EXTRACTION OF VARIOUS FEATURES OF ECG SIGNAL

Nishant Saxena[1], Kshitij Shinghal[2]
1Assistant Professor, Deptt. of EE&I Engg., MIT, Moradabad, U.P., India
2Associate Professor, Deptt. of E&C Engg., MIT, Moradabad, U.P., India

ABSTRACT

Principal component analysis (PCA) is one of the most valuable results oriented techniques of applied linear algebra. PCA is used abundantly in all forms of analysis from neuroscience to computer graphics because it is a simple, non-parametric method of extracting relevant information from confusing data sets. Extracting or decoding this information or feature from ECG signal has been found very helpful in explaining and identifying various pathological conditions. The feature extraction procedure can be accomplished straightforward by analysing the ECG visually on paper or screen. In addition, manual feature extraction is always prone to error. Therefore, ECG signal processing has become an indispensable and effective tool for extracting clinically significant information from ECG signals, for reducing the subjectivity of manual ECG analysis and for developing advanced aid to the physician in making well-founded decisions. ECG analysis systems are usually designed to process ECG signals measured under particular conditions, like resting ECG interpretation, stress test analysis, ambulatory ECG monitoring and intensive care monitoring. Noise reduction is closely related to data compression as reconstruction of the original signal usually involves a set of eigenvectors whose noise level is low, and thus the reconstructed signal becomes low noise, such reduction is, however, mostly effective for noise with muscular origin. The purpose of the thesis is to provide an overview of PCA in ECG signal compression.

KEYWORDS: PCA, ECG, feature extraction, data compression.

I. INTRODUCTION

Principal component analysis (PCA) is a statistical technique whose purpose is to condense the information of a large set of correlated variables into a few variables ("principal components"), while not throwing overboard the variability present in the data set [1]. The principal components are derived as a linear combination of the variables of the data set, with weights chosen so that the principal components become mutually uncorrelated. Each component contains new information about the data set, and is ordered so that the first few components account for most of the variability. In signal processing applications, PCA is performed on a set of time samples rather than on a data set of variables. When the signal is recurrent in nature, like the ECG signal, the analysis is often based on samples extracted from the same segment location of different periods of the signal. Signal processing is today found in virtually any system for ECG analysis, and has clearly demonstrated its importance for achieving improved diagnosis of a wide variety of cardiac pathologies. Signal processing is employed to deal with diverse issues in ECG analysis such as data compression, beat detection and classification, noise reduction, signal separation, and feature extraction. A recent application of PCA in ECG signal processing is robust feature extraction of various waveform properties for the purpose of tracking temporal changes due to myocardial ischemia. Historically, such tracking has been based on local measurements derived from the ST-T segment; however, such measurements are unreliable when the analyzed signal is noisy. With correlation as the fundamental signal processing operation, it has become clear that the use of principal components over a more robust and global approach to the characterization of the ST-T segment. Signal separation during atrial fibrillation is another recent application of PCA, the specific challenge being to extract the atrial activity so that the characteristics of this common arrhythmia can be studied without interference from ventricular activity.

II. ELECTROCARDIOGRAM

The electrical functionality of different organs in the human body is expressed by bioelectrical signals. The Electrocardiogram, also called ECG signal, is one important signal among all bioelectrical signals. The ECG reflects the performance and the properties of the human heart and conveys very important hidden information in its structure. This information has to be extracted and analyzed before any useful and meaningful interpretations can be started. Extracting or decoding this information or feature from ECG signal has been found very helpful in explaining and identifying various pathological conditions. The feature extraction procedure can be accomplished straightforward by analysing the ECG visually on paper or screen. However, the complexity and the duration of ECG signals are often quite considerable making the manual analysis a very time-consuming and limited solution. In addition, manual feature extraction is always prone to error. Therefore, ECG signal processing has become an indispensable and effective tool for extracting clinically significant information from ECG signals, for reducing the subjectivity of manual ECG analysis and for developing advanced aid to the physician in making well-founded decisions. Over the past few years automatic analysis of electrocardiograms (ECG) has gained more and more significance in the field of clinical ECG diagnosis.

The human heart is located in the chest between the lungs, behind the sternum and above the diaphragm. It weighs between 200 to 425 grams and is a little larger than the size of a fist [2, 3, 4]. The basis end and the apex end of the heart lie on its main axis which is oriented from the back-top-right to the front-bottom-left of the torso. Every day it beats in average 100000 times pumping about 7600 liters of blood to the body [5]. Like a sack, a double-layered membrane called the pericardium surrounds the heart. Its outer layer covers the roots of the heart's major blood vessels and is attached by ligaments to the spinal column, diaphragm, and other parts of your body. The inner layer of the pericardium is connected to the heart muscle. The layers are separated by a coating of fluid, letting the heart move as it beats and keeping it attached to the body. The normal periodic contractions and relaxations of the heart allow the human cells receiving the necessary amount of oxygen and nutrients and carrying away their end product of the metabolism.
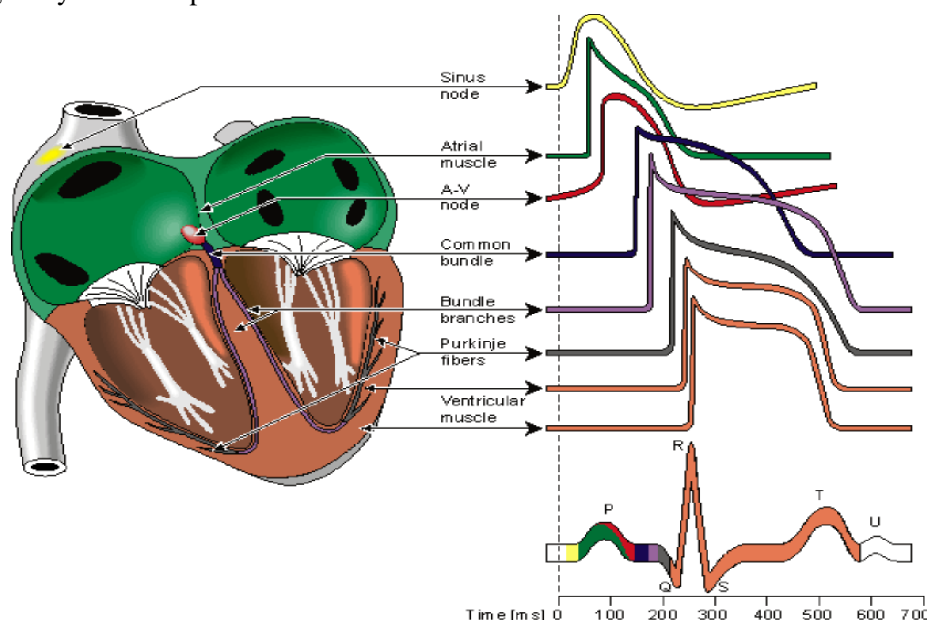

Figure 1. The Genesis of Electrocardiogram

The walls of the heart are composed of cardiac muscle, Myocardium. It is similar to skeletal muscle, because it has striations. The cardiac muscle consists of four chambers: the right and left atria and ventricles. The anterior aspect of the heart is the right ventricle, whereas the posterior aspect is the left atrium giving the heart its orientation. The endocardium is defined as the thin serous membrane that lines the interior of the heart, whereas the epicardium touches the inner layer of the pericardium that is

in actual contact with the surface of the heart. The left ventricle pumps blood to the systemic circulation, where pressure is considerably higher than for the pulmonary circulation, which arises from right ventricular outflow. The left ventricular free wall and the septum is much thicker than the right ventricular wall [6]. The tricuspid valve lays between the right atrium and ventricle, and the mitral valve is between the left atrium and ventricle. Between the right ventricle and the pulmonary artery lies the pulmonary valve, while the aortic valve is in the outflow tract of the left ventricle controlling blood flow to the aorta. Carried in the inferior and superior vena cava, the blood returns from the systemic circulation to the right atrium [7, 8, 9]. First, it has to go through the right ventricle, then it is ejected through the pulmonary valve and the pulmonary artery to the lungs. Oxygen-rich blood returns from the lungs to the left atrium and to the left ventricle. Finally blood is pumped through the aortic valve to the aorta and the systemic circulation. The left and right coronary arteries branch off the aorta. They are divided afterward into numerous smaller arteries supplying oxygen and nourishments to all heart muscles.

III.   THE NORMAL ECG WAVES, TIME INTERVALS, AND ITS NORMAL VARIANTS

The normal ECG signal represents a normal cardiac cycle. Figure 2.8 illustrates the normal ECG of one cardiac cycle along with its components. The normal variances and characteristics of its waves, durations and time intervals are described as follows:

3.1 The P Wave

Depolarization in the atria is registered as the P wave in the ECG. Duration of the P wave should not exceed 0.10 sec in limb leads, or 0.12 sec in chest leads. Its amplitude averages 0.1-0.3 mV. P wave is normally positive in limb leads except in a VR lead, where it is always negative. It is most pronounced in lead II. P wave is always positive in left precordial leads, often biphasic over the right chest wall. The autonomic nervous system activity plays a considerable role in the variation of P wave morphology. The amplitude of P wave may increase remarkably, above all in leads II, III and a VF, when the sympathetic tone is increased. On the other hand, when there is an increased parasympathetic tone, P wave becomes flat in leads II, III, and a VF. The spectral characteristics of a normal P wave is usually considered to be low-frequency, below 10-15 Hz [33].

3.2 The QRS Complex

The ventricular complex represents the initial ventricular depolarization. It usually comprises a Q, R and an S wave. Every positive wave is called R. The first negative wave preceding an R is always called Q and the first negative wave following R is always called S. A possible second or third R wave is called R' or R'', figure 2.9 It is also preferable to speak of a split ventricular complex, when several waves are presented. Notation of individual waves of the ventricular complex is different according to amplitude by using small or large letters.

3.3 The PR or PQ Interval

It is measured from the start of the P wave to the start of ventricular complex. It should not be shorter than 0.12 sec, nor longer than 0.20 sec. Prolonged AV conduction time at the rest is which becomes normal on exercise not necessarily a sign of abnormality.

3.4 The T Wave

It expresses repolarization of the ventricles. Its amplitude must always be taken in relation to the R wave. T wave is always positive in lead I and II, and it is always negative in aVR lead.

3.5 The U Wave

After T wave, an ECG can sometimes show a U Wave. It is of the same deflection as T Wave and similar to shape to P Wave. The U Wave is thought to represent late repolarization of the Purkinje fibers in the ventricles [63].

3.6 The PP Interval and the RR Interval

PP interval is defined as the duration of atrial cycle. It is useful as an indicator of atrial rate. Whereas, RR interval is defined as the distance in msec between two successive R waves. It is an indicator of

ventricular rate representing the length of a ventricular cardiac cycle. Moreover it is very important to characterize different arrhythmias and to study the heart rate variability.

3.7 The ST Segment

ST segment represents the period from the end of ventricular depolarization to the be-ginning of ventricular repolarization. The ST segment lies between the end of the QRS complex and the initial deflection of the T-wave and is normally isoelectric. It is clinically important if it is elevated or depressed as it can be a sign of ischemia and hyperkalemia [70]. In order to interpret ST segment correctly, the J point should be localized precisely. The J point, as definition, is the time instant in the ECG when the QRS complex curves into the ST segment.

IV.   PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA is a linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. PCA can be used for dimensionality reduction in a dataset while retaining those characteristics of the dataset that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones. Such low-order components often contain the 'most important' aspects of the data, but this is not necessarily the case, depending on the application. PCA is also called the (discrete) Karhunen-Lo`eve transform (or KLT, named after Kari Karhunen and Michel Lo`eve) or the Hotelling transform (in honor of Harold Hotelling). PCA has the distinction of being the optimal linear transformation for keeping the subspace that has largest variance. This advantage, however, comes at the price of greater computational requirement if compared, for example, to the discrete cosine transform. Unlike other linear transforms, the PCA does not have a fixed set of basis vectors. Its basis vectors depend on the data set.

4.1 Implementation and Result Analysis

First the signal is taken and divided into the blocks. Then the mean is taken of the particular beat. Then that mean is subtracted from the original signal. Thus we got the principal components. if the covariance is taken of the resultant we will get the eigenvectors and eigenvalues with the help of which we can reconstruct the original signal.

The procedure is performed with the help of MATLAB Software and results are being displayed. The principal component analysis is performed for each of the cases.

The key to obtaining data compression is signal representation, which concerns the representation of a given class of signals in an efficient manner. If a discrete signal comprises of n samples, then it can be thought of as being appoint in an n dimensional space. Each sampled value is then a component of the data n vector x, that represents a discrete signal in this space.

For an efficient representation of X, we secures an orthogonal transform of X, which results in Y=TX where Y denotes the transform vector and T represents the transformation matrix.

For data compression we will select a subset of m components of Y, where m is substantially less then n. The balance of (n-m) components will be discarded without introducing any serious error when the signal is reconstructed using the m saved components of the vector Y. To Whom It May Concern quantify this error of approximation an error criterion is needed and that is mean square error.

This chapter describes in detail the analysis experiments and their characteristics – i.e. the analysis environment setup for the experiment ECG signal analysis. The ECG signal has been picked from various sources and then analyzed. Basically all we are doing is to compress the ECG signal and then its reconstruction. The method used for this purpose is principal component analysis.

4.2 Simulation Experiment (Normal ECG)

In normal ECG the subject (human body) is in normal condition and measurement has been performed with the help of Lead-I. Figure shows the experimental result.
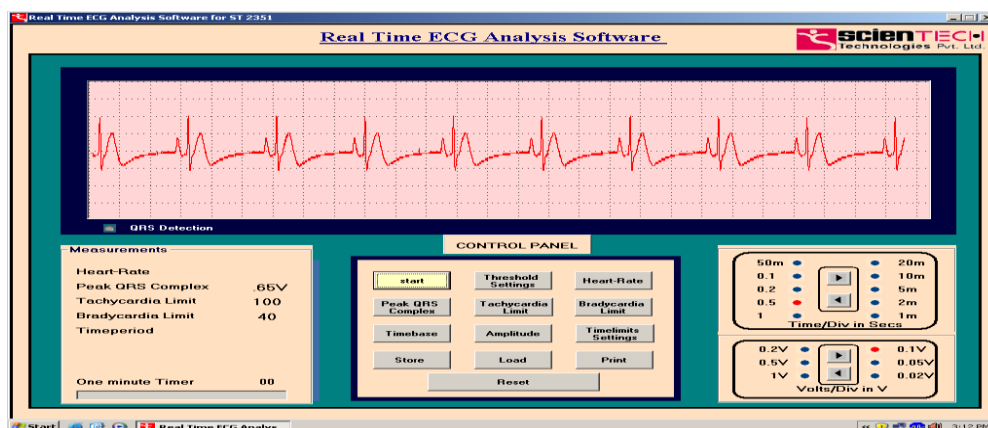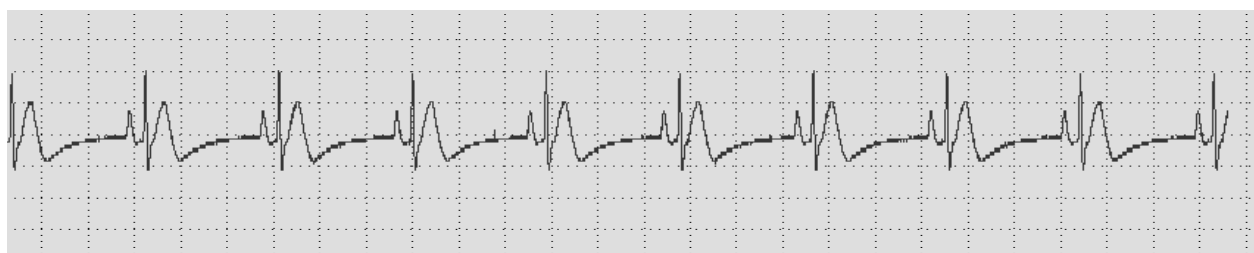
Figure 2 Normal ECG measurement

Figure 3. Extracted ECG waveform

As shown in figure 3 now we will take one heartbeat and will do the principal component analysis of that segment. So we will get the eigenvectors and Eigen values and the components with the highest eigen values will be selected.

The heart beat or the one cycle that has been selected is shown in figure 4
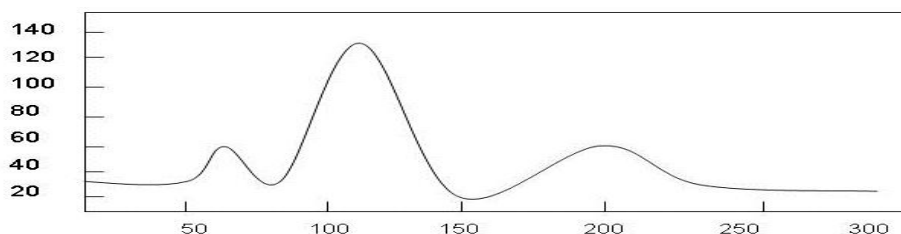

Figure 4. ECG waveform

Transform-based compression requires that the ECG first be partitioned into a series of successive blocks, where each block is subjected to data compression. The signal may be partitioned so that each block contains one beat. Each block is positioned around the QRS complex, starting at a fixed distance before the QRS, including the P wave and extending beyond the end of the T wave to the beginning of the next beat. Since the heart rate varies, the distance by which the block extends after the QRS complex is adapted to the prevailing heart rate. Hence, the resulting blocks vary in length, introducing a potential problem in transform-based compression where a fixed block length is assumed. This problem may be solved by padding too short blocks with a suitable sample value, whereas too long blocks can be truncated to the desired length. The use of variable block lengths has been studied in detail in [33, 34]; the results show that variable block lengths produce better compression performance than fixed blocks. It should be noted that partitioning of the ECG is bound to fail when certain chaotic arrhythmias are encountered such as ventricular fibrillation during which no QRS complexes are present.

The corresponding Eigen values of this waveform analysis with the MATLAB software has been shown here;
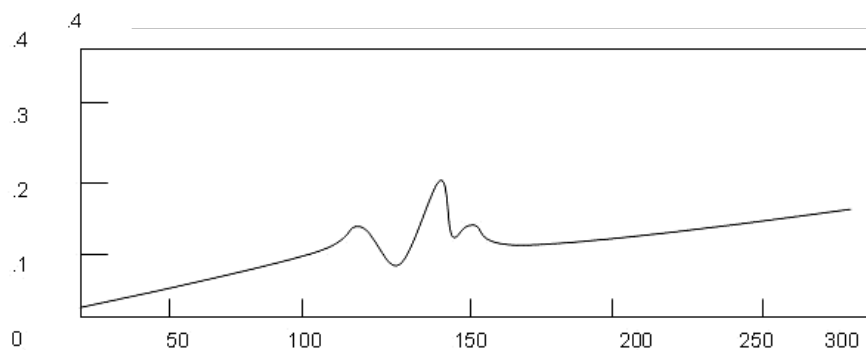
Table 1: Eigen values for the waveform

S.No.	Eigenvalues
1.	7.7696

2.	2.0530
3.	1.0886
4.	0.2225
5.	0.1388
6.	0.0407
7.	0.0213
8.	0.0177
9.	0.0099
10.	0.0098
11.	0.0082
12.	0.0072
13.	0.0057
14.	0.0049
15.	0.0037
16.	0.0033

Only three eigenvectors that   corresponds to the three dominant
eigenvalue are computed and are shown in figure 5 and 6.
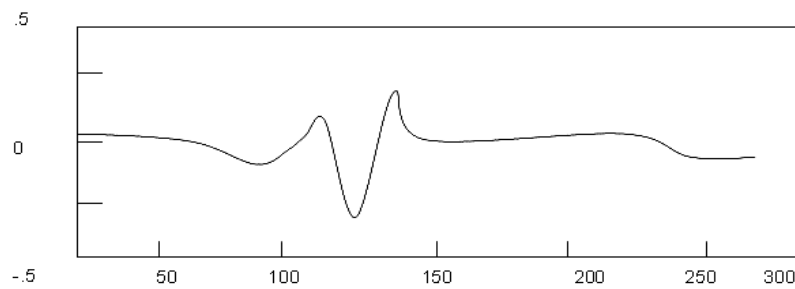
Figure 5: Eigenvector -1

Figure 6: Eigenvector-2

The reconstructed waveform corresponding to first cycle and given by the equation
$$X_m = y_1 \, \Phi_1 + y_2 \, \Phi_2$$
As shown in figure 5 and Figure 6 only two ECG Eigenvectors Φ_1, Φ_2 and two features y_1, y_2 we are able to reconstruct the original cycle of the waveform with fairly good accuracy.

Table 2: Comparison Table of original and reconstructed waveform

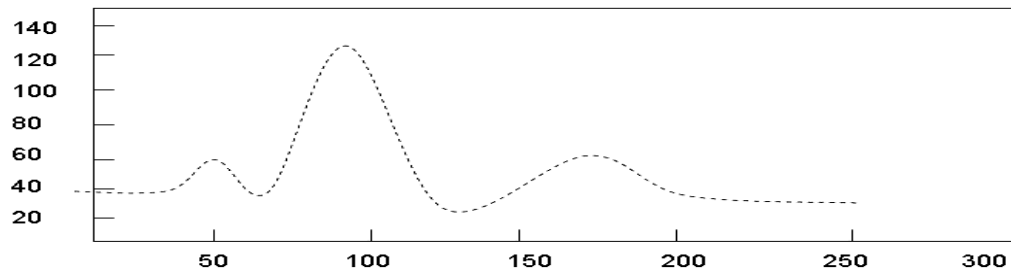| S.No.	PARAMETERS	ORIGINAL WAVE	RECONSTRUCTED WAVE	% Error
1.	P-Wave	0.25mv	0.23mv	8%
2.	QRS Complex	1.60mv	1.55mv	3.12%
3.	T-Wave	0.40mv	0.35mv	12.5%
4.	P-R Interval	0.20s	0.19s	5%
5.	Q-T Interval	0.40s	0.35s	5%

Figure 7: Reconstructed ECG waveform

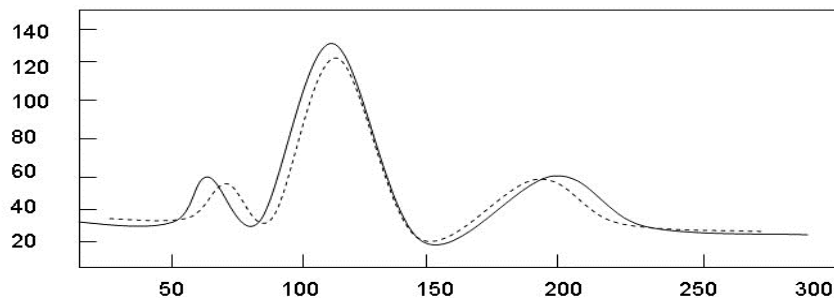As shown in above figure a reconstructed waveform has been found which shows the actual parameters.

Figure 8: Reconstructed and original ECG waveform

As shown in above figure:8, a comparison has been made been original and reconstructed waveform and these results are being found.

The comparison has been made and error has been calculated which is under the tolerance limit. It means to say that diagnosis can be made with the help of reconstructed ECG signal. The error has been shown with the help of graph drawn (Fig:9) below. On the X-axis the ECG parameters have been shown and on the Y-axis % Error has been shown.
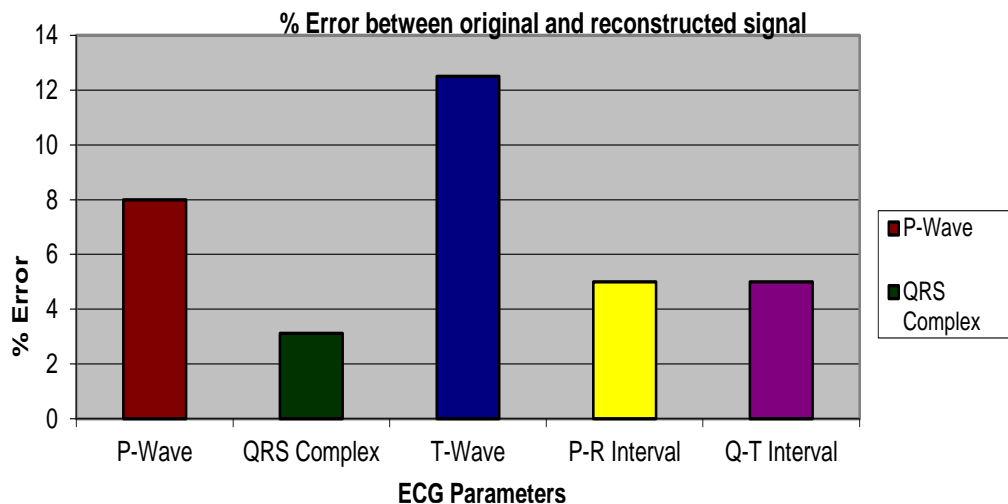

Fig 9: Error between Original & Reconstructed ECG Waveform

V.　CONCLUSION

The feature extraction of ECG signal is very helpful in the diagnosis of the signal. Although memory is not a very big problem nowadays but in the transmission of signal it is very useful as the distortion in the signal will be very less in the case of compressed signal. Several PCA-based strategies are available which exploit the fact that the ECG signal exhibits intrabeat and interlead redundancy. Although the underlying principle is the same in all ECG applications, the results are obtained and

interpreted in quite diverse ways. In some applications, the goal is to find a more compact representation of the signal, while in others it is to search for specific patterns or to extract a certain physiologic activity. In other applications, PCA may serve as a powerful, intermediate step when addressing problems related to noise reduction and beat classification. To date, PCA has been used to solve signal processing issues, most notably ECG data compression, as well as clinically oriented issues related to the characterization and diagnosis of myocardial ischemia, ventricular repolarization.

REFERENCES

[1]. Charles D. Deakin"From agonal to output: An ECG history of a successful pre-hospital thoracotomy" 0300-9572/$ — see front matter © 2007 Elsevier Ireland Ltd.

[2]. Dean Cvetkovic "Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: A pilot study" Digital Signal Process. (2007).

[3]. Barry R. Greene "Combination of EEG and ECG for improved automatic neonatal seizure detection" Clinical Neurophysiology 118 (2007) 1348–1359

[4]. Asiye Kanbaya, "Electrocardiography and Wells scoring in predicting the anatomic severity of pulmonary embolism"Respiratory Medicine (2007) 101, 1171–1176

[5]. Elif Derya Übeyli "Usage of eigenvector methods in implementation of automated diagnostic systems for ECG beats" Digital Signal Process. (2007),doi:10.1016/j.dsp.2007.05.005

[6]. N. Jafarnia-Dabanlooa,1, D.C. McLernona,_, H. Zhangb, A. Ayatollahic, V. Johari-Majdd "A modified Zeeman model for producing HRV signals and its application to ECG signal generation" Journal of Theoretical Biology 244 (2007) 180–189

[7]. Cheng Wen a, Ming-Feng Yeh a, Kuang-Chiung Chang a, Ren-Guey Lee "Real-time ECG telemonitoring system design with mobile phone platform" Measurement xxx (2007) xxx–xxx

[8]. M. Sabarimalai Manikandan , S. Dandapat "Wavelet threshold based ECG compression using USZZQ and Huffman coding of DSM"Biomedical Signal Processing and Control 1 (2006) 261–270

[9]. Johan Coosemans , Bart Hermans, Robert Puers "Integrating wireless ECG monitoring in textiles"Sensors and Actuators A 130–131 (2006) 48–53

[10]. Nurettin Acýr "A support vector machine classifier algorithm based on a perturbation method and its application to ECG beat recognition systems" Expert Systems with Applications 31 (2006) 150–158

[11]. *Yeong Pong Meaua, Fatimah Ibrahima*, *Selvanathan A.L. Narainasamyb, Razali Omarc* "intelligent classification of electrocardiogram (ECG) signal using extended Kalman Filter (EKF) based neuro fuzzy system" computer methods and programs in biomedicine 8 2 (2006) 157–168

[12]. H. Gholam Hosseini a,c,1, D. Luob,*, K.J. Reynolds c "The comparison of different feed forward neural network architectures for ECG signal diagnosis" Medical Engineering & Physics 28 (2006) 372–378

[13]. Gabriele Wehr, MDa,T, Ron J. Peters, MDb, Khalife´ , Adrian P. Banning, Volker Kuehlkamp, Anthony F. Rickards, Udo Sechtem, " vector-based, 5-electrode, 12-lead monitoring ECG (EASI) is equivalent to conventional 12-lead ECG for diagnosis of acute coronary syndromes" Journal of Electrocardiology 39 (2006) 22–28

[14]. Mehmet Engin, Yalc¸ýn Yamaner, Erkan Zeki Engin "A biotelemetric system for human ECG measurements" Measurement 38 (2005) 148–153

[15]. Marek Malik, PHD, MD, FACC, Velislav N. Batchvarov "Measurement, Interpretation and Clinical Potential of QT Dispersion"Journal of the American College of Cardiology Vol. 36, No. 6, 2000