International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:2, No:9, 2008

Kurtosis, Renyi's Entropy and Independent Component Scalp Maps for the Automatic Artifact Rejection from EEG data

Antonino Greco, Nadia Mammone, Francesco Carlo Morabito, and Mario Versaci

***Abstract*—**The goal of this work is to improve the efficiency and the reliability of the automatic artifact rejection, in particular from the Electroencephalographic (EEG) recordings. Artifact rejection is a key topic in signal processing. The artifacts are unwelcome signals that may occur during the signal acquisition and that may alter the analysis of the signals themselves. A technique for the automatic artifact rejection, based on the Independent Component Analysis (ICA) for the artifact extraction and on some high order statistics such as kurtosis and Shannon's entropy, was proposed some years ago in literature. In this paper we enhance this technique introducing the Renyi's entropy. The performance of our method was tested exploiting the Independent Component scalp maps and it was compared to the performance of the method in literature and it showed to outperform it.

***Keywords*—**Artifact, EEG, Renyi's entropy, Independent Component Analysis, kurtosis.

I. Introduction

THE goal of this work is to improve the efficiency and the reliability of the automatic artifact rejection, in particular from Electroencephalographic signals (EEG) recordings. EEG is a technique for monitoring the electrical activity of the brain: the brain cells communicate by producing electrical impulses and, by means of some electrodes holded on the scalp over multiple areas of the brain, we can detect and record patterns of this electrical activity. Some flat metal discs (electrodes) are applied in different positions on the scalp (see Fig. 1), these discs are held in places with a sticky paste, the electrodes are connected by wires to an amplifier and a recording machine.

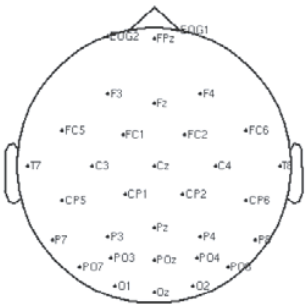

Fig. 1 The EEG electrode montage (32-channels)

All the authors are with Dipartimento di Informatica, Matematica, Elettronica e Trasporti (DIMET) of the University of Reggio Calabria, Via Graziella Loc. Feo di Vito Reggio Calabria, Italy (e-mails: {nadia.mammone, ngreco, versaci}@ing.unirc.it, morabito@unirc.it).

EEG monitoring helps to diagnose the presence and type of seizure disorders, to look for causes of confusion, and to evaluate head injuries, tumors, infections, degenerative diseases and metabolic disturbances that affect the brain.Artifacts are disturbances caused by eye movement, eye blink, electrode movement, muscle activity, movements of the head, sweating, breathing, earth beat, electrical line noise and so on. The occurrence of the artifacts may alter the analysis or completely obscure the EEG waves. When an artifact occurs during the acquisition of a signal, it generates some unwelcome signals that may overlap to the signals that we want to analyse, therefore it may undermine the results of the analysis. In this paper we focus on the automatic artifact rejection from EEG signals.

Many efforts have been devoted to artifact cancellation and in general two approaches have been adopted so far: the cancellation of the entire data segment affected by artifacts (in the following named *artifact-laden trials*), which implies throwing away both the useful and the artifactual information or, when we cannot afford to lose any information from the recorded data, the second approach consists in isolating the artifactual signals and then in cancelling them. In this paper we focus on the second approach.

It is well accepted that an artifact is independent from the rest of the signals, this is clear when the artifact is "external" (electrical line noise), it is also well accepted if it is "internal" (muscle activity, eye blink, etc.) because even though the triggering event starts from a brain area (for example the motor cortex) the time course of the artifact carries no information about the triggering event signal. The spectrum of the signals related to muscle activity contains high frequencies and it can spread from low frequencies to high frequencies; therefore, it can be overlapped to the spectrum of the brain waves, which goes from 0.5 up to 35 Hz, sometimes up to 50 Hz, this high frequency contribution can be weakened by a lowpass filtering.

Since it is accepted that the artifactual signals are independent from the rest, it is also well accepted that the artifactual signals can be extracted by Independent Component Analysis (ICA). ICA has been widely exploited as a tool for artifactual signals extraction, since it was proposed by Cichocki et al. [1] and Jung et al. [2]. The artifactual signal is isolated by ICA as a component independent from the rest.

In order to make the artifacts rejection procedure automatic we need some markers capable to measure how much an independent component (IC) is likely to be an artifactual signal. Delorme et al. [3] proposed the joint use of kurtosis and entropy as markers, while Vorobyov et al. [4] proposed the

Hurst exponent. Delorme et al. [3] proposed the joint use of kurtosis and entropy, in particular the Shannon's entropy, for the detection of the artifactual signals, once they have been isolated by means of ICA. This method had shown to be promising and, since we were interested in automatic artifact rejection, we decided to test and try to enhance it. The test showed some failures of the procedure in detecting some artifactual signals, thus we wondered whether another entropy definition could improve the performance of the method and we decided to introduce the Renyi's entropy. At first, the artifactual signals are isolated, by means of Extended-INFOMAX algorithm for ICA, then the joint use of kurtosis and Renyi's entropy, rather than Shannon's entropy, is proposed.

The paper is organized as follows: Section II describes the application of ICA to the EEG data and the Extended INFOMAX algorithm [5], Section III introduces the parameters for automatic detection and Section IV presents the results.

II. Independent Component Analysis

A. ICA and EEG Data

Independent component analysis was originally developed to deal with problems that are closely related to the cocktail-party problem. Since many real observations can be assumed to be a mixing of "sources", ICA was exploited for a lot of applications, for example for EEG data. The EEG data are presumably generated by mixing some underlying brain activity related signals and artifacts. This situation is quite similar to the cocktail-party problem: we would like to find the original components of the brain activity, but we can only observe mixtures of the components.

It is worth to wonder whether there are independent sources in the EEG data or not, we told above that it is a shared opinion that at least the artifacts can be considered independent from the rest. The second assumption for standard ICA is that each channel collects a linear combination of the sources, this assumption is well accepted for the EEG sensors, because the different electrical sources (brain activations, muscle activations) are supposed to sum linearly at the scalp electrodes.

B. The Extended-INFOMAX Algorithm

Given an input data vector $\mathbf{x}$, we consider it generated as a mixture of n statistically independent sources $(s_1,...,s_n)$:

$$\mathbf{x} = \mathbf{f}(\mathbf{s}) \tag{1}$$

If $\mathbf{f}$ is a linear function, we deal with the *linear* ICA problem:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{2}$$

ICA is a tool for multivariate data analysis which yields $(y_1,...,y_n)$ components as independent as possible so that $\mathbf{y} = \mathbf{W}\mathbf{x}$, and $\mathbf{W}$ is close to $\mathbf{A}^{-1}$.

Here we focus on the Bell-Sejnowski *INFOMAX* algorithm [5]. The input of the network is the observed data $\mathbf{x}$, the output is $\mathbf{y}$, whose elements are the estimated independent components $(y_1,...,y_n)$.

The *marginal entropy* of each estimated component is defined as follows:

$$H(y_i) = -E\{\log p(y_i)\} \tag{3}$$

The *joint entropy* has the following expression:

$$H(\mathbf{y}) = -E\{\log p(\mathbf{y})\} = -\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} p(\mathbf{y})\log p(\mathbf{y})d\mathbf{y} \tag{4}$$

The *mutual information* is defined as:

$$I(\mathbf{y}) = \int p(\mathbf{y})\log\frac{p(\mathbf{y})}{\prod_{i=1}^{N} p_i(y_i)}d\mathbf{y} \tag{5}$$

The relationship between marginal entropy, joint entropy and mutual information is:

$$H(\mathbf{y}) = H(y_1) + \dots + H(y_N) - I(y_1,\dots,y_N) \tag{6}$$

Therefore the components of $\mathbf{y}$ are as independent as possible when their mutual information is minimum. The marginal entropies are constant when $(y_1,...,y_n)$ have uniform distribution and are amplitude bounded random variables, so while minimizing mutual information we maximize joint entropy. Maximizing joint entropy leads to the INFOMAX learning rule [5]:

$$\Delta\mathbf{W} \propto \left[\mathbf{I} - \varphi(\mathbf{y})\mathbf{y}^T\right]\mathbf{W} \tag{7}$$

Where

$$\varphi(\mathbf{y}) = -\frac{\frac{\partial p(\mathbf{y})}{\partial \mathbf{y}}}{p(\mathbf{y})} \tag{8}$$

is called *score function* and p is the joint probability density function of $\mathbf{y}$.

INFOMAX has the limitation to assume that the sources have super-gaussian distributions and at most one source is normally distributed, therefore, Extended-INFOMAX was introduced to separate sources with a variety of distributions. Since Renyi's entropy, as we will describe in Section III, depends on sources distribution, we needed to know whether a source was sub-gaussian or super-gaussian, therefore we used the extended version of INFOMAX. A way of generalizing the learning rule is to consider an approximation of the estimated pdf of the sub-gaussian and the super-gaussian sources, this leads to extended learning rule is [5]:

$$\Delta\mathbf{W} \propto \left[\mathbf{I} - \mathbf{K}\tanh(\mathbf{y})\mathbf{y}^T - \mathbf{y}\mathbf{y}^T\right]\mathbf{W} \tag{9}$$

There are two different learning rules for sub-gaussian and super-gaussian sources, the switching criterion between the two rules is given by **K**: it is a N-dimensional diagonal matrix whose elements are:

$$k_{ii} = \begin{cases} 1 & \text{when } y_i \text{ is super-gaussian} \\ -1 & \text{when } y_i \text{ is sub-gaussian} \end{cases}$$

these elements are estimated as part of the algorithm. Extended-INFOMAX yields the independent components and the unmixing matrix **W**, to reconstruct the observed signals we compute:

$$\tilde{\mathbf{x}} = \mathbf{W}^{-1}\mathbf{y} \tag{10}$$

in order to reconstruct the cleaned signals, we have to suppress the alleged artifactual components and to compute the (10).

III. Automatic Artifact Detection

Artifacts are outlier data, in other words transient and unexpected events; therefore we need some markers to measure this oddness to detect outlier trials or outlier independent components (ICs) once we have extracted them by ICA. The markers proposed by Delorme et al. [3] are kurtosis and entropy, which are both related to the distribution of the signals.

A. Kurtosis as a Marker for Artifacts

In some trials, the distribution of the components is very peaky, for instance during a transient strong muscle activity, so the kurtosis can help us to detect these artifacts. Given a scalar random variable *x*, kurtosis has the following expression:

$$k = m_4 - 3m_2^2 \tag{11}$$

$$m_n = E\left\{(x - m_1)^n\right\} \tag{12}$$

where m_n is the *n*-order central moment of the variable and m_1 is the mean. Kurtosis is positive for "peaked" activity distributions, typical of eye blink and cardiac artifacts; kurtosis is negative for "flat" activity distributions, typical of noise [3].

In order to detect artifactual components, kurtosis was computed for each trial and for all the ICs and it was normalized, to 0-mean and standard deviation 1, with respect to all the ICs, the threshold was set at ± 1.64 [6], if a component exceeded the threshold in more than 20% of the trials, the component was marked for rejection because it was very likely to be an artifact.

Kurtosis can also be exploited to detect the components which were very likely to account for gaussian noise: we estimated the kurtosis of the pdf of each component (global kurtosis), the IC with the minimum positive coefficient was marked as gaussian noise.

B. Entropy as a Marker for Artifacts

As a second marker we used the differential entropy:

$$H(x) = -\int p_x(\xi)\log(p_x(\xi))d\xi \tag{13}$$

The entropy can be interpreted as a measure of randomness: if the random variable is concentrated on small temporal intervals, its differential entropy is small, indeed the variables whose probability densities take large values give a strong contribution to the integral in the (13), so their entropy is small. This feature of the entropy helps us to identify the signals which are concentrated in small temporal intervals with high probabilities and, therefore, which are very likely to be artifacts.

An approximation of the (13) was used:

$$H^i(j) = -\sum_{x \in j} f_j^i(x)\log\left(f_j^i(x)\right) \tag{14}$$

which is the entropy of the *ICi* in the trial *j*, and $f_j^i(x)$ is the pdf of the component *ICi* during trial *j*.

In order to detect the artifactual components, entropy was computed for each trial and for all the ICs and it was normalized, to 0-mean and standard deviation 1, with respect to all the ICs, the threshold was set at ± 1.64 [6], if a component exceeded the threshold in more than 20% of the trials, the component was marked for rejection because it was very likely to be an artifact.

We want to compare the performances of the Shannon's entropy (14) and the Renyi's entropy. Renyi's entropy depends on a parameter α and, for a random variable *y* with a pdf $f_y(y)$, it is defined as:

$$H_{R_\alpha}(y) = \frac{1}{1-\alpha}\log\int_{-\infty}^{+\infty} f_Y(y)dy = \frac{1}{1-\alpha}\log E\left\{f^{\alpha-1}(y)\right\} \tag{15}$$

The Parzen window pdf estimate of a random variable *y* for which only the samples $\{y_1, \ldots, y_N\}$ are given, is defined by:

$$\hat{f}(y) = \frac{1}{N}\sum_{i=1}^{N} k_\sigma(y - y_i) \tag{16}$$

where *i* = 1...*N* and k_σ is the gaussian kernel, whose size is specified by σ (we chose $\sigma = 0.25$).

When we substitute the real pdf with the approximation (16) in the (15) we have:

$$H_{R_\alpha}(y) = \frac{1}{1-\alpha}\log\left[\frac{1}{N^\alpha}\sum_j\left(\sum_i k_\sigma(y - y_i)\right)^{\alpha-1}\right] \tag{17}$$

The pdf's were estimated by the Parzen windowing method in both the Shannon's entropy and the Renyi's entropy computation.

All the properties stated in Section III.B are for both Shannon's entropy and Renyi's entropy and therefore they help us to identify the signals which are concentrated in small temporal intervals with high probabilities, for both artifact-laden trials and artifactual components. The automatic detection procedure, described in the Section III.B, was used for Renyi's entropy and each trial coincides with a Parzen window. We tested the automatic detection of the artifactual components for $\alpha = 2$ in order to study the behaviour of the Renyi's entropy because we know that, for super-Gaussian sources, entropy orders ≥ 2 should be preferred, whereas for sub-Gaussian sources, entropy orders smaller than 2, perhaps closer to 1 or even smaller than 1, should be preferred [7].

Entropy orders larger than 2 emphasize samples in concentrated regions of data, whereas smaller orders emphasize the samples in sparse regions of data. If the mixtures belong to different kurtosis classes, the quadratic entropy can be employed as it puts equal emphasis on all data points regardless of their probability density.

IV. Results

We used a benchmark [8] to test the procedure, it consists on 32-channels EEG data (Fig. 2) sampled at 128 Hz, we extracted the first 24 seconds data segment because it is known that INFOMAX algorithm provides N reliable components (for N channel data) if we process at least $3N^2$ data samples, in this case $N = 32$, therefore we have to process at least 3072 samples (24 seconds) [8]. We removed the electrical line noise by a notch filter at 60Hz and we set the width of each Parzen window at 200, therefore we had 15 data segments (trials). Thus, we had to reject an IC if the markers computed for that IC exceeded the fixed threshold in more than 20% of the trials, therefore in at least 4 trials.

The ICs and the scalp maps of the projection of the ICs on the electrode sites were obtained by EEGLAB [8].

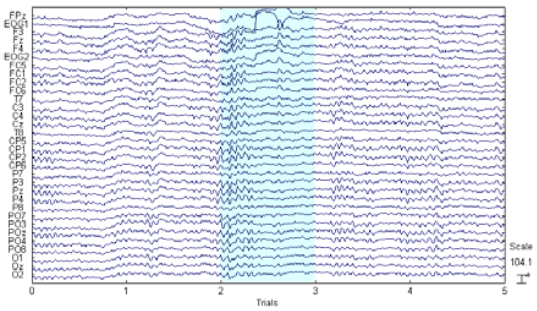

Fig. 2 The EEG data. During trial 3 (highlighted) the early channels are affected by ocular artifacts

We processed the data by Extended-INFOMAX and the ICs are plotted in Fig. 3. The scalp map projection of the ICs was estimated thanks to EEGLAB and is depicted in Fig. 4, this map helps us in localizing the activation of the ICs and in detecting those ICs whose activation is concentrated on one electrode, as the activation of the ocular artifacts is. Looking at the distribution of the activation of each IC, we can realize that IC10 has the typical eye blink artifact projection, IC9 has the typical eye movement projection. IC14 and IC20 seem to account for eye movement too. The ocular artifact can be identified thanks to scalp maps because the intensity of their activation is high in the frontal electrodes, in particular, the activation of the eye blink artifact is concentrated in a spot whereas the activation of the eye movement is a bit more spread.

Once the ICs were extracted, kurtosis, Shannon's entropy and Renyi's entropy were computed for each trial for each IC, as described in Section III, they are plotted in Fig. 5. Table I summarizes the automatic detection.

Kurtosis detected IC10, Shannon's entropy detected IC2, IC10, IC17, IC21, IC24, Renyi's entropy detected IC1, IC2, IC9, IC10, IC17, IC21, IC24. Global kurtosis detected IC14 which accounts for an amount of the eye movement artifact.

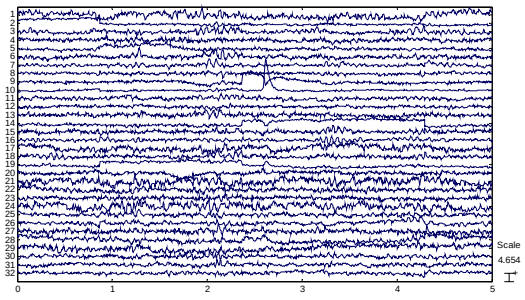

Fig. 3 The Independent Components. By a visual inspection we can realize that IC9 has the typical eye movement artifact shape and that IC10 surely accounts for eye blink

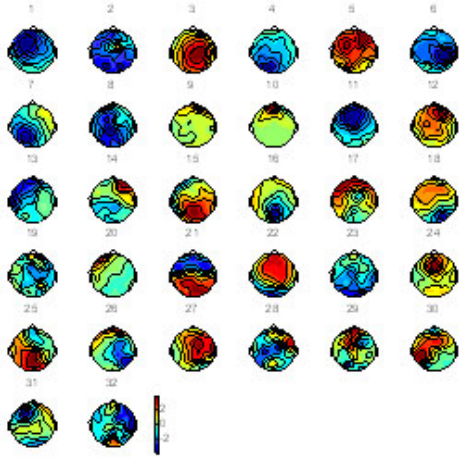

Fig. 4 The scalp map of the projection of the ICs. IC9 accounts for eye movement artifact, IC10 accounts for eye blink artifact, IC14 and IC20 are likely to account for a bit of the eye movement too

Thus, the joint use of kurtosis and Shannon's entropy detected IC2, IC10, IC14, IC17, IC21, IC24, whereas the joint use of kurtosis and Renyi's entropy detected IC1, IC2, IC9, IC10, IC14, IC17, IC21, IC24. IC9 was detected only by our method, and it is certainly an eye movement artifact. IC1 was detected only by our method too, and it looks like a muscular artifact. IC19 which accounts for a little amount of the eye movement too, but it was not detected by any of the two methods.

In Fig. 6, the cleaned dataset is depicted, it was reconstructed cancelling the ICs detected by the joint use of kurtosis and Renyi's entropy, the artifacts appear suppressed.

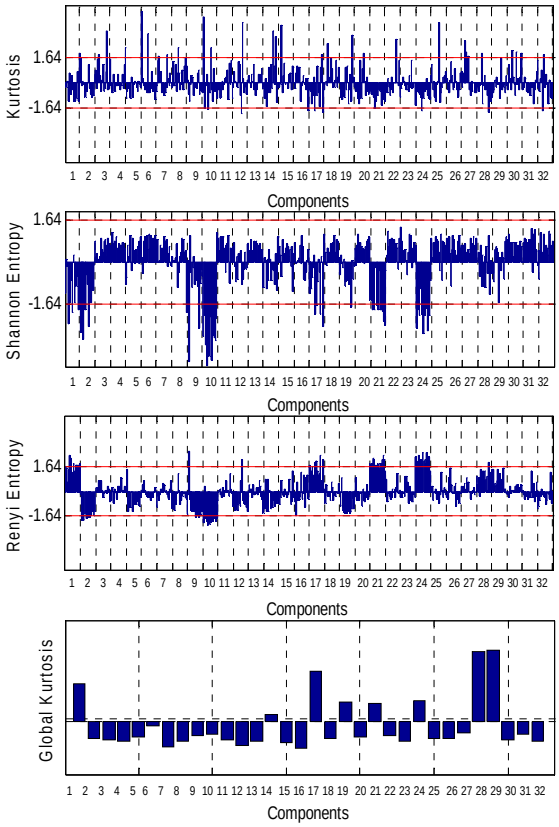

Fig. 5 The automatic artifact detection. Kurtosis detected IC10, which accounts for eye blink artifact. Shannon's entropy detected IC2, IC10, IC17, IC21, IC24. Renyi's entropy detected IC1, IC2, IC9, IC10, IC17, IC21, IC24. Global kurtosis detected IC14. IC9 was detected only by the proposed method, and it is certainly an eye movement artifact. IC1 was detected only by our method too, and it looks like a muscular artifact

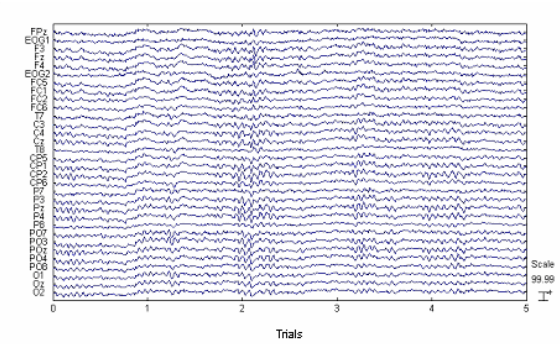

Fig. 6 The reconstructed EEG data. The ICs marked according to kurtosis and Renyi were suppressed and the data reconstructed by the Eq (10). Trial 3 appears free from the artifacts

V. Conclusions

Artifact rejection is a key topic in biomedical signal processing and in literature, Independent Component Analysis (ICA) has shown to be a suitable tool for extracting the artifactual signals and some higher order statistics, kurtosis together with Shannon's entropy, have been used as markers for the automatic artifact detection. In this paper we have proposed the joint use of kurtosis and Renyi's entropy, rather than Shannon's entropy, and we have tested the reliability of the procedure comparing its results with the results of the visual inspection of the shape of the independent components and of the component scalp maps. The joint use of kurtosis and Renyi's entropy showed to outperform the joint use of kurtosis and Shannon's entropy. Future efforts will be devoted to enhance the reliability of the procedure focusing on artifactual signal extraction.

References

[1] A. Cichocki, S. A. Vorobyov (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.

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

[3] A. Delorme, S. Makeig, T. Sejnowski, "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.

[4] S. Vorobyov, A. Cichocki, "Blind noise reduction for multisensory signals using ICA and subspace filtering, with application to EEG analysis", *Biol. Cybern.* 86, 293–303 (2002).

[5] T.-W. Lee. "Independent Component Analysis". *Kluwer Academic Publishers,* 1998.

[6] G. Barbati, C. Porcaro, F. Zappasodi, P. M. Rossini, F. Tecchio, "Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals", *Clinical Neurophysiology* 115 (2004) 1220–1232.

[7] D. Erdogmus, K. E. Hild II, J. C. Principe, "Blind source separation using Renyi's marginal entropies." *Neurocomputing* 49 (2002) 25-38.

[8] A. Delorme, S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis." *Journal of Neuroscience Methods.* http://sccn.ucsd.edu/eeglab/download/eeglab_jnm03.pd