Removal of Electrocardiogram Interference from Diaphragmatic Electromyogram Signals using Sliding Singular Spectrum Analysis

Muzammil Saeed, *Student Member, IEEE* and Stephen R. Alty†, *Senior Member, IEEE*

***Abstract*— Over recent years, Singular Spectrum Analysis (SSA) has gained popularity as an effective means to de-noise biologically sourced single channel signals, especially Electromyogram (EMG) and Electrocardiogram (ECG) signals amongst others. There are numerous applications whereby the signal acquisition process results in the mixing of both types of signals along with body motion artifacts and the inevitable electromagnetic interference. Both ECG and EMG signals are very useful to physicians, though preferably in isolation, though they rarely present themselves in this manner. Simple filtering techniques are ineffective in their separation as both signal spectra overlap in the frequency domain. In this paper, we propose a technique based on a *sliding* SSA algorithm which proves to be more successful in separating real mixed EMG and ECG signals than traditional block based approaches on single channel data. SSA is a non-parametric technique that decomposes the original time series into a number of additive components, each of which can then be readily identified based on statistical analysis as belonging to EMG or ECG signals. This approach could be applied equally to other signal types using different statistical methods as required, moreover, this technique is relatively straight-forward to implement and does not require any reference signals or training.**

***Index Terms*— Singular Spectrum Analysis, Electromyogram, Electrocardiogram, Signal Separation.**

I. Introduction

The digital acquisition of biomedical signals is becoming increasingly common amongst biomedical engineers, physicians and researchers. These signals are prone to interference due to their low voltage nature. Electromyogram (EMG) and Electrocardiogram (ECG) signals are frequently inadvertently mixed in the human body in applications from simple ECG monitors, to surface EMG (sEMG) [1] and diaphragmatic EMG (EMGdi) acquisition systems. In this paper, we focus our attention on the application of EMGdi recovery from a real mixture of signals that were acquired using an esophageal catheter based device, see Figure 1. As can be clearly observed, this device is susceptible to significant levels of interference principally from ECG signals the source of which are in close physical proximity to the diaphragm muscles. The focus of previous work [2] was the application of Blind Source Separation (BSS) techniques using Independent Component Analysis (ICA) [3] to unmix the original signals from multichannel data which has been mixed in an assumed non-convolutive manner. This technique works well in a synthetic environment when the mixtures are simply additive; real sampled data, however, often doesn't lend itself to un-mixing so well. Nevertheless, some success has been reported using ICA both on multichannel data [2], [4] and single channel data [5] where another mechanism is used to split the data into quasi-multichannel data using either FFT, Wavelet or Empirical Mode Decomposition (EMD) techniques.

The work presented in this paper proposes that an SSA technique based on kurtosis be applied alone to the EMGdi data requiring only a single channel input and not only simplifies the signal separation process it also improves on the previous ICA based technique. Moreover, it is shown that results can be improved further by applying the SSA method using a *sliding* window approach, this method will be detailed in the rest of the paper which is organised as follows. The next section introduces the theoretical basis of the Singular Spectrum Analysis technique. The Experimentation section entails a comparison of Block SSA versus Sliding SSA on synthetically mixed EMG and ECG data for quantitative evaluation. Subsequently, there is a section of Results based on the application of the proposed sliding SSA method to the effective removal of ECG interference from sampled EMGdi data from an esophageal catheter. Finally, a Conclusions section provides a summary of the work at the end of the paper.

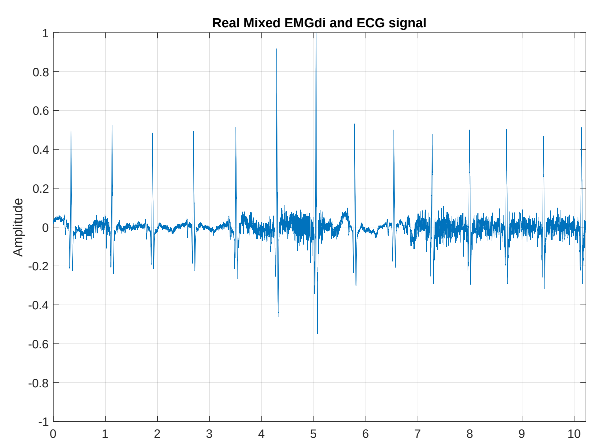

Fig. 1. Esophageal catheter acquired EMGdi data showing significant ECG interference.

Department of Electronic Engineering, Royal Holloway, University of London, Egham TW20 0EX, U.K.

†Please address all correspondence to `steve.alty@rhul.ac.uk`

II. Singular Spectrum Analysis

Singular Spectrum Analysis is a well established and powerful time-series analysis technique with relevance to multivariate statistics, dynamic systems and signal processing [6]. The applications of SSA are manifold, including signal source separation, financial modelling and biomedical signal denoising [7]. Essentially, SSA works by *embedding* a time-series into a Hankel matrix form, applying Singular Value Decomposition (SVD) to this matrix and then extracting the so-called *eigentriples*. Each *eigentriple* represent different components of the original signal including slowly varying trends, periodic components and unstructured noise. Hence, SSA can be used as an effective denoising or signal separation tool. There are typically two stages in the SSA process, *decomposition* and *reconstruction* and each stage consists of two steps.

A. Decomposition

The time series data is decomposed by first embedding it in a Hankel type matrix.

1) Embedding: Consider the real-valued non-zero time series $\mathbf{s}$ where

$$\mathbf{s} = (s_1, s_2, ..., s_{r-1}). \tag{1}$$

where $r > 2$ as a minimum. Embedding relates to the process of mapping $\mathbf{x}$ into k multidimensional lagged vectors of length l such that,

$$\mathbf{x}_i = [s_{i-1}, s_{i-2}, ..., s_{i+l-2}]^T \tag{2}$$

where $k = r - l + 1$, and the window length $1 \leq l \leq r$, and the superscript T denotes vector transpose operation. Selection of suitable window lengths l and lagged vectors k depends partially on prior knowledge of signals of interest and often ultimately relies on experimentation. To complete the embedding process a so-called *trajectory* matrix is formed from vectors of $\mathbf{x}_i$. The trajectory matrix is thus,

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k] \tag{3}$$

or

$$\mathbf{X} = \begin{bmatrix} s_0 & s_1 & s_2 & \cdots & s_{k-1} \\ s_1 & s_2 & s_3 & \cdots & s_k \\ s_2 & s_3 & s_4 & \cdots & s_{k+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{l-1} & s_l & s_{l+1} & \cdots & s_{r-1} \end{bmatrix} \tag{4}$$

$\mathbf{X}$ is clearly a Hankel matrix made up from the sample data whereby all its diagonal elements are equal.

2) Singular Value Decomposition (SVD): Here we define $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ and since $\mathbf{S}$ must be positive definite, we know that its eigenvalues $\lambda_1, \lambda_2, ..., \lambda_l$ must be positive valued also, further we place the eigenvalues of $\mathbf{S}$ in descending order so that they are monotonically decreasing in value from elements 1 to l, such that, $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_l \geq 0$ and $U_1, U_2, ..., U_l$ are the corresponding orthonormal eigenvectors such that $\|U_i\| = 1$.

If we now define $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$, then the trajectory matrix can be broken down into individual components such that,

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_d \tag{5}$$

where $d = arg(\max_i)\{\lambda_i > 0\}$ is often referred to as the *rank* of the matrix and $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$. The set of data elements $\{\sqrt{\lambda_i}, U_i \ \& \ V_i^T\}$ is often referred to as the ith *eigentriple* of the trajectory matrix $\mathbf{X}$ [8]. Each eigentriple contains progressively different statistical natured sub-components of the original time-series data and careful selection or *grouping* of the components can yield useful results in many source separation problems.

B. Reconstruction

The reconstruction stage comprises two parts, namely *grouping* and *diagonal averaging*.

1) Grouping: There is no general rule for grouping, it all depends on the application, the grouping rule will be defined by the special requirements of the problem and the statistical nature of the signal of interest or even its noise. In our case we are primarily interested in separating ECG signals from EMG signals and therefore we can use statistical means to select the eigentriples which correspond to the relevant signal distribution function. Experimentation has shown that the fourth standardized moment, *kurtosis*, (see Eq. 6) is an effective statistical measure to differentiate ECG and EMG type signals; typically ECG has a significantly higher kurtosis figure than that of EMG being more noise like in nature its kurtosis is typically less than 4. A simple threshold can be set to group all eigentriples with low kurtosis as EMG and the rest to be ECG. Further, a novel *sliding window* approach is adopted, which improves upon the standard block-based methods by avoiding arbitrary end effects caused by abrupt transitions of blocks. Whilst this method increases computational complexity significantly the results show marked improvement, the algorithm runs approximately one order of magnitude slower than real-time on a modern PC with Intel® i5-6600K processor running at 3.5GHz using Matlab® on a 64-bit Windows® 10 PC. Research is currently under way to reduce the computational burden of the SVD algorithm when using a sliding window.

$$\kappa(\mathbf{x}) = E\left[\left(\frac{\mathbf{x} - \mu}{\sigma}\right)^4\right] \tag{6}$$

Kurtosis is a useful and simple statistical measure that can be applied to each eigentriple $\mathbf{X}_n$ which can then be grouped accordingly.

$$\hat{\mathbf{X}} = \mathbf{X}_1 + \mathbf{X}_2 + \ldots \tag{7}$$

2) Diagonal Averaging: The final stage of the reconstruction process involves the reconversion from matrix back to vector format and this is achieved by averaging the diagonal elements of the reformed Hankel matrix $\hat{\mathbf{X}}$ which is simply a sum of the subset of desired eigentriples according to the grouping rules stated in the previous sub-section. Once the new Hankel matrix is determined it is straight forward to *de-embed* the new time-series data. Normally, one Hankel matrix is reconstructed of entirely ECG selected eigentriples and then the resultant time series is simply subtracted from the original mixed times series to yield the de-noised EMG signal trace.

III. Experimentation

In order to evaluate the performance of the proposed method under varying signal-to-noise conditions, synthetic data was generated using ECG and EMG data acquired from standard online corpora. The ECG data was specifically sourced from the PhysioNet hosted MIT-BIH database [9] and the EMG data was sourced from the SENIAM EMG database [10]. The two signals were mixed linearly at different levels of signal-to-noise ratio (SNR), here we considered the ECG to be the source of noise and the EMG the desired signal of interest. SNRs of 20, 15, 10, 5 and 0 dB were created and these were tested using both block and sliding based algorithms for different thresholds of kurtosis, (κ = 7, 6.5, 6, 5.5, 5, 4.5, 4, 3.5 and 3). The results were quantified by determining the relative root mean square error (RRMSE) accordingly [5]:

$$RRMSE = \frac{RMS(\mathbf{x}_{emg} - \hat{\mathbf{x}}_{emg})}{RMS(\mathbf{x}_{emg})} \tag{8}$$

where

$$RMS(\mathbf{x}) = \sqrt{\frac{\mathbf{x}^T\mathbf{x}}{N}} \tag{9}$$

and

$$SNR = 20log\left\{\frac{RMS(\mathbf{x}_{emg})}{RMS(\mathbf{x}_{ecg})}\right\} \tag{10}$$

Figure 2 shows the results of the simulated runs for five different values of SNR and nine different values of kurtosis threshold for block SSA. It can be seen that there is a downward trend in RRMSE as SNR increases as would be expected. It can also be seen that RRMSE value seems to plateau for Kurtosis threshold values between 4 and 5. This is consistent with the knowledge that typical EMG signals are generally akin to white noise and have a lower value of kurtosis close to 3, whereas typical ECG signals are super-Gaussian in nature and therefore have a high Kurtosis value usually in excess of 10. The results from the sliding SSA method, see Figure 3 show lower RRMSE values again at each SNR level, whilst confirming the optimal threshold for κ to be around 5. This information was then used to inform the application of the SSA algorithm to the real EMGdi data. Figure 4 shows the result of applying the block-based SSA algorithm to the synthetic test signal. The sliding window based SSA algorithm was then applied to the same signal and results shown in Figure 5. The comparison shows that the sliding window method has a clear advantage over the block based method.

IV. Results

The real EMGdi data used in this paper was acquired from patients at the Royal Brompton Hospital[1] suffering

[1]The Institution's Ethical Review Board approved all experimental procedures involving data obtained from human subjects.

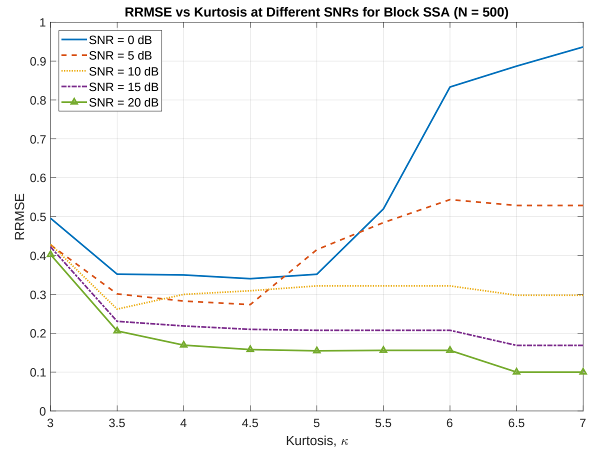

Fig. 2. RRMSE vs Kurtosis, κ, for Block Based SSA.

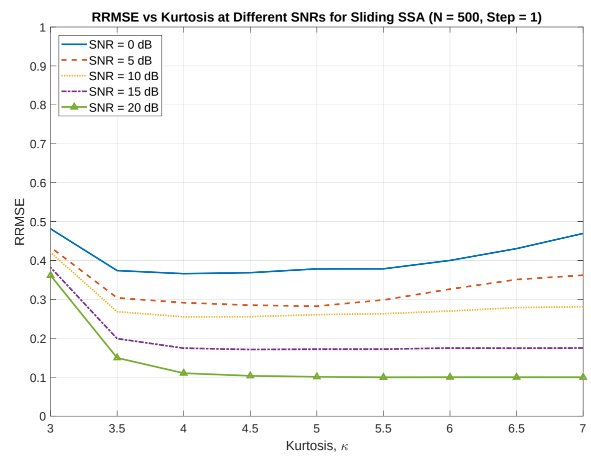

Fig. 3. RRMSE vs Kurtosis, κ, for Sliding SSA.

from some form of respiratory disease. Figure 1 shows a 10s segment of the test signal that originates from one channel of data acquired from the esophageal catheter which shows significant unknown mixing of EMG and ECG signals. Figure 6 shows the results from the application of the block-based SSA algorithm to the test signal. Figure 7 shows the results from the sliding window based SSA algorithm when applied to the same signal. Whilst the results for the block approach are quite reasonable, especially considering the challenging source of the data, some ECG artifacts remain. The sliding window method, however, shows a clear advantage over the block based method with virtually no evidence of ECG interference remaining. Our work improves upon the standard SSA result significantly, the initial settings of the SSA method remain unchanged; the window length used was, $N = 500$ samples, the depth or rank of the decomposition was set to, $r = 50$ for both methods as was the kurtosis threshold, κ, where experimentation on synthetic mixed data yielded a threshold of around 5 to be optimal.

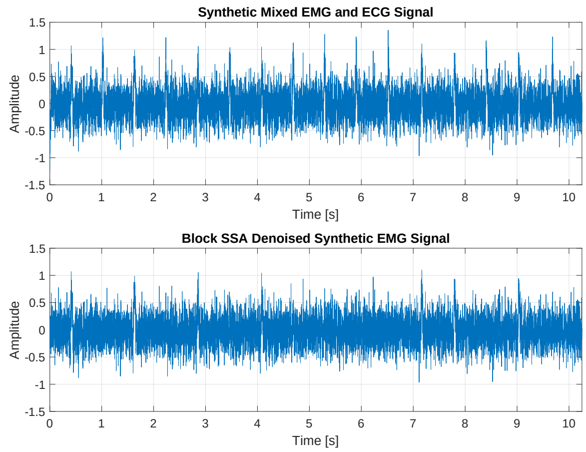

Fig. 4. Synthetic mixed EMG and ECG (5dB) de-noised using $\kappa = 5$ with block based SSA (RRMSE = 0.4192).

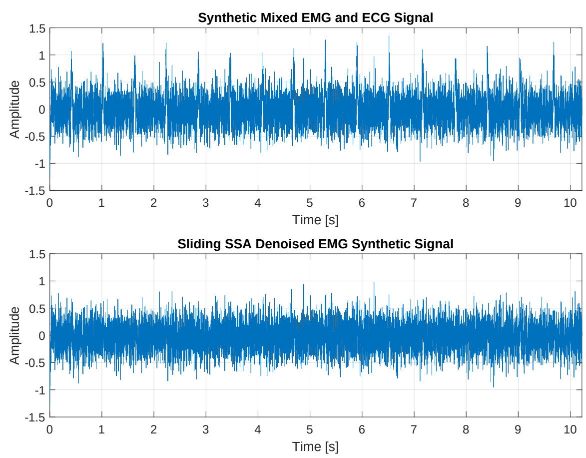

Fig. 5. Synthetic mixed EMG and ECG (5dB) de-noised using $\kappa = 5$ with sliding SSA (RRMSE = 0.2874).

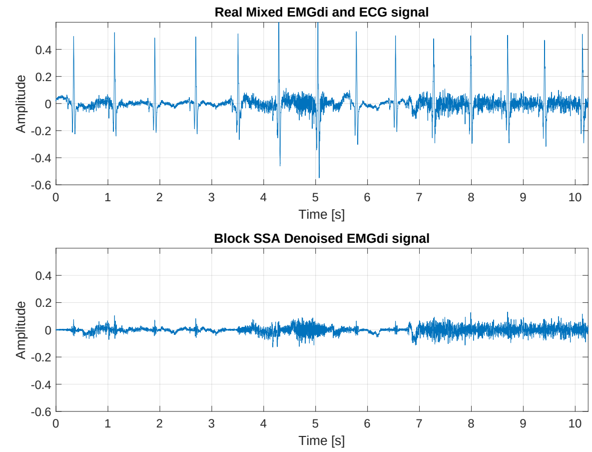

Fig. 6. Real mixed EMG and ECG data de-noised using $\kappa = 5$ with block based SSA.

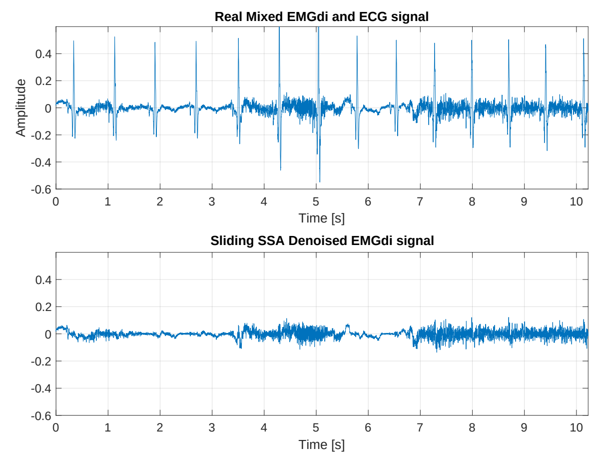

Fig. 7. Real mixed EMG and ECG data de-noised using $\kappa = 5$ with sliding SSA.

V. Conclusions

The results of the sliding window SSA approach shows clear improvements over the previously published work in this field. Future work could include computational improvements using an iterative SVD algorithm to reduce the computational complexity to facilitate real-time implementation.

References

[1] J. Barrios-Muriel, F. Romero, F. J. Alonso, and K. Gianikellis, "A simple ssa-based de-noising technique to remove ecg interference in emg signals," *Biomedical Signal Processing and Control*, pp. 117–126, 2016.

[2] S. R. Alty, W. D-C. Man, J. Moxham, and K. C. Lee, "Denoising of diaphragmatic electromyogram signals for respiratory control and diagnostic purposes," *30th Annual International Conference of the IEEE EMBS (EMBC)*, pp. 5560–5563, September 2008.

[3] E. Oja, A. Hyvärinen, and J. Karhunen, *Independent Component Analysis*, Wiley Press, 2001.

[4] J. D. Costa Junior, D. D. Ferreira, J. Nadal, and A. M. F. L. Miranda de Sá, "Reducing electrocardiographic artifacts from electromyogram signals with independent component analysis," *32nd International Conference of the IEEE EMBS (EMBC)*, pp. 4598–4601, 2010.

[5] Z. Zhang, H. Li, and D. Mandic, "Blind source separation and artefact cancellation for single channel bioelectrical signal," *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 2376–8894, 2016.

[6] H. Hassani, "Singular spectrum analysis: Methodology and comparison," *Journal of Data Science*, vol. 5, pp. 239–257, March 2007.

[7] S. Sanei and H. Hassani, *Singular spectrum analysis of biomedical signals*, CRC Press, 2015.

[8] T. Alexandrov, "A method of trend extraction using singular spectrum analysis," *REVSTAT Statistical Journal*, vol. 7, pp. 1–22, April 2009.

[9] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, Ivanov P.Ch, R.G. Mark, J.E. Mietus, G.B. Moody, C-K. Peng, and H.E. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/content/101/23/e215.full];*, 2000.

[10] H. Hermans et al, *SENIAM project CD*, Roessingh Research and Development, Enschede, Netherlands, 1999.