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

A new approach to improve the quality of Biosensor signals using Fast Independent Component Analysis: Feasibility study using EMG recordings

Ganesh R Naik, *Member, IEEE* Yina Guo, *Member, IEEE* and Hung Nguyen, *Senior Member IEEE*

Abstract — **The proposed signal processing technique uses Fast Independent Component Analysis (ICA) algorithm to improve the quality of the original biosensors recordings, which can be used as valuable pre-processing technique such as cross talk removal, artefact reduction etc. Initially, the ill conditioned original surface Electromyography (sEMG) recordings were separated using ICA methods and later they were reconstructed using modified un-mixing matrix. The simulation results showed huge improvement of the original recorded signal after reconstruction. The proposed method has potential applications in various biomedical signal processing techniques.**

I. INTRODUCTION

Cross talk removal and noisy source separation are one of the challenging tasks in biomedical signal processing techniques. During physiological signal recordings, noise and electrical activity from parts of the body other than the target muscles gets super-imposed. Moreover, the complexity of the human muscle anatomy and noise factors results difficulty in identifying the number of active sources from the multiple channel recordings.

Researchers have employed numerous techniques to reduce the cross talk and artefacts from different parts of the body. Most common methods among them include types of electrodes, careful selection of the electrode location and with suitably designed data acquisition equipment. However, there are number of situations where the cross talk and artefacts may be very strong and prevent the useful interpretation of the desired signals. Examples of such situations include surface Electromyography (sEMG) recordings, from the back and thorax area that have Electrocardiogram (ECG) and breathing artefacts. Another instance is, during sEMG recordings of hand and wrist muscles to identify different hand gestures for human computer interface. The cross talk due to the different muscles can result in unreliable recordings, also, at some instances the cross talk maybe of greater magnitude than the signal itself. Hence, it is essential to keep the original sensor recordings of very good quality, which would further help in other post processing methods [1].

There exist several signal processing methods to solve the above problem; the simplest and most commonly used technique to improve the quality of the recording is rejection [2], it is performed by discarding a section of the recording that has artefact exceeding a threshold. This method is simple, but causes a considerable loss of data and its reliability is debatable since it is primarily based on visual examination. Moreover, it is mainly reliant on the technician who is investigating the signal, making it less dependable. The other commonly used techniques to improve the quality of biosignals recordings include spectral filtering, gating and cross-correlation subtraction [3]. Spectral filtering is often not practical due to the overlap of the frequency spectrum of the desired signals and the artefact component. Conversely, gating and subtraction may introduce discontinuity in the reconstructed signal [1]. In the recent past, researchers have used techniques such as time domain [4], frequency domain and regression [5] methods. However, simple regression in time domain can over-compensate the artefacts [1], the regression techniques depend on the availability of a proper regressing channel (a separate channel) to record the corresponding artefact as a reference. This is often not feasible when recording sEMG data. As a result, better artefact removal and signal enhancement techniques are necessary to overcome the disadvantages of the previous methods. Also, in most of the cases, the quality or the signal separation solely depends on quality of the recordings, irrespective of the type of the algorithms used to solve the problem [6-8]. In general, sEMG signals are linear in nature. Hence, the signal originating from one muscle can normally be considered to be independent of other bioelectric signals such as ECG, Electro-oculargram (EOG) and signals from adjacent muscles. This justifies the usage of Independent Component Analysis (ICA) for this application [9].

II. THEORY

ICA is a data analysis technique whose rationale is to solve the so-called Blind source Separation (BSS) problem, which consists of recovering unobserved signals or "*sources*" from observed mixtures, i.e., from the output of an array of sensors. ICA is an iterative technique that can be used to separate signals from different sources. It is a very convenient technique for source separation as it requires very little information of the sources or the signals to be separated, and with the availability of easy to use software packages, is becoming very popular for numerous applications [6, 9].

Let $x_1, x_2, \ldots, x_n$ be a set of n observed random variables expressed as linear combinations of another n random variables $s_1, s_2, \ldots, s_n$, i.e.,

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \ldots + a_{in}s_n = \sum_{j=1}^{n} a_{ij}s_j \text{, where i = 1, . . . , n}$$

and $a_{ij} \in$R. The variables s_i are assumed to be statistically

Ganesh R Naik and Hung Nguyen are with Faculty of Engineering and Information Technology, The University of Technology Sydney, Ultimo, Broadway, NSW, Australia (e-mail: Ganesh.Naik@uts.edu.au).

Yina Guo is with Taiyuan University of Science and Technology, ShanXi Taiyuan 030024, China (e-mail: zulibest@gmail.com).

mutually independent. Let x and s be random vectors that contain the mixtures $x_1, x_2, \ldots, x_n$ and $s_1, s_2, \ldots, s_n$, respectively, and let A be the matrix with entries $a_{ij} = A_{ij}$. The aforementioned mixing model can then be written as

$$x = As \tag{1}$$

where x is an observed data vector, A is an unknown full rank mixing matrix, s, is an unknown non- Gaussian source process. The goal of ICA is finding the matrix, $W = A^{-1}$, so that the sources can be estimated from the vector x by optimizing a statistical independence criterion.

$$\hat{s} = Wx = W\,(As) \tag{2}$$

where $\hat{s}$ are the estimated sources up to permutation and scaling ambiguity. However in this process, the quality of the separation would also depends on the recorded signals; x. i.e. the quality of separation might be poorer for very low quality of recordings [6, 7, 9].

In general, physiological properties of the different sEMG recordings are unknown, and the purpose of the signal processing analysis is to find out the dissimilarities. Biosensor signals that need to be analysed have unknown features and properties. This makes it important to determine the reliability of the separation technique before employing the same on the bioelectric signals. Due to the unknown nature of the signal, it is difficult to measure the quality of separation using any features of the signal in the output. Hence, it is important to identify the objective quality of the separation, the technique would accomplish based on signal properties and mixing environment. The current techniques that have been developed for determining the quality of separation for audio based data are not suitable for biosensor signals because of the inability to relate it to information. For this reason, it is critical to reconstruct the separated sources for further analysis and post processing techniques.

In this research, we propose a novel method to reconstruct the improvised original recording, where it could be further used for source separation and other analysis. After computing the independent components (IC) as described before, a set of "*ICAreconstructed*" signals was obtained as follows. Initially, the frequency spectrum was computed for each IC. In the next step, the rows in the matrix W that corresponded to ICs with ic ≤ 0.10 (10% or less frequency content) were set to zero. After visual inspection of all IC waveforms and their spectrograms, a value of 0.10 was chosen as threshold because it was found that all ICs with ic ≤ 0.10 was adequate to keep up the power of the original recordings. The new matrix that was obtained, denoted by $\hat{W}$, is similar to the matrix W described before ($\hat{W} \approx$ W), with the difference that $\hat{W}$ does not contain information pertaining to ICs that contribute no or little to the content of the original signals x. After computing $\hat{A} = \hat{W}^{-1}$, we can likewise get rid of the columns in $\hat{A}$ analogous to the unwanted components and obtain a new matrix, denoted by Z, where Z $\approx \hat{A}$. By performing the operation, the original recorded signals (x) are reconstructed as

$$\hat{x} = Z\hat{W}\hat{s} = \hat{A}\hat{W}\hat{s} \tag{3}$$

which are similar to the original recorded signals x but with the difference that no information is included that comes from ICs that have ic ≤ 0.10.

III. Methods

To test the above theory, EMG data from the PhysioNet database (*courtesy of Seward Rutkove, MD, Department of Neurology, Beth Israel Deaconess Medical Center/Harvard Medical School*) was considered. They collected the data with a Medelec Synergy N2 EMG Monitoring System. They have used a 25mm concentric needle electrode into the tibialis anterior muscle of each subject. The patient was then asked to dorsiflex the foot gently against resistance. The needle electrode was repositioned until motor unit potentials with a rapid rise time were identified. Data were then collected for several seconds, at which point the patient was asked to relax and the needle removed [10].

The data were recorded at 50 KHz and then downsampled to 4 KHz. During the recording process, two analog filters were used: a 20 Hz high-pass filter and a 5K Hz low-pass filter. The EMG data set consist of the following categories:

- Healthy patients
- Patients with Neuropathy and
- Patients with Myopathy

The original recordings were initially separated using Fast ICA and later they were reconstructed using proposed method. The qualities of the original and reconstructed recordings were measured using the Amari index and Signal to Interference Ratio (SIR), which are one of the widely used evaluation methods in source separation problem [11]. Amari index and SIR is briefly explained below:

Amari performance index (API), it is also called as global rejection index defined by

$$A_\varepsilon(P) = \sum_{i=1}^{m}\left(\sum_{j=1}^{m}\frac{|p_{ij}|}{\max_k |p_{ik}|} - 1\right) + \sum_{j=1}^{m}\left(\sum_{j=1}^{m}\frac{|p_{ij}|}{\max_k |p_{kj}|} - 1\right) \tag{4}$$

where $P = p_{ij} = WA$. The API is a measure of diagonality of matrix. A perfect ICA separation will produce an identity confusion matrix (P) and therefore $A_\varepsilon(P) = 0$. Practically $A_\varepsilon(P)$ value should be very small for better separation [12]. SIR is the ratio of the power of the wanted signal to the total residue power of the unwanted signals. Generally, higher SIR value is desirable for better quality signals. Both SIR and Amari index were computed for several recordings.

IV. RESULTS AND DISCUSSIONS

The performance of the original recordings (x) and the reconstructed signal ($\hat{x}$) were tabulated in table I and II respectively.

TABLE I. AMARI INDEX VALUES FOR ORIGINAL AND ICARECONSTRUCTED SIGNALS

Samples	**ICAreconstructed**($\hat{x}$)	**Original** (x)
1	0.12	0.26
2	0.1	0.25
3	0.09	0.23
4	0.07	0.22
5	0.06	0.21
6	0.051	0.19
7	0.05	0.18
8	0.03	0.14
9	0.025	0.12
10	0.02	0.1

TABLE II. SIR VALUES FOR ORIGINAL AND ICARECONSTRUCTED SIGNALS

Samples	**ICAreconstructed**($\hat{x}$)	**Original** (x)
1	21	18
2	19	16
3	18.5	15.5
4	18	14
5	17	12.5
6	16	11.5
7	15.5	10
8	14.5	8.5
9	14	7
10	13	5.5

The above results were plotted for different sEMG signals (figure 1).

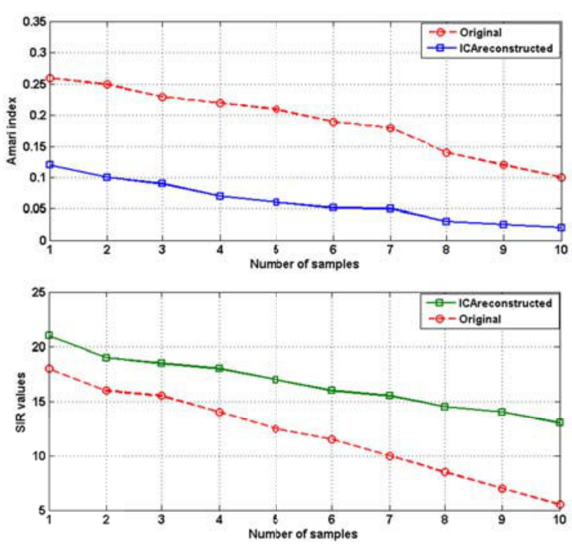

Fig.1. Amari index and SIR value plot for Reconstructed signal ($\hat{x}$) and original signal (x)

From the results listed in Table I, it can be seen that the "*ICAreconstructed*" signals showed lower Amari index values as compared to ICA separated sources. Also, from Table II it can be seen that "*ICAreconstructed*" signals have higher SIR values as compared to original ICA separated signals. It is evident from Figure 1 that the reconstructed signal ($\hat{x}$) showed lot of improvement as compared to the original signal (x). This has demonstrated the efficacy of the proposed technique.

V. CONCLUSIONS

There exist numerous ICA methods in literature for source separation. However, most of the instances the efficacy of the source separation depends on quality of the recordings. This research proposed a novel technique to reconstruct the original recordings by discarding the unwanted components in original recordings. This method could be considered as one of the filtering method, but it is arguable that by filtering some important information can be lost. Hence this method can be used for improving the quality of the original recordings (noisy or low quality) prior use of any ICA methods. The advantage of this technique is that it maintains the original quality of the recordings. Although, this technique is efficient in EMG signal analysis, it requires more experimental analysis to test this technique with other biomedical signals. Authors would like to conduct more experimental analysis to test this technique in near future.

REFERENCES

[1] D. Djuwari, D. K. Kumar, S. P. Arjunan, and G. R. Naik, "Limitations and applications of ICA for surface electromyogram for identifying hand gestures," *International Journal of Computational Intelligence and Applications,* vol. 07, pp. 281-300, 2008.

[2] J. S. Barlow, "Computerized Clinical Electroencephalography in Perspective," *Biomedical Engineering, IEEE Transactions on,* vol. BME-26, pp. 377-391, 1979.

[3] A. Bartolo, C. Roberts, R. R. Dzwonczyk, and E. Goldman, "Analysis of diaphragm EMG signals: comparison of gating vs. subtraction for removal of ECG contamination," *J Appl Physiol,* vol. 80, pp. 1898-902, 1996.

[4] R. Verleger, T. Gasser, and J. Möcks, "Correction of EOG Artifacts in Event-Related Potentials of the EEG: Aspects of Reliability and Validity," *Psychophysiology,* vol. 19, pp. 472-480, 1982.

[5] J. C. Woestenburg, M. N. Verbaten, and J. L. Slangen, "The removal of the eye-movement artifact from the EEG by regression analysis in the frequency domain," *Biol Psychol,* vol. 16, pp. 127-47, 1983.

[6] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Netw.,* vol. 13, pp. 411-430, 2000.

[7] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *Neural Networks, IEEE Transactions on,* vol. 10, pp. 626-634, 1999.

[8] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.,* vol. 7, pp. 1129-1159, 1995.

[9] G. R. Naik and D. K. Kumar, "An overview of independent component analysis and its applications," *Informatica: An International Journal of Computing and Informatics,* vol. 35, pp. 63-81, 2011.

[10] PhysioNet. (2013, Janauary). *Electromyography (EMG) Database*. Available: http://physionet.org/physiobank/database/emgdb/

[11] S. Haykin, *Neural Networks: A Comprehensive Foundation*: Prentice Hall PTR, 1998.

[12] M. Potter, N. Gadhok, and W. Kinsner, "Separation performance of ICA on simulated EEG and ECG signals contaminated by noise," in *Electrical and Computer Engineering, 2002. IEEE CCECE 2002. Canadian Conference on*, 2002, pp. 1099-1104 vol.2.