Taylor & Francis
Taylor & Francis Group

Parameterization and R-peak error estimations of ECG signals using independent component analysis

M. P. S. CHAWLA*

Department of Electrical Engineering, Indian Institute of Technology, Roorkee 247667 India

Principal component analysis (PCA) is used to reduce dimensionality of electrocardiogram (ECG) data prior to performing independent component analysis (ICA). A newly developed PCA variance estimator by the author has been applied for detecting true, actual and false peaks of ECG data files. In this paper, it is felt that the ability of ICA is also checked for parameterization of ECG signals, which is necessary at times. Independent components (ICs) of properly parameterized ECG signals are more readily interpretable than the measurements themselves, or their ICs. The original ECG recordings and the samples are corrected by statistical measures to estimate the noise statistics of ECG signals and find the reconstruction errors. The capability of ICA is justified by finding the true, false and actual peaks of around 25–50, CSE (common standards for electrocardiography) database ECG files. In the present work, joint approximation for diagonalization of the eigen matrices (Jade) algorithm is applied to 3-channel ECG. ICA processing of different cases is dealt with and the R-peak magnitudes of the ECG waveforms before and after applying ICA are found and marked. ICA results obtained indicate that in most of the cases, the percentage error in reconstruction is very small. The developed PCA variance estimator along with the quadratic spline wavelet gave a sensitivity of 97.47% before applying ICA and 98.07% after ICA processing.

Keywords: Electrocardiogram; Parameterization; Quadratic spline wavelet; PCA variance estimator; Feature extraction; Validation; Principal component analysis; Independent component analysis

1. Introduction

Principal component analysis (PCA) and independent component analysis (ICA) occupies a definite place in higher order statistical techniques for better feature extraction and electrocardiogram (ECG) interpretation [1,3]. PCA finds a set of the most representative projection vectors such that the projected samples retain the most information about original ECG samples [2,7,31]. ICA captures both second and higher-order statistics and projects the input ECG data onto the basis vectors that are as statistically independent as possible [3].

ICA opens new and useful windows into phenomena contained in multi-channel ECG records by separating ECG data recorded at multiple electrodes into a sum of independent

*Email: mpschawla@rediffmail.com; mpschawla@gmail.com

components. ICA is a new technique suitable for separating independent components from ECG complex signals. In many cases, the independent ICA components are also functionally independent. In particular, ICA appears to be a generally applicable and effective method for removing a wide variety of artifacts and noise from ECG records. However, PCA cannot completely separate artifacts and noise completely from ECG signals, especially when they have comparable amplitudes. ECG is oscillatory in nature, although not periodic in the strict mathematical sense. The analysis of ECG signal is extensively used as a diagnostic tool to provide information on the heart function. ECG signals are largely employed as a diagnostic tool in clinical practice in order to assess the cardiac status of an object. The ICA may actually be regarded as a refinement on the PCA results which provides a remarkable improvement in the source estimation [3,4,7]. Figure 1 shows the basic ECG waveform.

A number of ICA algorithms that incorporate the temporal structure of the ECG sources have been developed in the past and are available in literature. The most commonly used ICA algorithms are: Fast-ICA, Temporal Fast-ICA and Temporal decorrelation source separation [5,6,11,29,30].

In this paper, it is suggested that the ability of ICA for parameterization of ECG signals may be necessary at times for removal of redundancy if available in the ECG data. Properly parameterized ECG signals provide a better view to the extracted ECG signals, while reducing the amount of ECG data preserving diagnostic morphology [4,12,20,21].

In the present work, Jade algorithm is applied to three-channel ECG and three CSE database files, and the ECG waveforms are separated as independent components. |Kurt| and Varvar denote respectively the modulus of kurtosis and variance of variance as first proposed in [1] and used as an application in this study. It is verified and reimplemented in this analysis that the independent components whose |Kurt|⟨Kurt (Threshold) and Varvar⟩Varvar (Threshold) is taken as a noise or artifact component [2–4,31].

Thresholds for Kurtosis = 4.3 and Varvar = 0.4 are obtained after checking on around 20 CSE database files after parameterization of ECG signals, which show improvements
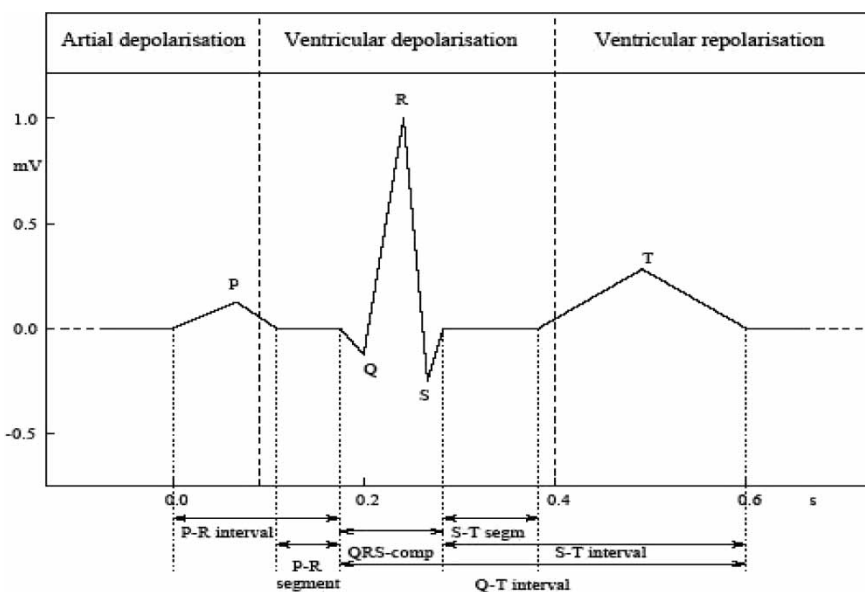

Figure 1.   ECG Waveform (Courtsey: Wouter. A.Th. Manintveld, UK, 1996, Thesis).

in results when compared to those presented in [1]. This study is a reimplementation of [1] as far as thresholds for kurtosis and variance of variance (Varvar) are concerned.

It is also confirmed after various case studies that ICA yields Independent Components (ICs) displaying more clearly the investigated properties of the original ECG sources. ICs of properly parameterized ECG signals [21] may also be more readily interpretable than the measurements themselves, or their ICs. The applied ICA algorithm in this study estimates the independence of the original signal, and the optimization based on the estimation searches an optimum-restoring matrix. In the proposed model, estimated results are in good agreement with the physiological view [4,9,10,21].

1.1 Existing classical ECG techniques: a review

In past years, several ECG analysers based on statistical methods, clustering methods, expert systems and Markov models have been developed and implemented to solve the problem of the time consuming ECG analysis. Their lesser reliability, together with their high sensitivity to noise and their failure to deal with new or ambiguous patterns, leads the research towards investigation of new analysis techniques.

Recently, there have been many approaches introduced involving techniques for computer processing of 12 lead ECG, in order to diagnose a certain disease. A first group of methods to interpret the ECG significance uses a morphological analysis. A second group of techniques for computer analysis of ECG uses statistical models. A third category of methods corresponding to neural models becomes a powerful concurrent to statistical ones for ECG signal classification. By integrating statistical and knowledge-based techniques, a system can be developed which is more robust than a system consisting of only the individual techniques [22–24].

Artificial neural networks have been often proposed as good classifiers when non linear separation borders are required and incomplete or ambiguous input ECG patterns can be found. In the last few years, the connectionist approach has been applied to the ECG analysis with promising results [18,20].

PCA can be used for reducing dimensionality in a dataset while retaining those characteristics of the dataset that contribute most to its variance by eliminating the later principal components (by a more or less heuristic decision). These characteristics may be the 'most important', but this is not necessarily the case, depending on the application. The linear PCA can be implemented with powerful, robust techniques as the singular value decomposition (SVD) that guarantees numerical accuracy and stability. A relatively successful method to deal with these problems is the use of patient-adaptive algorithms based on PCA. PCA is expected to set a milestone in this effort. PCA can provide descriptive contributions from each ECG lead and can represent a global measure of the atrial and ventricular activity [2,7,10,14].

Some undesired signals may be superimposed on ECG signals of interest and must be considered artifacts. These artifacts may have a biological origin, like reciprocal contaminations of muscle activity and heart cycles respectively in ECG and electromyogram (EMG) recordings. These motion artifacts produce base line drifts which can compromise vital signs parameters extrapolation [1,7,8].

1.2 Motivation of higher order statistics

The main motivation behind the use of higher order statistics (HOS) lies in their ability to suppress noise under certain conditions, without having to know the exact probability density

function governing the noise samples [7,10–12]. The limitations of second order statistics could significantly be reduced by the application of HOS to ECG, to facilitate accurate interpretation of the spatial sampled signal. However, the application of HOS to finite length and noisy ECG data poses a problem of its own and the use of particular order cumulant can help to investigate this problem [15,16]. One limitation that keeps the standard (3 or 12-lead system) ECG from providing a more comprehensive and transparent description of the electrical state of the heart is that it is a very sparse sampling of a complicated, spatially varied distribution of potential on the body surface. Another limitation is to do with the fact that the application of second-order statistics to ECG signals definitely presents a shortfall in analyzing and interpreting complicated interactions between the mechanical function of the heart, its internal electrical behavior, and externally recorded potentials.

1.3 Limitations of standard ICA

ICA is a multidimensional signal processing technique to separate signals from different 'sources' into distinct components. Once separated, components classified as noise may be discarded and the remaining components used to reconstruct the 'pure' signal [1,4]. However, due to the imperfect nature of the ICA technique, substantial data may be lost when the 'noise' components are removed or discarded [7,12,31].

When interpreting the ECG results using ICA, it is to be noted that:

(1) At times the energies of the ICs are undetermined; contributions of the ICs to the measurements can nevertheless be assessed by examining the corresponding elements of the mixing matrix **A**.
(2) On the contrary determined minor ICs may display results that produce otherwise hard to observe desired or expected information.

1.4 ICA classifiers

The purpose of this study was to evaluate the ability of the ICA technique to extract the noise-free ECG signal from ECG recordings, after removing artifacts and noise. The clean ECG signals are reconstructed dropping those components related to the artifacts and noise [1–3]. ICA is an extensive technique that extracts independent components from ECG mixed signals. A hypothetical clinical application is to remove artifacts and noise from ECG using ICA [11,12,22,25].

Why is ICA preferred in ECG analysis when compared to the other classifiers? The answer to this question is:

  i) ICA is a form of blind source separation
 ii) It can solve and handle time delay and ambiguity problems in the ECG data.
iii) ICA assumes the ECG sources as linear mixtures
 iv) ICA allows physicians an alternative higher order statistical technique for better ECG interpretation.
  v) It performs better at times in yielding a cleaned ECG signal when compared to higher order digital filters and gives comparable performance.
 vi) ICA is better at recovering specific points on the ECG such as the R-peak, RR interval which is necessary for obtaining the heart rate.

The above six points (i)–(vi), indicates the reasons and advantages of using ICA classifiers for ECG analysis [2–4,22].

2. Basics of ICA and need of parameterization

Based on trivial properties of convolution, some common signal parameters also fulfil the ICA mixing model, given that the original ECG measured signals comply with it. By ECG signal parameterization, it is meant that the construction of a new ECG signal is possible from any local or global properties, *i.e.* parameters, of the original ECG signal [4,22]. These properties may be related to *a priori* known features of the ECG sources. Even with proper ECG parameterizations, it may sometimes be hard or impossible to make an ICA algorithm converge, because of missing or otherwise bad ECG data due to, for example, bad electrode contacts [2–4]. It is also, otherwise, possible that the ICA does not converge, or that several runs of ICA are needed. R-wave reflects the intensity and direction of progressing ventricular heart muscle depolarization, *i.e.* contraction [11,12,22,31].

ICA is a statistical technique for obtaining independent sources, $\mathbf{s}$, from their linear mixtures, $\mathbf{x}$, when neither the original sources nor the actual mixing matrix [15–17,22], $\mathbf{A}$ are known as given by equation (1). This model is easily applied by exploiting higher order signal statistics and optimization techniques.

The basic noisy ICA ECG model is denoted by (1) as

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{N} \tag{1}$$

For a set of p random variables,

$$\mathbf{x}(t) = [x_1, x_2, x_3, \ldots, x_p]^T$$

assumed to be a linear combination (represented by a mixing matrix $\mathbf{A}$) of q unknown statistically independent sources,

$$\mathbf{s}(t) = [s_1, s_2, s_3, \ldots s_q]^T$$

where, $q < p$ such that:

$$\mathbf{x}(t) = \mathbf{A}\,\mathbf{s}(t) \tag{2}$$

ICA aims to find the de-mixing matrix $\mathbf{w}$, such that

$$\mathbf{s}(t) = \mathbf{w}\mathbf{x}(t) \tag{3}$$

where $\mathbf{w} = $ unmixing matrix.

The de-mixing matrix thus helps to find the sources $\mathbf{s}(t)$. To simplify the estimation of these independent sources, we start by decorrelating the mixtures, (whitening or sphering). This makes the covariance matrix of $\mathbf{x}$ diagonal and its components of unit variance. ICA then uses higher order statistical information (kurtosis, negentropy, etc.) to estimate the independent sources [1,4,22]. The prime requirement for applying ICA is that a number of simultaneously measured ECG signals carry linear combinations of the original source ECG signals, where

A is m × n complex mixing matrix
s represent the source ECG signals
N is a noise vector

Equation (3) can be reduced to equation (1), if **N** is zero or made to zero.

In certain situations, it is desired that the ICs of the parameterization of ECG signals is to be done in order to display diagnostic information carried by the parameters more clearly than what is observable from the original ECG measurements or their ICs. It is to be noted that in order for the ICA mixing model (1) to be valid, the parameters have to be derived from the ECG signal amplitudes. ECG parameters describing time durations, *e.g.* the time periods between consecutive R-waves, do not comply with the ICA mixing model [2,3,22,31]. In some cases, the appropriate ECG parameterization may be such that it greatly decreases the amount of ECG data, thus lowering the computational burden on the subsequent ECG analysis. As readily stated by the mixing model (1), each and every measured sample is a linear combination of the samples of the source ECG signals at the same time. Therefore, a set of ECG signals, in which each signal consists of samples from the corresponding original ECG measurement at the time points 's', *i.e.*

$$\boldsymbol{y'} = \boldsymbol{y}(\text{s}) \tag{4}$$

also satisfies the ICA mixing model (1). Also, it can be shown that ECG signals constructed from time averages of the original measurement ECG samples, or signals resulting from finite impulse response (FIR) filtering, fulfil the ICA signal model, and may thus be subjected to ICA. However, it is to be noted that even if the ICA algorithm converged and produced ICs, averaging will make the components more Gaussian. Thus, one must pay special attention to the proper interpretability of the components, and avoid lengthy averaging windows [4,21,22,31].

2.1 PCA preprocessing

Signal processing in general has tremendously changed during the last 20 years and it is expected to change even more in the years to come. What was earlier visualized as digital signal processing now forms only a small part of the new concept of signal processing which might be more adequately explained as the methods of analyzing, manipulating and conveying natural information. Feature extraction is basically reduction of the available information maintaining ECG morphology. Features are representatives of identification to a particular subject or specimen. Analysis and feature extraction from electrocardiograms is difficult until and unless artifacts and noise from the ECG are removed; there are many techniques available in the literature [1,2,13,14]. Figure 2(a) shows PCA based scheme for ECG data compression whereas figure 2(b) depicts the PCA scheme for ECG processing.

The ECG features of interest are the various intervals and segments of ECG. A preprocessing step based on PCA is recommended, since after pre-whitening, the real ECG sources and the whitened vectors are just related through an orthogonal transformation [2,7,10]. As long as the imposed conditions are fulfilled by the real ECG data, this method provides better performance than traditional ICA techniques because the new information is included in the proposed algorithm. The proposed method is tested on real ECG and CSE database signals. At times in the beginning of the ECG processing, PCA is used for base line wander removal, which could be better choice. In the present work, PCA offers significant advantages in removing base-line wander (BLW) from ECG as compared to the

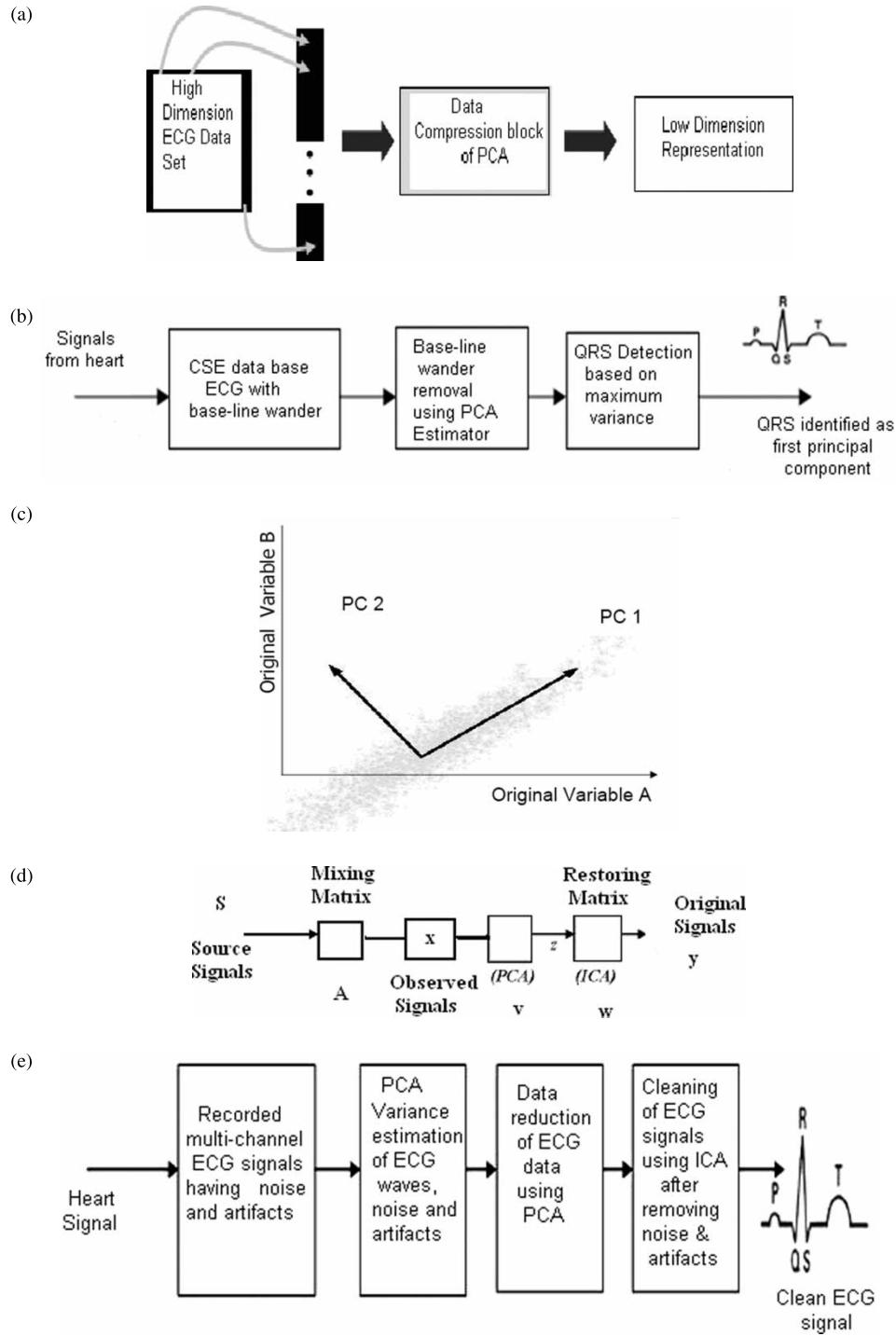(a)

(b)

(c)

(d)

(e)

Figure 2.   (a) PCA scheme for ECG data compression. (b) PCA scheme for ECG processing. (c) Basis for PCA vectors. (d) Combined PCA-ICA transformation scheme. (e) PCA based ICA scheme for dimension reduction and segment classification.

aforementioned methods [7,14,22,31]. Figure 2(d) depicts the PCA transformation scheme for ECG classification and figure 2(e) illustrates a PCA based scheme for ECG data dimension reduction.

In this paper, for classification purposes, a feature vector from ECG time samples is constructed from each ECG lead of a data set. This vector is of definite length and the R-wave peak is used as the reference point of this vector. The R-wave peak is the point where the difference between the next slope and the previous slope in QRS region is a maximum. Some QRS detection methods reduced the amount of information that medical practitioners needed to process in time but had to suffer critic scenarios, loosing diagnostic features. In the proposed method, first the peak of the QRS complex is detected with its high dominated amplitude in the signal using PCA variance estimator, followed by detection of Q and S-waves [2,7,22].

2.2 Modeling steps of PCA

PCA is the optimal linear technique which retains the maximum amount of variance (amongst all linear projections) within the projected feature space. The main drawback of PCA lies in its global linearity; since the algorithm finds only a linear subspace of the original data space, it is sub-optimal when the underlying structure in the data is inherently nonlinear. PCA uses projections onto an 'orthogonal basis' set to separate the ECG signal from the noise [3,7,22]. Figure 2(c) shows the basis for PCA vectors.

The initial stage referred to as PCA has three main purposes:

1. To estimate the number of ECG signals.
2. To remove the second order correlations between the temporal ECG waveforms.
3. To normalize the temporal ECG waveform vectors.

To estimate the number of ECG signals, the eigen values or singular values are compared against a detection threshold. The vectors relating to the noise subspace are then ignored. If the number of ECG temporal samples is large, it may be computationally more efficient to implement spatial eigen decomposition than SVD. The orthonormal temporal vectors can be constructed from the spatial eigenvectors [2,7,8].

The following simplified steps are required in modeling [7,22], the PCA based ECG analysis:

Step 1: get the ECG data
Step 2: subtract the mean
Step 3: calculate the ECG covariance matrix
Step 4: calculate eigenvectors and eigenvalues of the covariance matrix
Step 5: Choose principal components and form ECG feature vectors

$$\text{Feature vector } \mathbf{v} = (\text{eig}_1, \text{eig}_2, \text{eig}_3, \cdots \text{eig}_p)$$

.
Step 6: deriving the new ECG data set

$$\text{Final data} = \text{Row feature vector} \times \text{Row data adjust}$$

Step 7: getting the old ECG data back, *i.e.* reconstruction of ECG data

Step 8: reconstruct the original dimensionality of the ECG data.

2.3 Extraction of independent components

An ICA method is proposed where additional knowledge about the time and statistical structure of the ECG sources is incorporated. ICA yields equations describing the behavior of the various ECG segments as a function of cardiac cycle time. ICA can be used to synthesis an ECG signal which is a realistic reproduction of the original signal, and also can control parameters such as QRS complex amplitude, rise-time, fall-time and the relative amplitudes of the P- and T-waves. The duration of each component will automatically track the selected heart rate in a non-linear fashion, reflecting its true behavior. This will provide an invaluable as well as cost-effective tool for testing, calibrating and maintaining electrocardiographic equipment in hospitals and clinics, and for the design/improvement of new and existing instrumentation [17–19].

PCA is not a very appropriate technique for the visualization of ECG data and nonlinear dimensionality reduction algorithm from the ECG morphological point of view. This is due to the fact that it can only uncover linear relationships in the ECG data, and is designed to find the directions in the ECG data with the highest variance, which may not always be the most informative directions. The resulting low-dimensional PCA descriptors can be used for exploratory ECG data analysis, visualization and subsequent ECG data modeling. It has been reported in the literature that the ICA source estimates promote some inverse process, acting on the ECG observations. In a standard ICA model, this inverse function is a generalized linear function, *i.e.* a function of the pseudo-inverse of the mixing matrix, the so-called unmixing matrix and the inferred noise model. ICA is expected to remove noise from the ECG with known characteristics [20–22]. Figure 3 shows the proposed scheme of ICA for ECG analysis.

In this paper, a method for removing artifacts and noise from ECG signals acquired by blind source separation (BSS) and ICA is presented. Since in some situations, both artifacts and signals of interest show common features in different ECG channels, hence correction of ECG signals by ICA was felt for diagnostic information.

3. ICA steps

First step involves the determination of independent components by removing the mean values of the variables also known as 'centering the ECG data'. The second step is to 'whiten the ECG data' also known as 'sphering the ECG data'. In the third step, independent components are obtained by applying a linear transformation to the whitened ECG data [1,22–24].
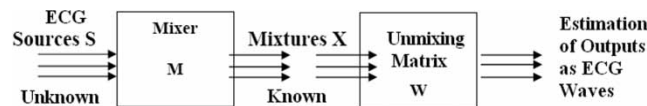

Figure 3.   Basic understanding of ICA application for ECG analysis.

To estimate one of the independent components, a linear combination of the x_i is considered. Let us denote this by

$$y = \mathbf{w}^T x = \sum_{i=1}^{n} w_i x_i, \tag{5}$$

Where the column vector $\mathbf{w}$ is to be determined [2,19,22,31].

The independent components are determined by applying a linear transformation to the whitened data. A given component can be obtained using the linear transformation

$$\mathbf{i_c} = \mathbf{b_i^T x}, \tag{6}$$

where

$\mathbf{i_c}$ gives independent components which is an estimate of the original signal.
$\mathbf{b}$ is an appropriate vector to reconstruct the independent components [25–27].

In order to employ the ECG signal for facilitating interpretation and medical diagnosis, ICA is used to clean the ECG signal by removing some or all the sources of noise. By using ICA, the basic idea is to 'project out' the noise and artifacts from ECG signals and to represent noise and artifacts as independent components [22,28–30].

The ICA algorithmic principle is

$$\mathbf{y} = \text{Est}(\mathbf{s}) = \mathbf{w}_{\text{ICA}}\mathbf{x}, \tag{7}$$

$$\mathbf{y} = \mathbf{w}\mathbf{x}. \tag{8}$$

The idea of ICA is to recover the original signals by assuming that they are statistically independent. $\mathbf{y}$ is independent and it is desired to find how $\mathbf{w}$ maximizes the independence of $\mathbf{y}$. After estimating $\mathbf{A}$, computation of $\mathbf{w} = \mathbf{A}^{-1}$ is done, which gives

$$\mathbf{s} = \mathbf{w}\mathbf{x} = \mathbf{A}^{-1}\mathbf{x}. \tag{9}$$

Using ICA, it is required to derive a 'clean ECG signal' from the source ECG signal to find the noise reduction factor and compare the proposed model with the existing methods and algorithms [2,22,30]. ICA steps for artifacts/noise removal and feature extraction is depicted in figure 5.

4. Simulations

The simulation results of this paper suggest that the integration of PCA and ICA techniques can efficiently remove the noise and artifacts from the ECG signals. The results of PCA and ICA when used together showed good efficacy and informative ECG classification.
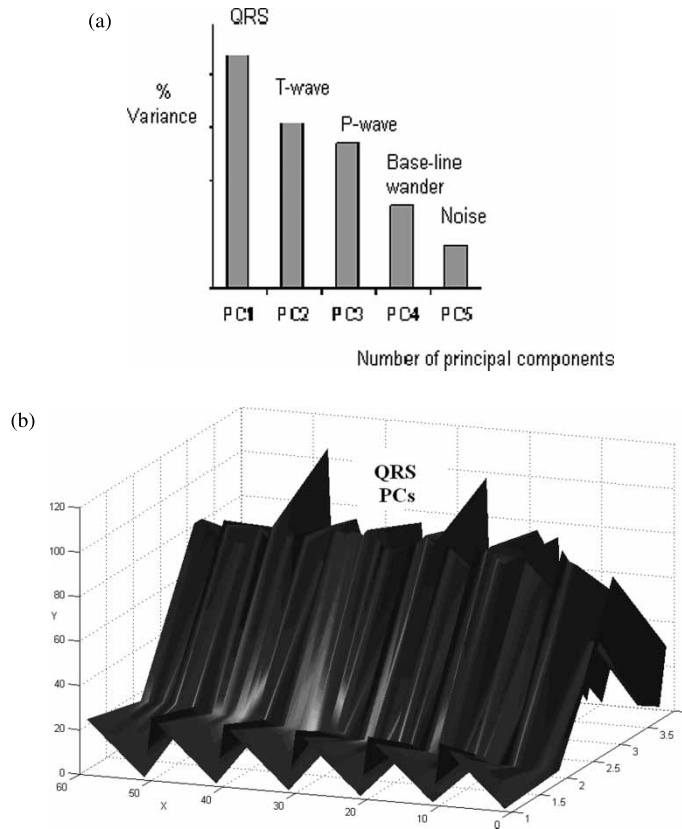
Figure 4. (a) Representation of principal components in an ECG lead in the order of their variance magnitudes. (b) Plot of no of QRS PCs obtained in a lead of an ECG database file.

4.1 PCA results

The idea behind PCA is to have an effective linear coding method for multivariate ECG data [2,7,22]. The various algorithmic steps used are:

Step 1: Pre-process the input ECG data and identify features as ECG segments.

Step 2: Choose the initial set of ECG features.

Step 3: Select features as essential and non-essential features based on PCA scatter or scree plots for ECG data set.

Step 4: Find variances with ECG data set for different combinations of features.

Step 5: Choose the best possible set of ECG features based on the ability to reach the desired mean square error (MSE), by developed PCA variance estimator.

Step 6: Vary the threshold for the chosen ECG feature set.

Step 7: Choose the best principal components based on morphological basis and an ability to reach the predefined MSE.

Step 8: Final choice is based on generalization ability of the variance estimator to the test ECG data and separate noise and artifacts.

Step 9: Test performance of variance estimator for more noisy data and check the ability of PCA to separate useful ECG components from noisy components.
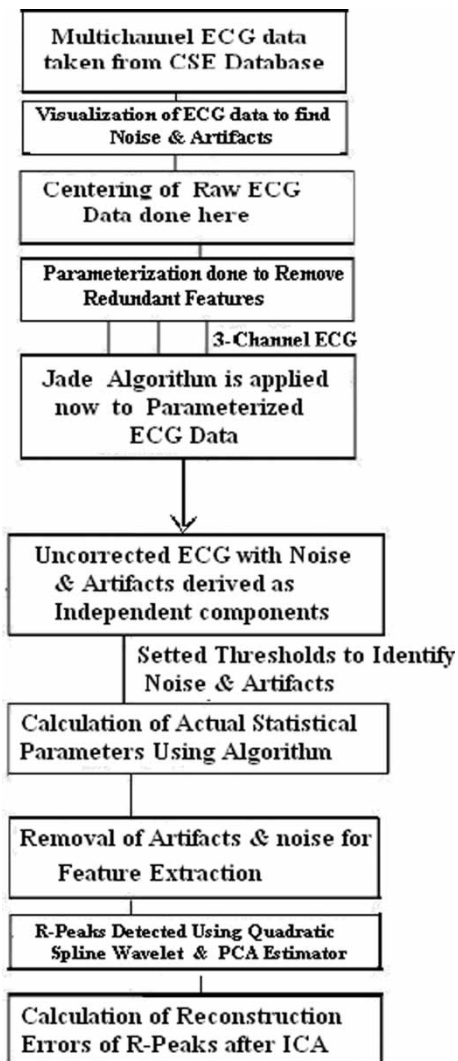
Figure 5.   ICA flowchart for R-peak detection.

Step 10: Select the useful ECG components which have more variance as compared to noisy components.

A different newly developed algorithm known as 'PCA variance Estimator' is suggested by the author, based on the decreasing values of eigen vectors/eigen values [7,22]. This variance estimator works on the principle that detection of various segments of an ECG is done in the following sequence:

(1) First QRS complex is detected.
(2) T-wave detection is the second step.
(3) P-wave detection is based on the fact that first positive slope is found followed by negative slope.
(4) Lastly, BLW and noise components if available in the ECG signal are detected.

The performance of the PCA variance Estimator is evaluated on the ECG data sets by computing the percentages of:

(1) Sensitivity (SE)
(2) Specificity (SP)
(3) Correct classification (CC)

The calculations of sensitivity, specificity and classification will indicate the appearance of false positive and negative peaks, which can be evaluated using PCA. The results of the PCA variance estimator is validated by calculating the correct classification, sensitivity and specificity for leads AF and AR of various CSE based ECG data records before and after applying PCA for detection of true and false peaks. These validation parameters are defined as:

$$\text{Sensitivity (SE)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{Specificity (SP)} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{Correct classification, } i.e. \text{ Accuracy (AC)} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{TP} + \text{FN}}$$

where, TP is true positive, FN is false negative, TN is true negative and FP is false positive respectively of the R-peaks.

Tables 1(A) and (B) give the results of correct classification, sensitivity and specificity for two leads of an ECG database file for noise conditions, baseline wander and their combinations using PCA variance estimator.

4.2 ICA results

In this work, an approach of ICA to separate the 3-channel ECG waveforms is discussed and reimplemented as first discussed in [1]. Jade algorithm is applied to remove noise and artifacts from ECG recordings. Jade algorithm is a statistically based technique for better visualization of any ECG data. Before applying the algorithm to the ECG data, the data is to be centered and the whitening of the data is done by the algorithm itself, which is the main reasoning of using it in this analysis. The reconstructed ECGs after cleaning using ICA are

Table 1(A). Values of correct classification, sensitivity and specificity for one lead of an ECG database file.

Conditions of ECG data sets	Correct classification (%)	Sensitivity (%)	Specificity (%)
Free of noise and base-line wander	94.83	86.63	94.42
With noise only	91.34	81.33	91.92
With noise and base-line wander both	88.25	76.17	88.95

Table 1(B). Values of correct classification, sensitivity and specificity for other lead of the same ECG database file.

Conditions of ECG data sets	Correct classification (%)	Sensitivity (%)	Specificity (%)
Free of noise and base-line wander	93.81	85.77	94.47
With noise only	90.55	80.38	90.89
With noise and base-line wander both	87.49	75.917	87.91

compared with the source ECG signals. The noisy ECG recordings were cleaned using statistical measures kurtosis and variance of variance to find out the noise characteristics. In the present work, Jade algorithm for ICA is applied to obtain the independent components. It is proposed that the independent component whose |Kurt|⟨Kurt (Threshold) and Varvar⟩Varvar (Threshold) is taken as a noise or artifact component [1].

For the three special cases of the CSE database discussed, the R-Peak magnitudes of the ECG waveforms before and after applying ICA are shown in respective tables. From these tables, it is clear that in most of the cases, the percentage error in the reconstruction is very small.

Case studies: Three different case studies are discussed to show the capability of ICA in removing artifacts and noise as well a competent tool for ECG feature extraction in collaboration with other tools like PCA and wavelet transforms.

Case 1: 3-channel ECG with baseline drift in one channel and noise in other channel.
Case 2: 3-channel ECG with baseline drift in two channels.
Case 3: 3-channel ECG with high frequency noise in one of the channels.

The simulation and graphical results are shown in the respective figures. The results have been obtained as: Number of points $= 5000$, Sampling frequency $= 500\,Hz$.

Thresholds for Kurtosis $= 4.3$ and Varvar $= 0.4$ were obtained after checking on around 20 CSE database files after parameterization of ECG signals, which shows improvements in results as compared to those presented in [1]. This study is reimplementation of [1] as far as thresholds for kurtosis and variance of variance (Varvar) are concerned.

Case 1: CSE data file for leads L_1, L_2 and L_3 of an ECG are used in the simulation. This case deals with 3-channel ECG with baseline drift in one channel and noise in other channel.

It is apparent that channel-2, *i.e.* figure 8(b) is having baseline wander and channel-3, *i.e.* figure 8(c) is having noise. It is obvious from figure 8 and table 2 that ICA-2 is noise component since it is having |Kurt|⟨4.3, and ICA-3 is an artifact component that it is having Varvar⟩0.4, after parameterization of source ECG signals, hence these components are made zero [1].
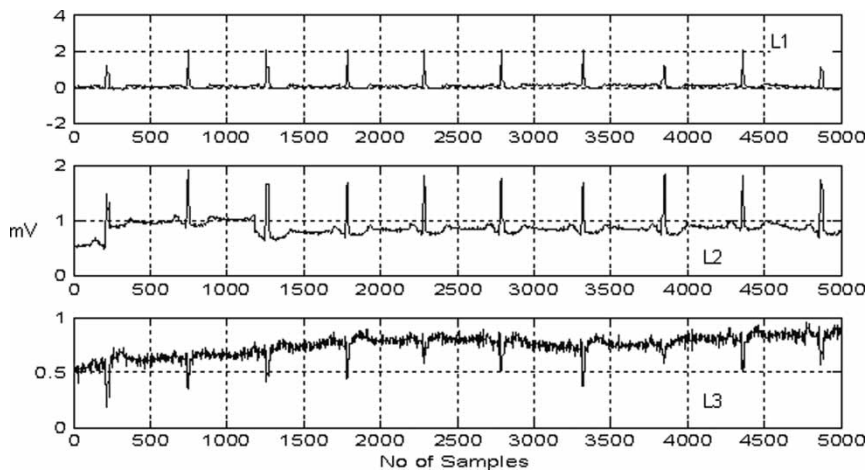

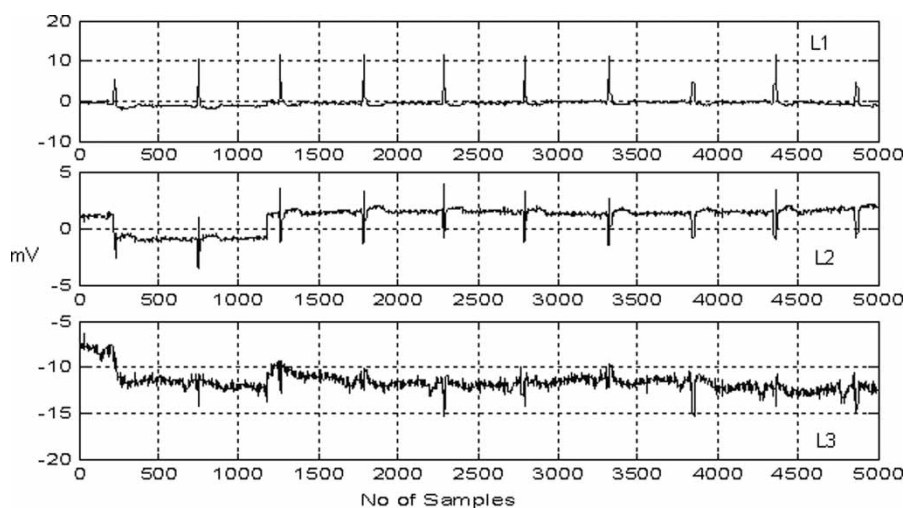Figure 6. 3-channel ECG for case-1.

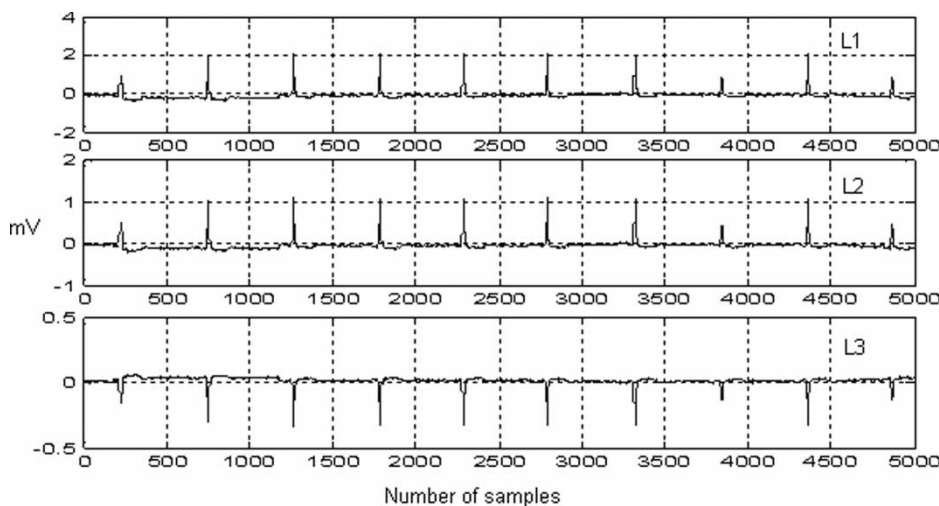Figure 7. Plot of extracted independent components.

Figure 8. Plot of reconstructed clean ECG signals.

Case 2: A CSE data file for leads L_1, L_2 and L_3 of an ECG are used in the simulation. ECG leads are having baseline drift in two out of the three channels.

Table 3 indicates that ICA-1 is noise component that is having $|Kurt| < 4.3$, and ICA-3 is an artifact component having Varvar > 0.4, hence these components are made zero.

Table 2(A). |Kurt| and Varvar for the three ICA components for case-1.

Index	ICA1	ICA2	ICA3
|Kurt|	39.1544	3.3999 (Noise)	8.3203
Varvar	0.0254	0.1683	0.6913 (Artifact)

Table 2(B).   R-Peak magnitudes before and after ICA processing of case-1.

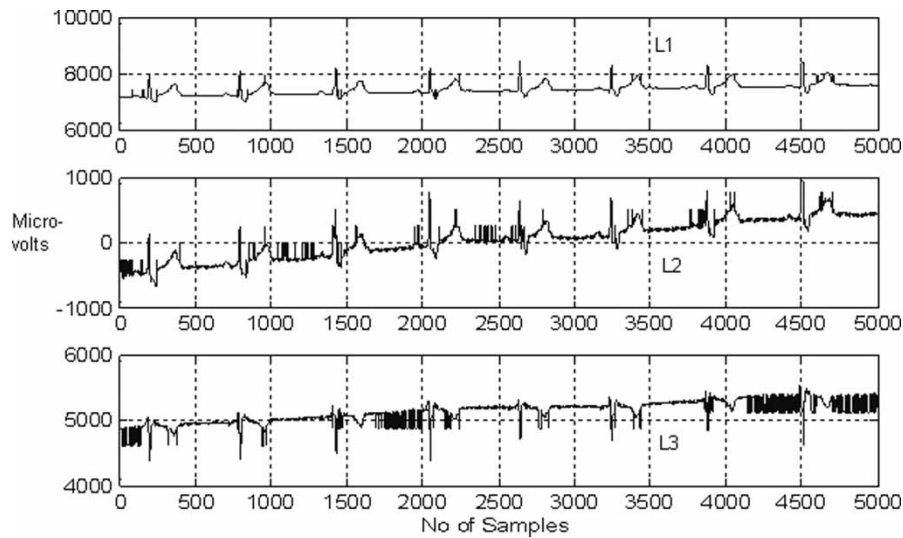| ECG lead	Original ECG in mV {Peak−Peak} (R-wave)	Reconstructed ECG in mV {Peak−Peak} (R-wave)	(%) Reconstruction error
L_1	1.224	1.220	0.326
L_2	0.999	0.989	1.001
L_3	0.335	0.330	1.492

Figure 9.   3-channel ECG for case-2.

Case 3: A CSE data file for leads L_1, L_2 and L_3 of an ECG are used in the simulation. This case deals with 3-channel ECG with high frequency noise in one of the channels L_1, L_2 and L_3.

Independent components extracted after ICA processing are shown in figure 13. It is apparent from table 4, ICA-3 is a high frequency noise component that is having |Kurt| < 4.3
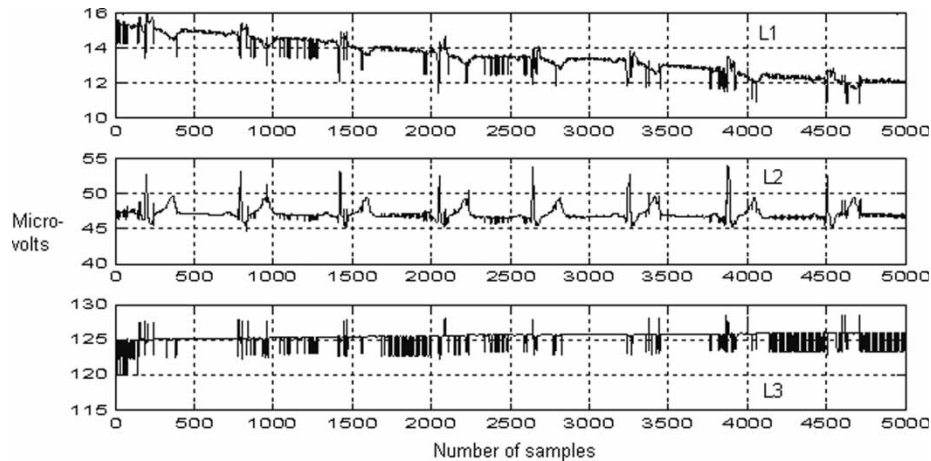

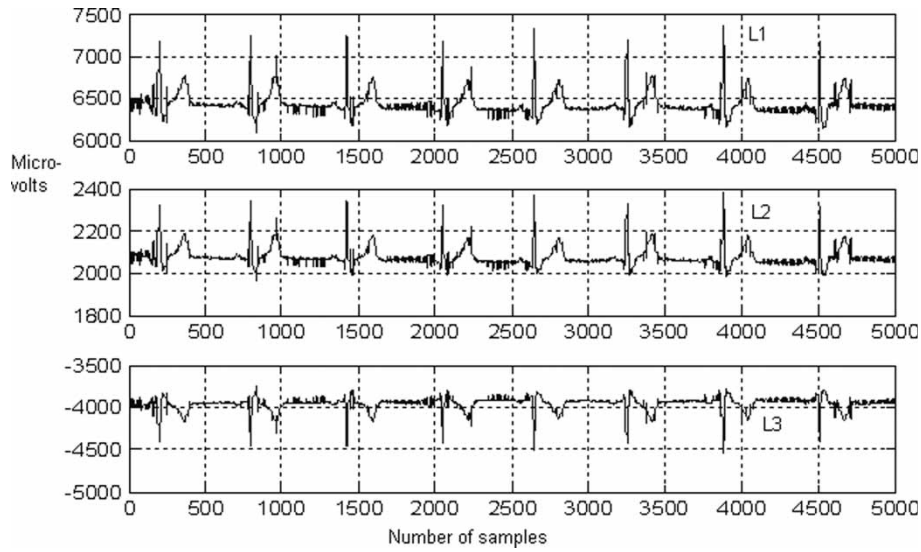Figure 10.   Plot of extracted independent components.

Figure 11. Plot of reconstructed clean ECG signals.

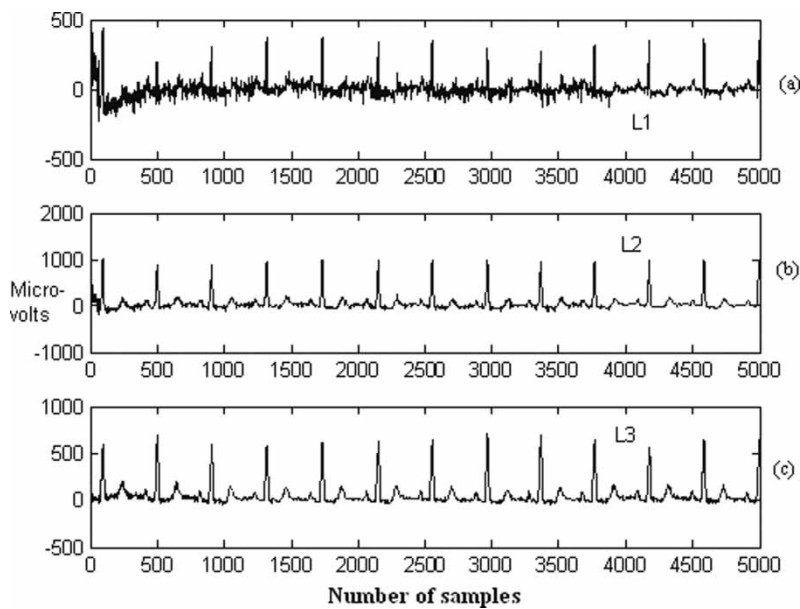Figure 12. ECGs with high frequency noise in one channel for case-3.

Table 3(A). |Kurt| and Varvar for the three ICA components of case-2.

Index	ICA1	ICA2	ICA3
|Kurt|	2.1111(Noise)	11.9471	6.5778
Varvar	0.0017	0.0812	0.5196 (Artifact)

Table 3(B).   R-Peak magnitudes before and after ICA processing of case-2.

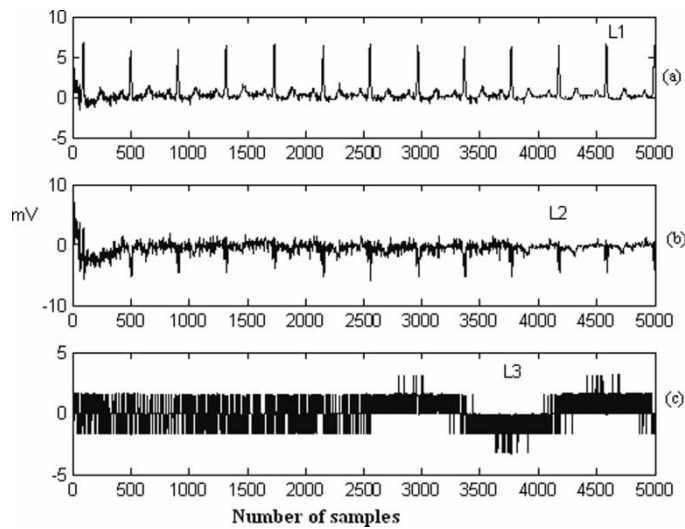| ECG lead	Original ECG in mV {Peak−Peak} (R-wave)	Reconstructed ECG in mV {Peak−Peak} (R-wave)	(%) Reconstruction error
L_1	0.858	0.842	1.864
L_2	0.571	0.565	1.050
L_3	0.607	0.600	1.153

Figure 13.   Extracted independent components.

Table 4(A). |Kurt| and Varvar for each of the three ICA components for case-3.

Index	ICA1	ICA2	ICA3
|Kurt|	19.5467	10.3303	3.1682(Noise)
Varvar	0.1085	0.9487(Artifact)	0.0246

Table 4(B).   R-Peak magnitudes before and after ICA processing of case-3.

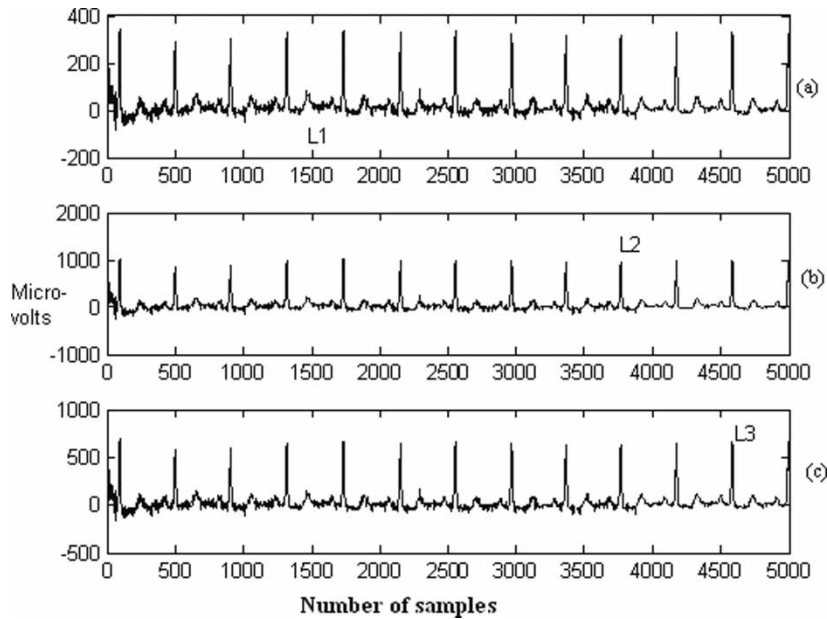| ECG Lead	Original ECG in mV {Peak−Peak} (R-wave)	Reconstructed ECG in mV {Peak−Peak} (R-wave)	(%) Reconstruction error
L_1	0.389	0.378	2.827
L_2	1.212	1.212	0
L_3	0.612	0.610	0.326

Figure 14. Plot of reconstructed clean ECG signals for case-3.

and ICA-2 is an artifact component that is having Varvar > 0.4. Clean ECG signals are reconstructed after making ICA-2 and ICA-3 equal to zero as shown in figure 14.

5. Validation of ICA simulations

The results of the proposed method is validated by calculating the sensitivity to the various CSE based ECG data records before and after applying ICA for detection of true and false peaks using quadratic spline wavelet along with the PCA variance estimator.

Table 5. Results of QRS detection and R-peaks before applying quadratic spline wavelet along with PCA variance estimator to various ECG data records.

Lead name	Number of ECG records	Total number of samples	Actual number of R-peaks	True positive (TP)	False positive (FP)	False negative (FN)
AvF	50	250,000	502	422	17	10
AvL	50	250,000	502	418	19	5
AvR	50	250,000	502	412	12	17
L1	50	250,000	502	417	18	10
L2	50	250,000	502	406	17	10
L3	50	250,000	502	409	5	9
V1	50	250,000	502	405	12	12
V2	50	250,000	502	408	4	14
V3	39	250,000	502	403	7	13
V4	50	250,000	502	405	10	9
V5	50	250,000	502	400	6	8
V6	50	250,000	502	403	8	10
	Total			4908	135	127

Table 6. Results of QRS detection and R-peaks after applying ICA to ECG data using quadratic spline wavelet and PCA variance estimator to various ECG data records.

Lead name	Number of ECG records	Total number of samples	Actual number of R-peaks	True positive (TP)	False positive (FP)	False negative (FN)
AvF	25	125,000	227	214	6	4
AvL	25	125,000	227	207	9	5
AvR	25	125,000	227	212	7	6
L1	25	125,000	227	212	6	4
L2	25	125,000	227	217	2	3
L3	25	125,000	227	216	3	3
V1	25	125,000	227	212	4	4
V2	25	125,000	227	215	3	3
V3	25	125,000	227	214	5	4
V4	25	125,000	227	212	4	6
V5	25	125,000	227	209	4	4
V6	25	125,000	227	213	4	4
	Total			2553	57	50

5.1 Estimation of sensitivity before applying ICA

$$\text{Sensitivity (SE)} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{4908}{4908 + 127} = 97.47\%$$

5.2 Estimation of sensitivity after applying ICA

$$\text{Sensitivity (SE)} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{2553}{2553 + 50} = 98.07\%$$

Jade algorithm and quadratic spline wavelet along with PCA variance estimator gave a sensitivity of 97.47% before applying ICA and 98.07% after ICA processing to several CSE based ECG data files.

6. Discussions

PCA gives optimal compression performance, and exceeds wavelet transform performance, though it requires marginally more processing overheads. The performance is slightly poorer than neural-network compression but the processing overhead is significantly lower. For PCA, the measure which is used to discover the axes is 'variance' and leads to a set of orthogonal axes. ICA decomposition gives constant correlation coefficient whereas PCA decomposition exhibits varying correlation coefficient. In ICA, since some components are more emphasized or some components are either hidden or reduced, it can be used as a good potential framework for automated heart monitoring. Abnormal signals are more easily detected in the independent components compared to the original measured ECG signals, reducing the need for trained doctors' opinions. Additional benefits indicate the PCA approach is more suitable as the basis for a complete ECG analysis, classification and diagnosis. It is expected by the authors after this comparative analysis that ECG researchers

will use the proposed methods and algorithms as a convenient platform to integrate new research into ECG triggering, compression, clustering, analysis, presentation, classification, and interpretation techniques.

In this paper, the author has reimplemented a method for cleaning of ECG signals by ICA approach, a special case of blind source separation technique. Starting from the multi-channel nature of the ECG acquisitions, ECG signals as a convolutive mixture of independent components are modeled in this study. The ECG, noise and artifacts components are separated from multi-channel acquisitions by exploiting the well known Jade algorithm available for instantaneous ICA model. The original ECG data and the results achieved by this method are also used to find the reconstruction errors of R-peaks.

The important assumptions made in applying the ICA are: components are independent and have non-Gaussian distributions. The ambiguities of ICA on the other hand, such as energies of the extracted ICs at times which cannot be determined, and their order remaining undetermined, are taken care in this analysis. ICs of parameterized ECG signals display the desired aspects of the question at hand much more clearly than the ICs of the original ECG signals, while possibly reducing the amount of ECG data to a fraction of the original ECG signal preserving morphology. Therefore, author feels to conclude that employing the time structure information in ICA calculations can potentially improve artifacts and noise removal and enhance the overall ECG classification and feature extraction.

7. Conclusion

The proposed method provided a cleaning procedure for the noisy and corrupted ECG signals. Jade algorithm to the 3-channel ECG data is applied and it is observed that the approach is quite effective in identifying and removing noise/artifacts. The present study deals with the capability of ICA in general to identify and remove baseline wander, abrupt baseline changes, high frequency noise, dc drifts, sudden spikes, etc. from the ECG. ICA processing of four cases has been discussed graphically as well as computationally. The results demonstrate that there is significant improvement in signal quality, *i.e.* signal to noise ratio is improved. In all the cases discussed, it is found that the independent components whose $|Kurt|\langle$Kurt (Threshold) and Varvar$\rangle$Varvar (Threshold) are taken respectively as noise and artifact components. From the results presented in tabular form, it is clear that in most of the cases, the percentage error in the ECG reconstruction is very small maintaining diagnostic morphology. However, standard ICA fails to separate the mixtures if more than one of the sources has a Gaussian amplitude distribution. The capability of the proposed ICA algorithm is validated by finding the true, false and actual peaks for about $25-50$ CSE database ECG files. Jade algorithm with PCA variance estimator gave a sensitivity of 97.47% before applying ICA and 98.07% after ICA processing to several CSE based ECG data files.

References

[1] He, Taigang, Clifford, Gari and Tarassenko, Lionel, 2005, Application of independent component analysis in removing artefacts from the electrocardiogram. *Neural Computing and Applications* (London: Springer-Verlag), pp. 1–19.

[2] Chawla, M.P.S., Verma, H.K. and Kumar, V., 2007, A new statistical PCA-ICA algorithm for location of R-peaks in ECG, *International Journal of Cardiology*, (in press, available online 25th July 2007).

[3] Chawla, M.P.S., Verma, H.K. and Kumar, V., 2007, Artifacts and noise removal in electrocardiograms using independent component analysis, *International Journal of Cardiology*, (in press, available online 8th August 2007).

[4] Chawla, M.P.S., 2007, Parameterization and correction of electrocardiogram signals using independent component analysis, *Communicated in International Journal of Mechanics in Medicine and Biology (JMMB)*, WSPC (in press, to appear December 2007).

[5] Urrestarazu, Elena, Iriarte, Jorge, Alegre, Manuel, Valencia, Miguel, Viteri, Cesar and Artieda, Julio, September 2004, Independent component analysis for removing artifacts in ictal recordings, *Epilepsia*, **45**, 1071–1078.

[6] Jorge, Iriarte, Elena, Urrestarazu, Miguel, Valencia, Manuel, Alegre, Armando, Malanda, Cesar, Viteri and Julio, Artieda, July/August 2003, Independent component analysis as a tool to eliminate artifacts in EEG: a quantitative study, *Journal of Clinical Neurophysiology*, **20**(4), 249–257.

[7] Chawla, M.P.S., Verma, H.K. and Kumar, V., 2006, ECG modeling and QRS detection using principal component analysis. Paper no. 04, in *IET Proceedings, International Other, MEDSIP-06*, Glasgow, UK, July.

[8] Draper, B., Baek, K., Bartlett, M.S. and Beveridge, J.R., 2003, Recognizing faces with PCA and ICA, *Computer Vision and Image Understanding*, **91**(1–2), 115–137.

[9] Sornmo, L. and Laguna, P., 2005, *Bioelectrical Signal Processing in Cardiac and Neurological Applications* (London: Elsevier Academic Press).

[10] Chawla, M.P.S., Verma, H.K. and Kumar, V., 2006, Data reduction and removal of base-line wander using principal component analysis, Communicated in DSP, Elsevier in May 2006 (Under review).

[11] Chawla, M.P.S., Verma, H.K. and Kumar, V., 2006, Independent component analysis: a novel technique for removal of artifacts and base-line wander in *ECG*, *National Other CISCON-2006*, MIT, Manipal, November, pp. 14–18.

[12] Chawla, M.P.S., Verma, H.K. and Kumar, V., 2006, Modeling and feature extraction of ECG using independent component analysis, Paper no.040, *IEE (IET), Proceedings, International Other* APSCOM-2006, Hong Kong, October 30–November 2.

[13] Vigneron, V., Ionescu, A.Paraschiv, Azancot, A., Sibony, O. and Jutten, C., 2003, Fetal electrocardiogram extraction based on non-stationary ICA and wavelet denoising, *Proceedings, ISSPA-2003*.

[14] Gao, P., Chang, E.C. and Wyse, L., 2003, Blind separation of fetal ECG from single mixture using SVD and ICA, *Proceedings of ICICS-PCM* 2003, 15–18 December 2003, Singapore.

[15] Marossero, D.E., Erdogmus, D., Euliano, N.R., Principe, J.C. and Hild II, K.E., 2003, Independent components analysis for fetal electrocardiogram extraction: a case for the data efficient Mermaid algorithm, *Proceedings, IEEE International Workshop on Neural Networks for Signal Processing*, Toulouse, France, September 2003.

[16] Vrins, F., Jutten, C., Verleysen, M., *et al.*, 2004, Sensors array and electrode selection for non-invasive fetal electrocardiogram extraction by independent component analysis, *Proceedings*, ICA-2004, LNCS **3195**, pp. 1017–1024.

[17] Azzerboni, B., Finocchio, G., Mammone, N. and Morabito, F.C., September 2002, A new approach to detection of muscle activation by independent component analysis and wavelet transform. *Lecture Notes in Computer Science* (London: Springer-Verlag), pp. 109–116.

[18] Azzerboni, B., Carpentieri, M., La Foresta, F. and Morabito, F.C., 2004, Neural-ICA and wavelet transform for artifacts removal in surface EMG, *Proceedings of International Joint Other on Neural Networks* (IJCNN 2004), pp. 3223–3228.

[19] James, C.J. and Hesse, C.W., 2005, Independent component analysis for biomedical signals, *Physiological Measurement*, **26**, R15–R39.

[20] Azzerboni, B., La Foresta, F., Mammone, N. and Morabito, F.C., April 2005, A new approach based on wavelet-ICA algorithms for fetal electrocardiogram extraction, ESANN'2005, *Proceedings of European Symposium on Artificial Neural Networks* Bruges (Belgium), pp. 193–198.

[21] Jarno, M.A.T., Viik, J.J. and Hyttinen, J.A.K., 2006, Independent component analysis of parameterized ECG signals, *Proceedings of 28th IEEE, EMBS, Annual International Other*, New York City, USA, August 30–September 3, pp. 5704–5707.

[22] Chawla, M.P.S., Verma, H.K. and Kumar, V., 2006, PCA-ICA Method for Detection of QRS Complexes and Location of R-Peaks in Electrocardiograms, Communicated in DSP, Elsevier in May, 2006 (Under review).

[23] Bai, J., Zhang, Y., Shen, D., Wen, L., Ding, C., Cui, Z., Tian, F., Yu, B., Dai, B. and Zhang, J., 1999, A portable ECG and blood pressure telemonitoring system, *IEEE Engineering in Medicine and Biology Magazine*, **18**, 63–70.

[24] Liang, H., 2005, Extraction of gastric slow waves from electrogastrograms: combining independent component analysis and adaptive signal enhancement, *Medical and Biological Engineering and Computing*, **43**, 245–251.

[25] Tikkanen, P., 1999, Characterization and application of analysis methods for ECG and time interval variability data, PhD dissertation, University of Oulu.

[26] AbedMeraim, K., Amin, M.G. and Zoubir, A.M., May 2001, Joint anti-diagonalization for blind source separation, *Proceedings of ICASSP-01*.

[27] Paul, J.S., Reddy, M.R. and Kumar, V.J., 2000, A transform domain SVD filter for suppression of muscle noise artefacts in exercise ECGs, *IEEE Transactions on Biomedical Engineering*, **3**, 654–663.

[28] Irimia, A. and Bradshaw, L.A., 2005, Artifact reduction in magneto-gastrography using fast independent component analysis, *Physiological Measurement*, **26**, 1059–1073.

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

[30] Lee, T.W., Girolami, M. and Sejnowski, T.J., 1999, Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources, *Neural Computation*, **11**(2), 606–633.

[31] Chawla, M.P.S., 2007, A comparative analysis of principal component and independent component techniques for electrocardiograms, *International Journal of Neural Computing and Applications (NCA)*, (in press).