposed to four different groups of emotional pictures (each group contains 40 trials), selected from International Affective Picture System (IAPS) [15] presented on a PC monitor; so 108 datasets have been obtained in total. From these 108 dataset, only 86 have been chose. The rest were excluded from this analysis, because they have many artifacts introduced by external sources like electrode movements or bad tangencies. Each dataset lasts almost two minutes and it was recorded by nineteen scalp electrodes placed according to the International 10-20 system. More specific sensors were placed at Fp1, Fp2, F3, F4, F7, F8, Fz, C3, C4, Cz, T3, T4, T5, T6, P3, P4, Pz, O1 and O2 sites.

Data Pre-processing

The earlobe montage has been used for our analysis [16]. According to this montage electrodes with odd indices were referenced to the left mastoid while electrodes with even indices were referenced to the right mastoid. Central electrodes (Fz,Cz, Pz) were referenced to the half of the sum of left and right mastoid. The signals were digitized at a rate of 500Hz and further filtered (band pass filter at 0,5-40Hz and notch filter at 50Hz). ICA was applied in filtered datasets and the result was nineteen ICs. Each IC was separated into 40 trials (according to the presented pictures) and kurtosis was computed for each trial. The mean kurtosis value was computed for each IC and the Naïve Bayes Classifier (NBC) classifier was adopted in order to classify the ICs as artifactual or normal. Then three independent observers manually marked the contaminated ICs (by EOG and ECG activity) as artifactual while the rest were marked as normal in order to quantify the NBC's classification efficiency.

Automatic Artifact Rejection System

According to the herein proposed model, EEG signals have to be decomposed into statistical ICs using ICA. Kurtosis value is then computed for each IC separately and a NBC automatically recognizes the artifactual components. Finally the system isolates the recognized as artifactual components and reconstructs the signals free of ocular and heart artifacts (Fig 3).

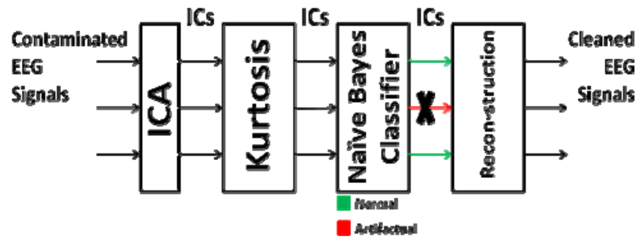

Fig. 3 **Automatic Artifact Rejection System**. According to this approach contaminated EEG signals are decomposed using ICA, then kurtosis was computed for each IC separately. NBC recognizes and the system removes the artifactual components (both EOG and ECG). The rest, normal ICs are re-projected back reconstructing with that way the cleaned EEG signals

III. Results

The classification using the NBC was performed by means of a 10-fold cross-validation. The accuracy rates, as well as, other statistical measurements regarding the classifier's performance, are reported in the next table (Table 1).

Table 1 Summary Statistics of the NBC classification performance used for artifactual component identification

Artifactual component identification task. Stratified cross-validation Summary	
Total Number of Instances	1634
Correctly Classified Instances	1611/1634 (98.5924 %)
Incorrectly Classified Instances	23/1634 (1.4076 %)
Kappa statistic	0.8912
Mean absolute error	0.0192
Root mean squared error	0.1142

More specifically, the first and the second line show the number and the percentage of the cases that were correctly and incorrectly classified respectively. The third line illustrates the kappa statistic which measures the agreement of predictions with the artifactual and normal instances. Finally the last two lines demonstrate the Mean Absolute Error and the Root Mean Squared Error which provide measurements of the difference between the predicted values and the real ones.

The detailed accuracy for each class is presented in Table 2. In specific, Table 2 illustrates the percentage of correctly classified items (TP rate), as well as, the percentage of the instances that were wrongly classified as items of the class under consideration (FP rate). Moreover, the precision feature is derived by dividing the number of elements were correctly classified with the total amount of instances that were classified in the class under consideration, whereas recall is the number of the correctly classified elements divided by the total number of the real elements of the class under consideration environment.

Table 2 Detailed Accuracy for each Class

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Artifactual IC	0.962	0.012	0.843	0.962	0.899	0.997
Normal IC	0.988	0.038	0.997	0.988	0.992	0.998

IV. CONCLUSSIONS AND DISCUSSION

It is logical beyond the obvious (Fig 2.) that normal ICs are more than artifactual. So the probability of an IC to be normal is much higher than to be artifactual. In this sense, the probability of an IC to be normal is 0.935 while a probability of an IC to be artifactual is 0.065. The mean±SD kurtosis value for normal components is 0.232±0.334 and 5.153±3.670 for artifactual.

It has also to be mentioned that the herein proposed system has correctly recognized 102/106 artifactual components, which means that only 4 artifactual ICs didn't recognized properly (0.03%). On the other hand our approach has successfully identified 1509/1518 normal components and only 19 normal components identified as artifactual and rejected (0.01%).

One limitation of the current approach is that kurtosis is not able to recognize if an artifactual component is in due to heart or ocular activity. Despite this, the results suggest that kurtosis is more than enough for the separation of artifactual components from normal ones.

ACKNOWLEDGMENT

This work is partially funded by the ICT Policy Support Programme (ICT PSP) as part of the Competitiveness and Innovation Framework Programme by the European Community (project LLM).

REFERENCES

[1] Blankertz B., Muller K.-R., Curio G., Vaughan T.M., Schalk G., Wolpaw J.R., Schlogl A., Neuper C., Pfurtscheller G., Hinterberger T., Schroder M., Birbaumer N., "The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials", IEEE Transactions on Biomedical Engineering; Volume 51, Issue 6, June 2004 Page(s):1044 – 1051.

[2] Fisch B.J., 1991. Artifacts. In: Fisch, B.J. (Ed.), Spehlmann's EEG Primer, 2nd edition. Elsevier, Amsterdam, The Netherlands, pp. 108-124.

[3] Anderer P, Roberts S, Schlogl A, Gruber G, Klosch G, Herrmann W, et al. Artifact processing in computerized analysis of sleep EEG – a review. Neuropsychobiology 1999;40:150–7.

[4] T-P. Jung, C. Humphries, T.W. Lee, M.J. McKeown, V. Iragui, S. Makeig, and T.J. Sejnowski. Removing electroencephalographic artifacts by blind source separation. Psychophysiology, 37:163-178, 2000.

[5] D.A. Overton and C. Shagass. Distribution of eye movement and eye blink potentials over the scalp. Electroencephalography and Clinical Neurophysiology, 27:546, 1969.

[6] McFarland DJ, McCane LM, David SV, Wolpaw JR. Spatial filter selection for EEG-based communication. Electroencephalogr Clin Neurophysiol 1997;103:386–94.

[7] Cichocki A, Vorobyov S.A. (2000), "Application of ICA for automatic noise and interference cancellation in multisensory biomedical signals", Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation, Helsinki, Finland, June 19–22, pp, 621–626.

[8] Jung T. P., Makeig S., Humphries C., Lee T. W., McKeown M. J., Iragui V. and Sejnowski T. J. "Removing electroencephalographic artifacts by blind source separation". Psychophysiology, 37(2):163-178, 2000.

[9] Delorme A., Makeig S., Sejnowski T. "Automatic artifact rejection for EEG data using high-order statistics and independent component analysis". Proceedings of the 3rd International Workshop on ICA, San Diego, December. 2001. p. 457–62.

[10] Antonino Greco, Nadia Mammone, Francesco Carlo Morabito, and Mario Versaci, Semi-Automatic Artifact Rejection Procedure based on Kurtosis, Renyi's Entropy and Independent Component Scalp Maps, World Academy of Science, Engineering and Technology 7 2005

[11] Giulia Barbati, Camillo Porcaro, Filippo Zappasodi et.al., Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals, Clinical Neurophysiology 115 (2004) 1220–1232

[12] Delorme A, Makeig S, Sejnowski T. Automatic artifact rejection for EEG data using high-order statistics and independent component analysis. Proceedings of the 3rd International Workshop on ICA, San Diego, December. 2001. p. 457–62.

[13] Bell A. J., Sejnowski T. J. 1995. An information-maximization approach to blind separation and blind deconvolution. Neural Computation, 7, 1129-1159.

[14] 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, 606-633.

[15] P.J. Lang, M.M. Bradley and B.N Cuthbert International Affective Picture System (IAPS): Technical Manual and Affective Ratings , NIMH Center for the Study of Emotion and Attention, 1997.

[16] Manousos A. Klados, Christos Frantzidis, Ana B. Vivas, et al., "A Framework Combining Delta Event-Related Oscillations (EROs) and Synchronisation Effects (ERD/ERS) to Study Emotional Processing," Computational Intelligence and Neuroscience, vol. 2009, Article ID 549419, 16 pages, 2009. doi:10.1155/2009/54941

Author: Manousos A. Klados
Institute: Laboratory of Medical Informatics, School Of Medicine, Aristotle University of Thessaloniki
Street: P.O. Box 323 54124
City: Thessaloniki
Country: Greece
Email:mklados@med.auth.gr