ISSN:0975-9646

Hemant P. Kasturiwale et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012,3544-3547

Component Extraction of Complex Biomedical Signals and Performance analysis

Hemant P. Kasturiwale,
Electronics Engineering Department ,
Thakur College of Engineering & Technology,Kandivali(E) , Mumbai

P.V.Ingole
Electronics & Telecommunication Department
G. H. Raisoni College of Engineering & Management, Amravati

Abstract- Biomedical signals can arise from one or many sources including heart, brains and endocrine systems. Multiple sources poses challenge to researchers which may have contaminated with artifacts and noise. The Biomedical time series signal like electroencephalogram (EEG), electrocardiogram (ECG), etc. The morphology of the cardiac signal is very important in most of diagnostics based on the ECG. The immense scope in the field of biomedical-signal processing Independent Component Analysis (ICA) is gaining momentum due to huge data base requirement for quality testing. The diagnosis of patient is based on visual observation of recorded ECG, EEG, etc, may not be accurate. To achieve better understanding, PCA (Principal Component Analysis) and ICA algorithms helps in analysing ECG signals .This paper describes some algorithms of ICA in brief, such as Fast-ICA, Kernel-ICA, MS –ICA, JADE, EGLD-ICA, Robust ICA etc. The quality & performance of some of the ICA algorithms are tested and analysis of each can be done with respect to Noise/Artifacts, SIR (Signal Interference Ratio), PI(performance Index). The most common bioelectric signals are EEG and ECG. The experimental results presented in the paper show that the proposed here to indentify the various components with higher accuracy in the particular algorithm based on classifying biomedical data.

Keywords- **CBS(complex biomedical signals), EEG(electroenphalogram*),* ECG(Electrocardiograph), PCA(Principal Component Analysis) ,ICA(Independent Component Analysis), Algorithms, SIR , Signal processing.**

I. INTRODUCTION

The Processing biomedical data becomes more and more important now days, because of its relevance and support for the decisions of the specialists. Due to its complex nature, the process of acquisition for medical information about a certain patient usually supplies data that contains a significant amount of noise. The detection of electrical properties of the heart can be valuable clinical tool. The ECG is major tool used in clinical situations. In recorded ECGs many types of noise and artifact are present. Early work on noise and artifact reduction in the ECG used either temporal or spatial averaging techniques. Apart, many methods to filter out the noise and artifacts from ECG are only partially successful. On the other hand the filters often lead to a reduction in the amplitudes of the components waves, Q-R-S wave. (shown in fig 1.) Hence filters to remove interference when it is same frequency range as the cardiac signal. In the particular problem of classifying certain types of biomedical signals, many techniques that extract relevant information from the data have been used. In order to detect abnormalities in informational data, Principal Component Analysis (PCA) is used. The idea behind extracting the principal components is to find the spatial directions of the data set, that have the maximum data variance ([1],[2]). PCA is a very efficient technique used widely to obtain relevant statistical description of the data. It is extensively applied in preprocessing and classification steps of the information in several domains. Its popularity comes from the fact that it uses first and second order statistics in order to characterize the data sets, thus giving a high level of confidence in the obtained results [3].Although PCA is very useful in the case of extracting relevant information from some data sets, there are other techniques that use higher-order statistics in characterizing the data. Independent Component Analysis (ICA) may be regarded as a particular case of blind source separation problem ([4], [5], [6]) that attempts to separate all underlying sources contributing to the data without knowing these sources or the way that they are mixed. In order to achieve the goal of separating independent components from mixed signals, the ICA model does not need any prior knowledge about each source. The most important assumption that is associated with the ICA model is the independence of the sources to be estimated.

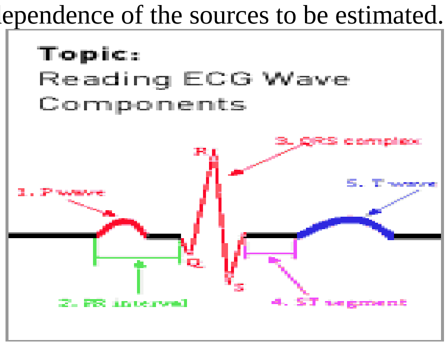

Fig.1 Standard ECG with components

Another assumption used in the ICA model is the following: the signal from each sensor (that is, an observable variable) has different mixing ratio of the independent components. ICA has been introduced to mechanical dynamic signal analysis in the last few years ([7], [8]). ICA is more suitable when the purpose is to find a component from a mixture of many independent sources. However, in some circumstances, there is limitation to install too many sensors to satisfy the requisition for ICA. In our application, the signals received from the sensors are formally represented by the examples in the data set that is taken into consideration. The sources are, in fact, the independent components estimated by the ICA model or the latent variables depending on which the recorded signals can be expressed. In biomedical data, noise is almost always present, because of the residuals in the signals, coming from other body activities [9]. The main disadvantage of the various types of approaches in the problem of noise reduction is that they are time consuming and computationally expensive. That is why the idea of preprocessing the data is much more popular nowadays. The noise reduction may be done by applying suitable low pass filters in time domain before implementing ICA algorithm [10]. ECG appears to satisfy some of the conditions for ICA: 1) mixing happens linearly at electrodes 2) Time delays in signal transmission are negligible: 3) There appears to be fewer sources than mixtures 4) sources have non-Gaussian voltage distributions. Some violations can happen just because of movements of the heart such as contraction of the chambers.

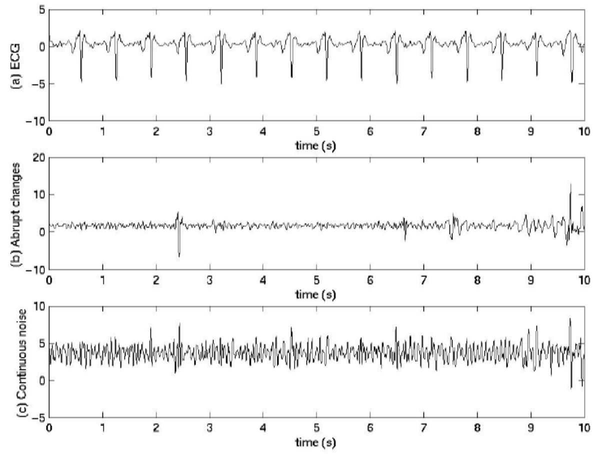

Fig. 2. Typical waveforms of (a) the ECG, (b)abrupt changes and (c) continuous noise

The ECG signal is a very weak time varying signal (about 10 microvolt) and has a frequency between 0.5Hz to 100Hz. The waveforms thus recorded have been standardized in terms of amplitude and phase relationships and any deviation from this would reflect the presence of an abnormality. Abnormal patterns of ECG may be due to undesirable artifacts Normally ECG is contaminated by power line interference of 60Hz .So it is desired to eliminate this noise and to find how best the signal can be improved.

II.THEORETICAL FRRAMEWORK: DECOMPOSITION

A. *PCA (Principal Component Analysis)*

Principal component analysis (PCA) has widespread applications because it reveals simple underlying structures in complex data sets using analytical solutions from linear algebra which provides a brief summary for implementing PCA. A primary benefit of PCA arises from quantifying the importance of each dimension for describing the variability of a data set. In particular, the measurement of the variance along each PCA is completely nonparametric: any data set can be plugged in and an answer comes out, requiring no parameters to tweak and no regard for how the data was recorded. From one perspective, the fact that PCA is non-parametric (or plug-and-play) can be considered a positive feature because the answer is unique and independent of the user. From another perspective the fact that PCA is agnostic to the source of the data is also a weakness. PCA to certain extent due to PCA worked mainly on dimension reduction. So, discrimination of signal (ECG) and noise dimension is not straightforward.

B. *Independent Component Analysis (ICA)*

ICA has become an important signal processing and data analysis technique; it is a particular case of blind source separation and it is used on a wide range of data, such as biomedical, acoustical and astrophysical signals. Generally speaking, ICA is viewed (Jutten [14] , Cardoso [15], Jutten and Herault [16], Comon [17] , Hyvarinen et al. [6]) as a statistical signal processing technique that models a set of observations, x, with an instantaneous linear mixing of independent latent variables, s

$$x(t) = As(t) + ns(t) \quad (1)$$

where ns is additive noise.

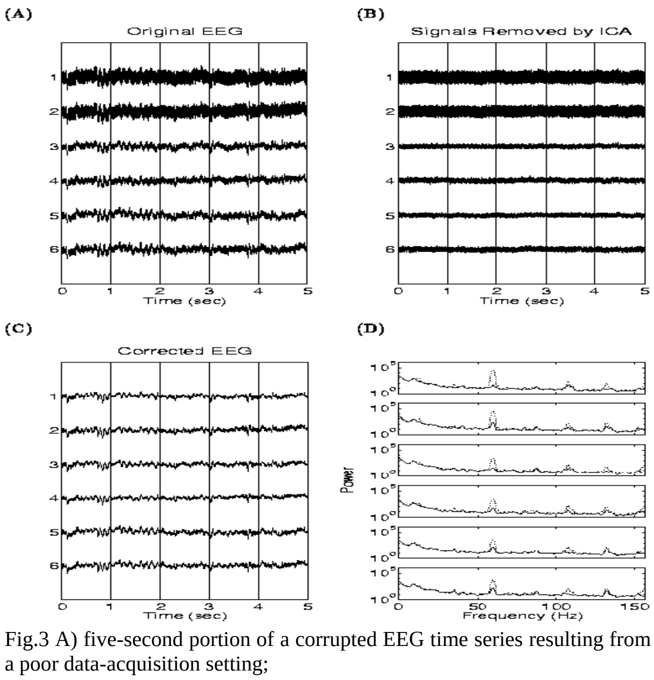

Fig.3 A) five-second portion of a corrupted EEG time series resulting from a poor data-acquisition setting;
B) Noise components extracted by ICA (*right panel*).
C) The same EEG signals corrected for artifacts by ICA by removing the six selected components, and,
D) Spectral analysis of the original and artifact-corrected EEG recordings. Note that EEG activity is more visible than in (A), particularly in channels 1 and 2, and the line noise (60 Hz) and aliased line noise frequencies (near 12 Hz, 105 Hz, 135 Hz) are reduced

ICA supplies a series of techniques allowing the decomposition of a random vector in linear components which Are as independent as possible, where the independence should be understood in its strong statistical sense. The problem of recovering sources from their linear mixtures without knowledge of the mixing channel can be expressed in its simplest form as the problem of identifying the factorization of the N-dimensional observations x into a mixing channel A and M-dimensional sources s, a large body of work being devoted to the case when the statistical independency of sources is assumed. The goal of ICA is to recover the latent components from the observations. If noise is negligible, this can be achieved by the determination of an inverse linear mapping from x to s, say

$$x = As \qquad (2)$$

The equation represents a simplified ICA model, resembling the one presented in equation 1, in which the noise is considered to be negligible and the time component is implicit. Provided the model is used to estimate the sources, or the latent variables, denoted by s, the simplified ICA model may be written in the following form [4]:

$$s = B\,x \qquad (3)$$

In order to determine the matrix B, usually, an intuitively justified criterion function is selected, yielding to an unconstrained optimization problem. Most algorithms, directly or indirectly, minimize the mutual information, I, between the component estimates. It can be shown (Hyvarinen [5], [18], Hyvarinen et al. [6]) that minimizing I corresponds to the maximization of the negentropy, a measure of non-Gaussianity of the components. Most of the existing ICA algorithms can be viewed as approximating negentropy through simple measures, such as high-order cumulants (Cardoso [15], Hyvarinen [19], Hyvarinen et al. [6]).In our work; we used the FastICA algorithm introduced by [4]. In [4] the negentropy is approximated by

$$F(y) = [EG(y) - EG(y)]^2 \qquad (4)$$

Where G is an nonquadratic function, and $y = w^T z$ are Gaussian variables of zero mean and unit variance, yielding to the constraint optimization problem [4]

C. Data Processing Steps

1) Centering- It is nothing but observed data to be used with ICA should have zero mean.

2) Whitening- Whitening means to transform the observed vector X such that its components become uncorrelated and have unit variances. This is also called sphering. The preprocessing makes the data of ICA estimation simpler and the main measures of non-gaussianity include- Kurtosis, Negentropy.

III. ALGORITHM AND PARAMETERS

Algorithms like FASTICA, Complex ML ICA, JADE, and Newton ML, kernel algorithms can able to find components within the mixture signal, i.e. noise or artifacts it has to be decided on the facts that mixture need to multichannel or single channel. But in spite of all there are some parameters which will calculate the amount of noise in the mixture are SNR, PI, and Psi etc.

1) Signals to Noise Ratio (SNR)

SNR is an engineering term for the power ratio between a signal (meaningful information) and the background noise. ECG signals normally have a wide dynamic range. SNRs are usually expressed in terms of the logarithmic decibel scale. In decibels, the SNR is 20 times the base-10 logarithm of the amplitude ratio, or 10 times the logarithm of the power ratio:

SNR=10 log (Es/En),

where Es is a average signal amplitude and En is average noise amplitude measured within the system bandwidth. With FASTICA SNR is better over with normal filtering.

2) Performance Index (PI)

$$PI = \sum_{i=1}^{n}\{(\sum_{k=1}^{n} \frac{|g_{ik}|^2}{\max_j g_{ij}} - 1) + ((\sum_{k=1}^{n} \frac{|g_{ki}|^2}{\max_j g_{ji}} - 1)\}$$

Where gij is an ij-th element of the global matrix G=W*A0, W calculated demixing matrix. Script for computing performance index is available

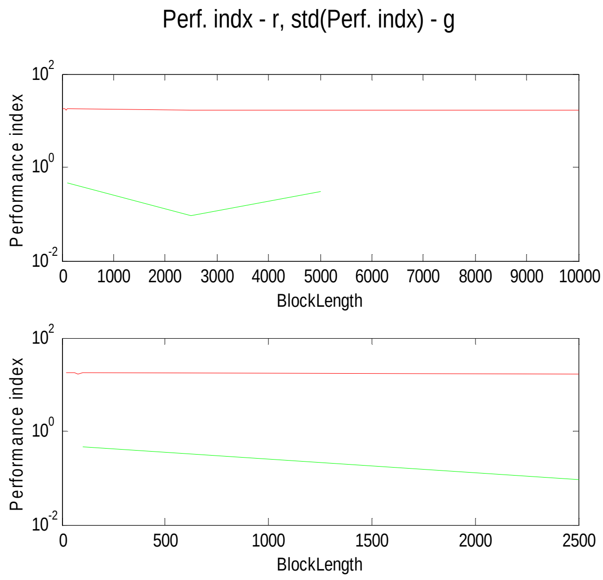

Fig.4 Performance of ICA algorithm for given mixture.

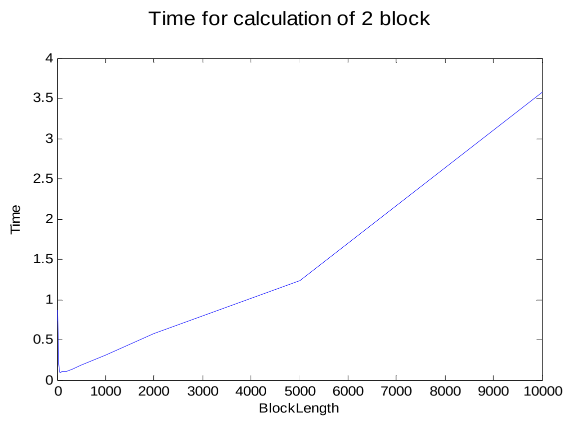

Fig.5 Time for calculation for given signal mixture using ICA

IV. CONCLUSIONS

1) Different Algorithm has different performance. The FAST-ICA & MS- ICA has acceptable results. When sources have same (lambda) distribution the algorithm of EGLD-ICA has good results.
2) FAST- ICA overall performance is good for less number of blocks (preferably less than 2000 samples).
3) FAST-ICA is much more time consuming for longer blocks but less than MS-ICA
4) SNR is much better than normal way of filtering.
5) KICA worked better for more datasets than any other ICA algorithm.
6) JADE does not depend on gradient optimization techniques, neither on choice of unmixing matrix. This makes it more attractive over others.

ACKNOWLEDGMENT

The author wishes to thanks Dr.Shardul Dongaonkar (M.D.),India for giving in-depth analysis of biomedical signal specially ECG signal.

REFERENCES

1 R. M. Rangayyan. Biomedical signal analysis: A case-study approach. IEEE Press Series in Biomedical Engineering, 2002.
2 A. Hyvarinen, J. Karhunen, and E. Oja. Independent component analysis. Wiley, 2001
3 Wisbeck JO, Barros AK, Ojeda R. Application of ICA in the Separation of Breathing artefacts in ECG Signals. International Conference on Neural Information Processing, (ICONIP'98), Kyushu, Japan, 1998
4 T.-P. Jung, S. Makeig, T.-W. Lee, M. McKeown, G. Brown, A. Bell, and T. J. Sejnowski. Independent component analysis of biomedical signals. In Proc. Int. Workshop on Independent Component Analysis and Blind Separation of Signals (ICA'00), pages 633 {644, Helsinki, Finland, 2000.
5 FastICA.-The FastICA MATLAB package. 1998.
6 Molegedy, H. Schuster, "Separation of Independent Signals using Time delayed Correlations", Physical Reviews letters .Vol72pp.
7 R Batch, Michael Jorden," Kernel Independent Component Analysis," Journal of Machine Learning Research, Vol.3 pp-1-48, 2002.
8 JADE.- The JADE MATLAB package. 1999.
9 Hyvarinen, A. and Oja, E. (1997) A fast fixed-point algorithm for independent component analysis. Neural Computation, 9:1483-92.
10 The MIT-BIH Noise Stress Test Database.
11 B. D. Moor, Database for the Identification of Systems (DaISy).