NOVEL TECHNIQUES FOR CARDIAC ARRHYTHMIA DETECTION

By

Eng. Mohamed Ibrahim Ismail Owis
Systems and Biomedical Engineering Department
Faculty of Engineering, Cairo University

A Thesis Submitted to the
Faculty of Engineering, Cairo University
In Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

In

SYSTEMS AND BIOMEDICAL ENGINEERING

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
November 2001

NOVEL TECHNIQUES FOR CARDIAC ARRHYTHMIA DETECTION

By

Eng. Mohamed Ibrahim Ismail Owis
Systems and Biomedical Engineering Department
Faculty of Engineering, Cairo University

A Thesis Submitted to the
Faculty of Engineering, Cairo University
In Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

In

SYSTEMS AND BIOMEDICAL ENGINEERING

Under the Supervision of

Prof. Dr. Abou-Bakr Mohamed Youssef

Department of Systems and Biomedical Engineering, Faculty of Engineering, Cairo University.

Dr. Yasser Moustafa Kadah

Department of Systems and Biomedical Engineering, Faculty of Engineering, Cairo University.

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
November 2001

Novel Techniques for Cardiac Arrhythmia Detection

By

Eng. Mohamed Ibrahim Ismail Owis
Systems and Biomedical Engineering Department
Faculty of Engineering, Cairo University

A Thesis Submitted to the
Faculty of Engineering, Cairo University
In Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
In
SYSTEMS AND BIOMEDICAL ENGINEERING

Approved by the
Examining Committee:

Prof. Dr. Abd El-Moneim Abd El-Zaher Wahdan, Member

Prof. Dr. Mohamed Emad Mousa Rasmy, Member

Prof. Dr. Abou-Bakr M. Youssef, Thesis Main Advisor

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
November 2001

Acknowledgments

First and foremost, thanks to God the most merciful, and most gracious.

I wish to thank Prof. Abou Bakr M. Youssef for his invaluable time and his insightful comments that improved the quality of this dissertation.

I also wish to thank Dr. Yasser M. Kadah, for his kind supervision and creative suggestions through my Ph.D. study. Without his thought-provoking guidance and never-ending encouragement, this dissertation would not have been possible.

Also, I wish to thank my parents for their ceaseless love and firm support and for instilling in me a love of learning. I would like to thank my wife for enduring a seemingly endless ordeal, for sacrificing some of her best years so that I could finally finish this Ph. D. research.

Finally, I would like to thank the administrative team and workers in the department for their cooperation.

This work was supported in part by IBE Technologies, Egypt.

Abstract

The heart is responsible for supplying blood to the whole body. Synchronization is very important in stimulating the cardiac muscle. Normally, the electrical signal starts at the sinoatrial (SA) node, which is located in the right atrium. The stimulation spreads from the SA node into the right atrium and then to the left atrium. This activates both atria to contract and pump the blood simultaneously into the ventricles. Then, the electrical signal is transferred to the ventricles by the way of the atrioventricular (AV) junction and the bundle branches. The AV junction consists of the AV node and bundle of His. The stimulus spreads into the ventricular myocardium through specialized conducting cells called Purkinje fibers. That leads to ventricular contraction with pumping blood to the lungs and the general circulation.

An arrhythmia is a change in the regular beat of the heart (i.e., abnormal rhythm). The heart may seem to skip a beat or beat irregularly, at a rate that may be too fast, or too slow. Besides the normal causes of arrhythmia that include stress, caffeine, tobacco, and some medicines, arrhythmias are an excellent indicator of the presence of a heart disease.

Arrhythmias are classified according to their origins, ventricular or supraventricular and by the kind of rate change. In general, ventricular arrhythmias are the most serious and in fact can be life threatening in some cases. Therefore, the detection of such arrhythmias is essential to the life of the patient.

In this thesis, we consider the problem of detection and classification of cardiac arrhythmias of ventricular origin. In particular, we study the characteristics of such conditions as ventricular couplet, ventricular bigeminy, ventricular tachycardia, and ventricular fibrillation. The problem is divided into two steps: feature extraction and classification. Novel techniques of feature extraction are applied to characterize ECG signals that include features based on dynamical modeling (correlation dimension and largest Lyapunov exponent), principal component analysis (PCA), and independent component analysis (ICA). Regarding classification, we applied the following

classifiers: minimum distance classifier, k-nearest neighbor classifier (k-NN), and Bayes minimum-error classifier. The statistical analysis of the results discusses the robustness of the new approaches and suggests their implementation for clinical use.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction and Literature Review

1.1 Introduction

The early detection of abnormal heart conditions is vital for intensive care unit patients. Sudden cardiac death remains a major unresolved clinical and public health problem. There are more than 300,000 sudden cardiac deaths each year of which ventricular fibrillation (VF) is implicated in the vast majority. The detection of such conditions is possible through continuous monitoring of electrocardiographic (ECG) signals to detect the presence of arrhythmia. The accurate detection of such conditions at an early stage is essential for timely management of the case.

Conventional methods of monitoring and diagnosing arrhythmia rely on detecting the presence of particular signal features by a human observer. Due to the large number of patients in intensive care units and the need for continuous observation of such conditions, an automated arrhythmia detection is needed. Several techniques for automated arrhythmia detection have been developed in the past ten

years to attempt to solve this problem. Such techniques work by transforming the mostly qualitative diagnostic criteria into a more objective quantitative signal feature classification problem.

1.2 Thesis Objective

The feature classification problem of arrhythmia detection is divided into two main steps. First the extraction of robust features out of the measured ECG signals. Second is the detection of the existence of an arrhythmia and classification of the arrhythmia type.

- The main objective of this thesis is proposing the application of three novel techniques of feature extraction on ECG signals. These techniques are:

  a) Features based on nonlinear dynamical modeling (chaos theory). We address the problem of characterizing the nonlinear dynamics of the ECG signal and its variation with different arrhythmia types. Two important chaotic system parameters namely, the correlation dimension and largest Lyapunov exponent, form feature vector to be used in discriminating between different ECG signal types. The resultant values are compared to detect statistically significant differences among different arrhythmia types.

  b) Two blind source separation (BSS) techniques: principal component analysis (PCA) and independent component analysis (ICA). ECG signal samples are utilized to compute the basic components of the ECG signals using PCA and ICA. The signals are preprocessed to obtain the magnitude of their Fourier transformation in order to reduce the number of components resulting from different shifts of the same signal (will be explained in details in chapter 4). The ECG signal window at hand is projected onto such components and the projection magnitudes are considered as signal features.

- The suggested parameters are evaluated for a large number of real ECG signals within each class (four arrhythmia classes besides the normal case) from the MIT-BIH Arrhythmia Database [MIT-BIH 1997] and the results are reported. Feature vectors from all signals in the training database are collected and used to define the feature space of the problem. Each of the feature extraction techniques is treated separately; i.e. chaos-based features, PCA components, and ICA components. Subsequent features from the test set are classified to the closest type within that feature space using statistical classification techniques: minimum distance classifier, Bayes minimum-error classifier, and voting k-nearest neighbor (k-NN) classifier.

Arrhythmia types under study in this thesis are ventricular in origin, since the ventricular arrhythmias are more seriously life-threatening arrhythmias. These arrhythmia classes are:

- Ventricular Couplet (VC).
- Ventricular Tachycardia (VT).
- Ventricular Bigeminy (VB).
- Ventricular Fibrillation (VF).

1.3 Literature Review

Many techniques to detect arrhythmias have been reported in the literature that include, in general, time-domain, frequency-domain, time-frequency analysis techniques. Our literature review also includes the review of previous works using the methods we use; namely chaos theory, PCA, and ICA.

1.3.1 Time Domain Techniques:

Time domain are usually computationally less intensive, use simple algorithms and suitable for real-time execution. But the performance of these algorithms is not ideal and each can be improved. Examples of time domain algorithms are:

- The analysis of ECG signals for arrhythmia detection using the autocorrelation function. Peak analysis of the autocorrelation function of the ECG signals is used to select some parameters for discrimination between VT and VF. It has been based on the period stability and repeatability of wave patterns [Guillén *et al.* 1989].

- Sequential hypothesis technique was applied by Thakor *et al.* on the discrimination of VT and VF [Thakor & Pan 1990, Thakor *et al.* 1990]. A sequential hypothesis test procedure is developed that repeats the test to detect VF or VT until a decision is reached. The algorithm consists of three steps:

    1. The ECG signals are converted to binary sequences by comparison with a threshold (20% of the peak value of the signal for each 1 sec. record).
    2. A probability distribution is constructed from threshold crossing intervals (TCI) for VF and VT.
    3. VF is discriminated from VT by sequential hypothesis testing (applied on subsequent 1-sec data segments).

- The sequential tests have been generalized for simultaneous discrimination of several rhythms (supraventricular tachycardia and ventricular tachycardia) from normal sinus rhythm using atrio-ventricular delay (AV_d) as a test parameter [Thakor *et al.* 1994]. The AV_d parameter is typically determined as the positive delay between each atrial depolarization and the subsequent ventricular depolarization.

1.3.2 Frequency Domain Techniques

- An algorithm based on power spectrum of QRS wave is reported that [Minami *et al.* 1999]:

    1. Extracts individual QRS complexes from ECG/EGM (intracardiac electrogram) signals,

2. Converts each QRS complex to a Fourier spectrum,
3. Classifies the spectrum into three kinds of rhythms: supraventricular rhythm (SVR), ventricular rhythm (VR) including VT and premature ventricular contraction (PVC), and ventricular fibrillation (VF).

Regarding the QRS extraction of this algorithm, it relies on detection of local maxima as R-waves. For VF beats, where there is not an obvious R-wave, not all VF peaks can be reliably detected except for relatively large peaks.

- The aim of another research group is to show how an adaptive recurrent filter structure detects cardiac arrhythmia [Thakor & Zhu 1991]. Their idea is to build an impulse response of the QRS complex and to detect as arrhythmias the signals whose impulse response deviates from normal.

1.3.3 Time-Frequency Techniques

- A time-frequency analysis is reported in [Afonso & Tompkins 1995]. This paper illustrated that time-frequency distributions such as the smoothed pseudo Wigner-Ville distribution (SPWVD), and the cone-shaped kernel (CKD) method have better time and frequency resolution than the short time Fourier transform (STFT) of normal and ventricular rhythms. It demonstrates that accurate methods of computing the time-frequency domain should be found for ECG signals [Afonso & Tompkins 1995].

- Also, wavelet theory was introduced by another research group as a time-frequency representation technique to provide a method for enhancing the detection of cardiac arrhythmias [Khadra *et al.* 1997]. An extension of this work using artificial neural networks in classification is published in [Al-Fahoum & Howitt 1999].

1.3.4 Nonlinear Dynamical Systems (Chaos Theory)

- One research group studied the estimation of Lyapunov exponents of the ECG time series aiming to characterize the dynamics of cardiac electrical activity

through the analysis of the Lyapunov spectra of ECG signals [Casaleggio *et al.* 1997]. They reported that in order for their tool to be used as a diagnostic tool it is necessary to repeat the analysis on a large data set.

- The other study discussed the chaotic behavior of the ECG signals. They reported some results of the correlation dimension (D_2) and the largest Lyapunov exponent (λ_1). The data set used was only two signals for each ECG signal type [Govindan *et al.* 1998].

Given that such techniques are particularly sensitive to parameter variations, it is not possible to directly utilize these results or attempt to draw conclusions based on these studies about the robustness of their implementations. Therefore, we propose a study that involves the analysis of ECG chaotic behavior based on a large number of signals of different arrhythmia types using a more detailed implementation of the feature extraction steps. That would be useful to show the advantages and the limitations of such class of nonlinear analysis.

1.3.5 Blind Source Separation: Principal Component Analysis (PCA) and Independent Component Analysis (ICA)

Blind source separation techniques have been applied in several aspects of ECG signal processing. Examples of such applications include:

- Separating the fetal and maternal ECG signals [Zarzoso *et al.* 1997, Lathauwer *et al.* 2000].
- Analysis of ST segment for ischemia detection [Stamkopoulos *et al.* 1998].
- Identification of humans using ECG [Biel *et al.* 2001].
- They have been also used in related areas to diagnose peripheral vascular disease from multiple blood flow measurements [Panerai *et al.* 1988].

While the blind source separation based techniques were shown to be successful in the above applications, their use in ECG signal classification has not been addressed in the literature.

1.3.6 Discussion of the Literature

Even though fairly good results have been obtained using previous techniques, we may put our comments as follows:

1. Only 1-3 classes of arrhythmias are included (mostly VT & VF). That was based, mostly, on the application of automatic arrhythmia detectors in automatic implantable defibrillators (AIDs), which are mainly concerned with VF. On the other hand, we in this thesis include four ventricular arrhythmias (VC, VB, VT, VF), which are included in most practical intensive care unit (ICU) monitors.

2. Small data sets are used that makes, statistically, results inaccurate. Due to the stochastic of signals, such studies did not allow the extraction of a general statistical description of the different features of the cardiac arrhythmia types. In this thesis, a large data set is used that is divided into 320 ECG signals in the training data set (64 in each class) and 160 ECG signals in the testing data set (32 in each class).

3. Techniques that ignore the underlying nonlinear signal dynamics, seem to provide only a limited amount of information about the signal.

4. Moreover, the details of implementation of feature extraction techniques were not discussed.

1.4 Thesis Organization

The thesis consists of eight chapters and an appendix organized as follows:

- Chapter 1 titled “Introduction and Literature review,” gives an introduction to the problem and a literature review of previous work.

- Chapter 2 titled “Medical Background,” presents the medical background of this thesis. Medical information of the mechanism of the heart and its electrophysiology along with the ECG is presented. Arrhythmia is introduced as the abnormal cardiac rhythm and ventricular arrhythmias are explained in details.

- Chapter 3 titled “Nonlinear Dynamical Systems (Chaos Theory),” contains the computation of two parameters: correlation dimension and the largest Lyapunov exponent. It starts with the state space reconstruction in which those parameters are extracted.

- Chapter 4 titled “Blind Source Separation Techniques (PCA and ICA),” covers the two blind source separation techniques used in this thesis: principal component analysis (PCA) and independent component analysis (ICA).

- Chapter 5 titled “Significance Test,” reviews the t-test, which is used in several parts of this work.

- Chapter 6 titled “Statistical Classifiers,” reviews the three statistical classifiers, which are used for detection and classification between different ECG signal types.

- Chapter 7 titled “Results and Discussion,” demonstrates and discusses the results of applying the proposed features to discriminate between different ECG signal types.

- Chapter 8 titled “Conclusions and Future Work,” concludes the thesis and gives some suggested future work as extension to our work.

Chapter 2

Medical Background

2.1 Mechanism of the Heart

The heart functions as a pump propelling blood throughout the body and collecting blood circulating back from the body. The role of the circulation mechanism is to deliver oxygen and essential metabolites to the tissues of the body and eliminate waste products and carbon dioxide. The heart has four chambers divided into left and right, and upper and lower sides. Two upper chambers are called atria (receiving chamber), while two lower chambers are called ventricles (pumping chamber). The blood can flow from atrium to ventricle via atrioventricular valves. Following the path of the blood, the right atrium receives unoxygenated blood from lower and upper part of the body and heart itself via inferior vena cava, superior vena cava, and coronary sinus. After the collected blood fills the right atrium, the tricuspid valve (right atrioventricular valve) opens while right atrium pumps blood to right ventricle. After a short delay for filling, the right ventricle pumps low oxygen blood into the pulmonary arteries and to the lungs. The oxygenated blood circulates back to the left atrium via

pulmonary veins, and passes to the left ventricle via the bicuspid valve (left atrioventricular valve) by pumping of the left atrium. Finally, the left ventricle contracts to eject oxygenated blood to the aorta, which distributes it to the peripheral tissues. (Note: the right and left atriums function simultaneously, as do the right and left ventricles.)

2.2 Electrophysiology of the Heart

2.2.1 Action Potential of Heart Cells

The mechanical mechanism of the heart is founded on and tightly related to electrical activity of heart cells. The activity of each heart cell is influenced by the flow of sodium and potassium ions across the cell membrane. Furthermore, the electrical activity of one cell or part of one cell can interact with its surroundings. In the resting state of the cell, the difference in potential between inside the cell and the outside, called the resting membrane potential, is in range of – 60 mV to – 90 mV (inside cell has lower potential than the outside) [Wagner 2001]. The electrical potential remains relatively unchanged, without any external stimulus. An electrical stimulus from an external source or from neighboring cells or parts of cells with potential higher than a threshold level increases the cell membrane permeability to sodium ions, and causes a rapid sequence of changes in the membrane potential, called an action potential. The underlying electrophysiology in an action potential is as follows [Hobbie 1988].

- Initially, the electrical stimulus cause a rapid change in permeability of the membrane to sodium; sodium ions flow into the cell abruptly increasing the membrane potential to 20 mV. This phenomenon is called depolarization and is shown as phase 0 in figure 2.1.

- Eventually the membrane potential will decrease and return to resting state, called repolarization. This process happens slower than depolarization, and it can be divided into three phases.

- Phase 1 is a slight decrease in potential due to the opening of potassium channels.

- In phase 2, the membrane potential remains almost constant in the plateau phase, due to balance of sodium and potassium ions.

- Finally, in phase 3, the membrane potential decreases continuously to level of membrane resting potential [Wagner 2001].

An action potential in one part of a cell stimulates an action potential in another cell or part of the same cell dependent on its electrical influence or ability to meet the threshold. This mechanism is what is used to "conduct" electrical activity through all the muscle cells of the heart.

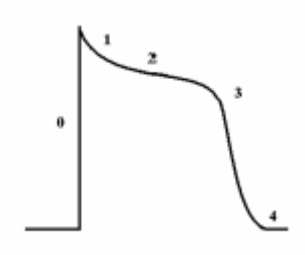

Figure 2.1 Schematic of action potential in ventricle cell. (0 = depolarization, 1,2,3 = repolarization, 4 = diastolic phase)

2.2.2 Normal Cardiac Rhythm

The electrical stimulation of the heart normally starts in the pacemaker cells of the sinoatrial (SA) node (also called sinus node). The location of SA node is at the right atrium near the opening of the superior vena cava. The SA node functions as a pacemaker, and automatically generates electrical pulses at 60 to 100 cycles per minute. The electrical stimulus distributes through the right atrium and then into the left atrium; see figure 2.2. As a result, right and left atria pump blood simultaneously to right and left ventricles.

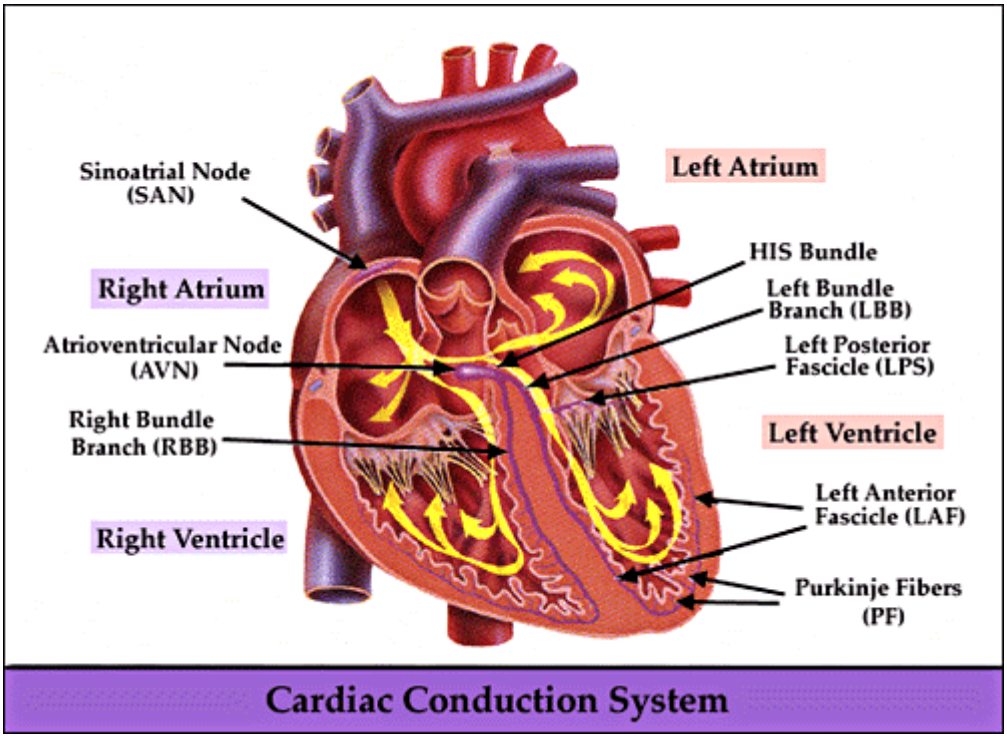

Figure 2.2 Cardiac Conduction System.

The spread of electrical stimulus stops at the junction between atrium and ventricle except at small conduction tissues located at the end of the interatrial septum. This collection of conducting tissue is called the atrioventricular (AV) junction. The AV junction acts as an electrical bridge connecting the atria and ventricles. Besides conducting an electrical stimulus to the ventricles, the AV junction also delays the stimulus to ensure that the blood completely flows from atria to ventricles. The AV junction includes the AV node, which is the distal (upper) part of AV junction, and the bundle of His, which is the proximal (lower) part of AV junction. The transmission of the stimulus is conducted to left and right ventricular myocardium (ventricular muscle) via left and right branches of the bundle of His, respectively. The electrical stimulus spreads out broadly over ventricular muscle by way of Purkinje fibers connected to the branches.

The purpose of the bundle and Purkinje fibers is that they conduct faster than in the AV node or regular cardiac muscle cell distributing electrical potential to many

parts of the ventricle at once. In response, the ventricles pump blood into pulmonary arteries (for right ventricle) and aorta (for left ventricle) [Goldberger 1999].

2.3 Electrocardiogram

An electrocardiogram (ECG) is a graphical recording of electrical voltages generated by heart (atrium and ventricle muscles). The measurement is performed by means of patch electrodes placed on the surface of the body. The ECG measurement can be grouped as extremity leads, and chest leads as shown in figure 2.3 [Goldberger 1999].

- The six-extremity (limb) leads record voltages from the heart that are directed onto the frontal plane of the body. The extremity leads record six voltage differences from electrodes on the limbs, and each lead has two subgroups, unipolar and bipolar. The extremity leads include
  - Three bipolar extremity leads (I, II, and III). A bipolar lead records the difference between voltages from the heart detected at two extremities. The bipolar extremity leads can be represented by Einthoven's triangle. They are related by the equation II = I + III.
  - Three augmented unipolar extremity leads (aV_R, aV_L, and aV_F). A unipolar lead records voltages at one point relative to zero potential. The unipolar extremity leads can also be represented by a triaxial diagram. They are related by the equation $aV_R + aV_L + aV_F = 0$.

  As a general rule, the P-QRS-T pattern in lead I resembles that in lead aV_L. Leads aV_R and II usually show reverse patterns. Lead aV_F usually resembles lead III.
- The six chest-leads (V_1 to V_6) record voltages from the heart as directed onto the horizontal plane of the body, from the front and the left side. The chest leads attach to six positions on the chest overlying the 4^{th} and 5^{th} rib spaces as shown in figure 2.3.

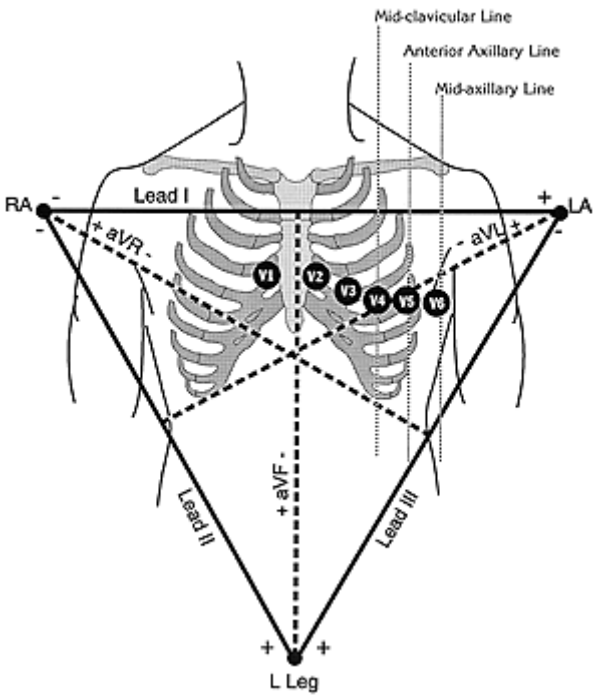

Figure 2.3 Twelve ECG leads.

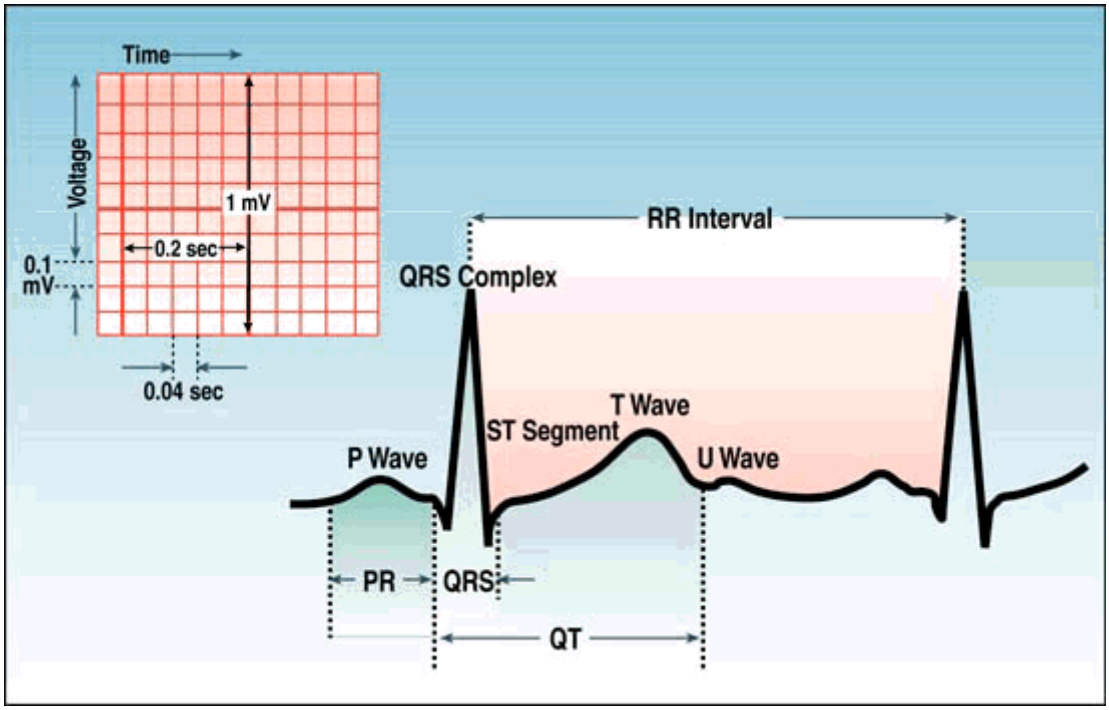

Figure 2.4 A drawn schematic of sinus rhythm ECG and intervals in PQRST complex.

The waveform of one cycle of normal heart rhythm, called sinus rhythm, can be labeled corresponding to each deflection as shown in figure 2.4. The deflections of the ECG represent events from certain parts of the heart. The P wave represents atrial depolarization, and QRS complex is ventricular depolarization. The interval from Q to S is the time required for an electrical stimulus spread through the ventricles. The T wave represents ventricular repolarization. The last phase of the repolarization may appear as a small round deflection, called a U wave [Hampton 1999]. Each electrical event corresponds to a mechanical event where the atrial contraction follows the P wave and ventricular the QRS.

Four basic intervals are measured on the normal ECG:

1. The heart rate (based on RR interval).
2. The PR interval is measured from the beginning of the P wave to the beginning of the QRS complex. The normal PR interval varies from 0.12 - 0.2 seconds.
3. The QRS complex is normally 0.08 - 0.1 second or less in width.
4. The QT interval is measured from the beginning of the QRS complex to the end of the T wave. It varies with the heart rate, becoming shorter as the heart rate increases.

These four intervals can be considered the “vital signs” of the ECG because they give essential information about the electrical stability (or instability) of the heart.

2.4 Arrhythmia

The term “arrhythmia” refers to all rhythms other than regular sinus rhythm. Even the slight variation in sinus rate caused by altered autonomic balance during the respiratory cycle is termed sinus arrhythmia. The term “dysrhythmia” has been proposed by some as an alternative, but “arrhythmia,” meaning “imperfection in a regularly recurring motion,” is the commonly accepted term for rhythms other than

regular sinus rhythm. Generally, the term arrhythmia refers to any heart rhythm other than strictly normal sinus rhythm.

The presence of an arrhythmia does not necessarily reflect cardiac disease, as indicated by the broad array of abnormal rhythms that commonly occur in healthy individuals of all ages. Even in healthy people the SA node may not pace the heart at a perfectly regular rate, and a slight beat-to-beat variation of the sinus rate occurs with sinus arrhythmia. If the heart rate becomes greater than 100 beats/min, it is recognized as sinus tachycardia and sinus bradycardia if less than 60 beats/min. Sinus bradycardia and sinus tachycardia have multiple causes, and both rhythms may occur normally. The most common cause of sinus arrhythmia is respiration. The rate of variation is typically phasic with respiration (respiratory sinus arrhythmia). On the other hand, SA node dysfunction (because of either pacemaker cell failure or exit block) may lead to sinus pauses or, in extreme cases, SA arrest. Prolonged sinus arrest causes fatal asystole unless normal sinus rhythm resumes or escape beats from other foci in the atria, atrioventricular junction, or ventricles take over.

The causes of cardiac arrhythmia are usually one or a combination of the following abnormalities in the conduction system of the heart:

1. Abnormal rhythmicity of the pacemaker.
2. Shift of the pacemaker from the sinus node (SA) to other parts of the heart (e.g. atria, A-V junction, or ventrilcles).
3. Blocks at different points in the transmition of the impulse through the heart (e.g. 2^{nd} and 3^{rd} degree A-V junction block).
4. Abnormal pathways of impulse transmission through the heart.
5. Spontaneous generation of abnormal impulses in almost any part of the heart.

The sinus node, or sinoatrial (SA) node, is the intrinsic pacemaker of the heart and normally initiates each heartbeat. However, pacemaker stimuli can arise from other parts of the heart, atria, atrioventricular (AV) junction, or ventricles. The terms ectopic pacemaker and ectopic beat are used to describe these non-sinus heartbeats. Ectopic beats are often premature; that is, they come before the next sinus beat is due. Thus atrial premature beats (APBs), AV junctional premature beats (JPBs), and ventricular premature beats (VPBs) may be found. Ectopic beats can also come after a pause in the normal rhythm, as in the case of AV junctional or ventricular escape beats. Ectopic beats originating in the AV junction (node) or atria are referred to as supraventricular (i.e., coming from above the ventricles). Generally, arrhythmias are primarily classified according to their rate and site of origin as follows.

- According to the heart rate, there are two kinds of arrhythmias bradyarrhythmia and tachyarrhythmia. The term “bradyarrhythmia” is used to identify any rhythm with a rate < 60 beats/min, and “tachyarrhythmia” is used to identify any rhythm with a rate > 100 beats/min. There are also many arrhythmias that do not alter the rate beyond these normal limits.

- According to the site of origin, an arrhythmia is either have ventricular or supraventricular origins. Ventricular arrhythmias are discussed in details in the next section, as it’s the most life-threatening arrhythmias and then the arrhythmias of interest for our work. Types of supraventricular arrhythmias are Paroxysmal Supraventricular Tachycardia (PSVT), Atrial Flutter, and Atrial Fibrillation (AF). Paroxysmal Supraventricular Tachycardia (PSVT) is a sudden run of three or more supraventricular beats. Atrial flutter and atrial fibrillation are two distinct but related tachyarrhythmias, like Paroxysmal Supraventricular Tachycardia (PSVT), they are ectopic. In other words, with all three types of arrhythmia the atria are being stimulated not from the sinus, or sinoatrial (SA) node, but from ectopic site or sites located in the atria. With PSVT the atria are stimulated at a rate generally between 140 and 250 beats/min. With atrial flutter the atrial rate is even faster, generally 250 to 350 beats/min. Finally, with Atrial Fibrillation (AF) the atrial depolarization rate

is typically between 400 and 600 beats/min. The ECG of a patient with AF shows the rapid irregular fibrillatory waves instead of P waves and ventricular rate that is usually grossly irregular.

There are many different atrial/ventricular relationships among the cardiac arrhythmia according to whether the atria and ventricles are associated or not and what is the rate of each. When the atrial and ventricular rhythms are associated, but the atrial rate is faster than the ventricular rate, the rhythm must originate in the atria. When the atrial and ventricular rhythms are associated, but the ventricular rate is faster than the atrial rate, the rhythm must originate in the ventricles. When the atrial and ventricular rhythms are dissociated, names should be given to both of the rhythms (e.g. atrial tachyarrhythmia with ventricular tachyarrhythmia).

2.5 Ventricular Arrhythmias

When ectopic (non sinus) beats arise in the ventricles themselves, it could produce premature ventricular beats (or complexes), ventricular tachycardia (VT), and sometimes ventricular fibrillation (VF). During ventricular arrhythmia, the normal synchronous sequence of activation described above does not occur. This compromises the pumping function of the heart, causing reduced delivery of oxygen to the body. In some cases, no oxygen is delivered to the cells, and they die.

Some arrhythmias require immediate treatments, while no treatment is necessary for others. However, ventricular arrhythmias almost always require treatment immediately due to its control of pumping blood to the body. Otherwise, a patient will reach unconsciousness or death, because improper contraction of the ventricles decreases the amount of oxygenated blood flow to the body and the brain. A myocardial infarction and ischemia or heart muscle dysfunction can potentially cause arrhythmias, especially ventricular fibrillation, leading to sudden cardiac death [Goldberger 1999].

In this thesis, we study detection and classification of ventricular arrhythmias based on their life-threatening nature. We will consider four types of arrhythmia:

1. Ventricular Couplet (VC)
2. Ventricular Bigeminy (VB).
3. Ventricular Tachycardia (VT).
4. Ventricular Fibrillation (VF).

2.5.1 Premature Ventricular Contractions

Premature ventricular contractions (PVC) are also termed ventricular premature beats (VPB), ventricular ectopics, ventricular premature complexes, and ventricular extrasystoles. When a focus within the ventricles prematurely begins myocardial depolarization, the electrocardiogram shows a beat of abnormal morphology. The premature depolarization arise in either the right or left ventricle. Therefore the ventricles are not stimulated simultaneously, and the stimulus spreads through the ventricles in an aberrant direction. Thus the QRS complexes are wide with PVCs. The wide QRS complex will either have initially a high-energy negative, or positive, deflection. The direction of this deflection is dependent upon the exact location of the focus as shown in figure 2.5.

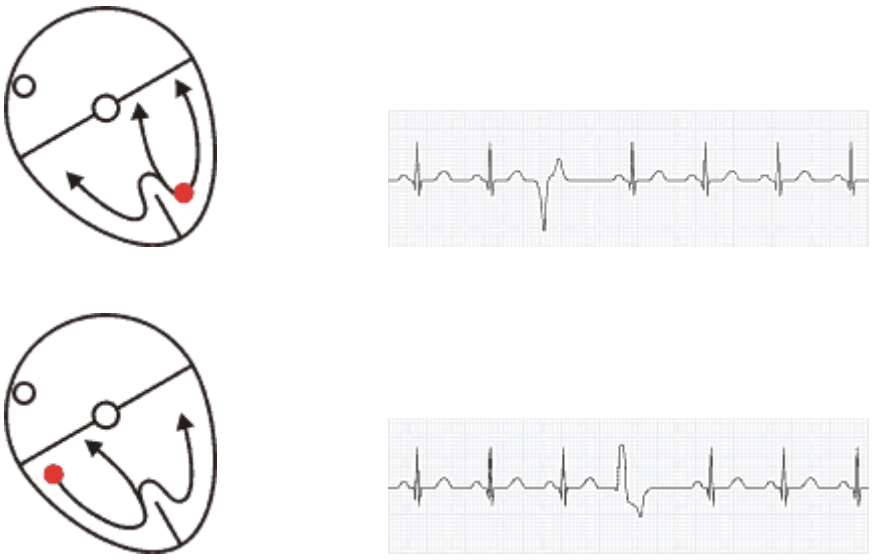

Figure 2.5 Conduction mechanisms and the corresponding ECGs of two examples of PVC.

PVCs have two major characteristics [Wagner 2001].

- PVCs occur before next normal beat.
- PVCs have abnormal shape. The QRS interval is usually 0.12 second or greater, which is abnormally wide. T wave and QRS complex usually orient in opposite direction, and have a fixed coupling interval between PVC and preceding normal beat.

An upright P wave may follow a PVC due to retrograde (reverse) conduction through the AV node. Occasionally, PVCs may arise after a P wave, but before a QRS complex. PVCs may be combined in various fashions. Bigeminy is PVCs occurring every other beat, and every third beat is referred to as trigeminy. If two PVCs appear in a row is referred to as ventricular couplet. Figures 2.6 and 2.7 show examples of ventricular bigeminy and couplet. Moreover, PVCs can develop into more severe arrhythmias, such ventricular tachycardia or ventricular fibrillation [Goldberger 1999].

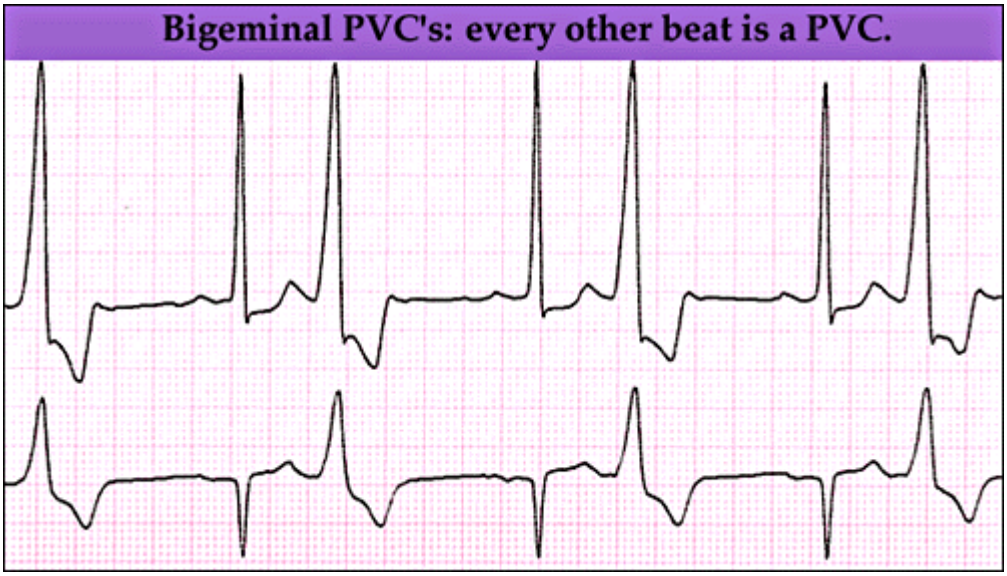

Figure 2.6 Example of ventricular bigeminy.

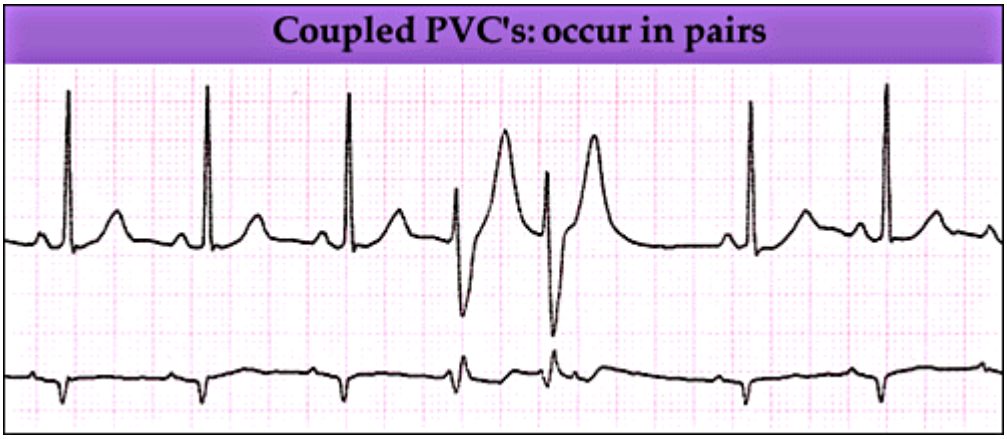

Figure 2.7 Example of ventricular couplet.

2.5.2 Ventricular Tachycardia

Ventricular Tachycardia (VT) is formed out of electrical impulses originate from a site in the ventricles causing the heart rate to become excessively rapid. Usually initiated by premature ventricular complexes, ventricular tachycardia is continued by re-entry mechanisms as shown in figure 2.8. The condition is dangerous if the rate increases to the point when the R wave overlaps with the T wave. In this case the arrhythmia can degenerate into Ventricular Fibrillation.

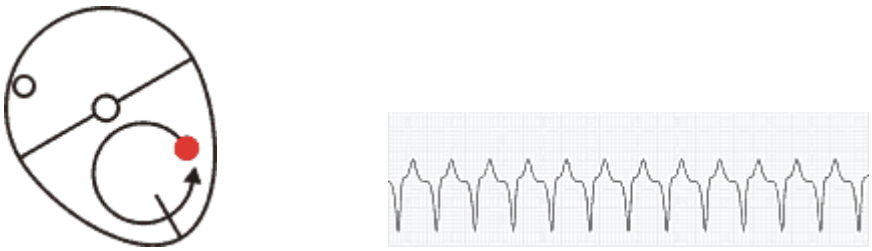

Figure 2.8 Conduction mechanism and the corresponding ECG of VT.

Ventricular tachycardia may give rise to symptoms such as palpitations, shortness of breath, or lightheadedness, or, depending upon the rate of the arrhythmia, its duration, and the underlying heart disease. With faster heart rates and underlying

heart disease loss of consciousness (syncope) or sudden death may occur. Episodes lasting only a few beats may produce no or minimal symptoms. Tachycardia rates between 110 and 150 may be tolerated even if sustained for minutes to hours. However, faster rates (>180 beats per minute) may cause drops in arterial pressure and produce syncope. Very fast rates (>220) are imminently dangerous because they rarely terminate spontaneously and invariably cause drops in blood pressure and low cardiac output. Most commonly, sufferers of ventricular tachycardia have underlying cardiac disease. Majority of the patients suffer from coronary artery disease. Although most patients having ventricular tachycardia will have underlying coronary disease or severely depressed heart function some have no demonstrable disease of the heart muscle or coronary arteries. Long term treatment for ventricular tachycardia includes medications, implanted defibrillators (ICDs), catheter ablation, or surgery(rarely used).

Ventricular Tachycardia is usually classified by duration (sustained and non-sustained) and/or by morphology (mono-morphic and poly-morphic). Sustained VT typically lasting for more than 30 sec., it is usually a life-threatening arrhythmia. VT whether sustained or not, can also be characterized as monomorphic or polymorphic depending on whether consecutive VPBs have the same or variable appearance in a single lead. Very rapid VT with a sine wave appearance is some times referred to as ventricular flutter. This arrhythmia often leads to ventricular fibrillation (VF).

According to the above VT classification, there is many types of VT among them: the so called Accelerated Idioventricular Rhythm (AIVR), a slow VT with a rate between 50 and 100 beats/min, and the Torsade de pointes (a French term meaning “twisting” of the points) which is a polymorphic VT in which the QRS complexes in the same lead appear to twist and turn in the opposite direction. It is important because of its diagnostic and therapeutic implications. Torsade de points VT is generally seen in the setting of delayed ventricular repolarization.

In our work, we take monomorphic ventricular tachycardia as the featuring pattern for different kinds of ventricular tachycardias. Monomorphic ventricular tachycardia (will be called just VT from now on) is an arrhythmia consisting of three

or more consecutive PVCs with uniform beat-to-beat QRS appearance. The shape of the QRS complex is abnormal, and duration of each complex is 0.12 seconds or greater (usually greater than 0.14 seconds). Occasionally, a P wave will occur following a QRS complex due to retrograde conduction through AV node. The rate usually ranges from 150 to 250 beats per minute. In some cases, the heart rate may be low as 120 beats per minute. Monomorphic ventricular tachycardia can be treated by DC voltage pacing or defibrillation [Goldberger 1999]. An example of monomorphic ventricular tachycardia (VT) is shown in figure 2.9.

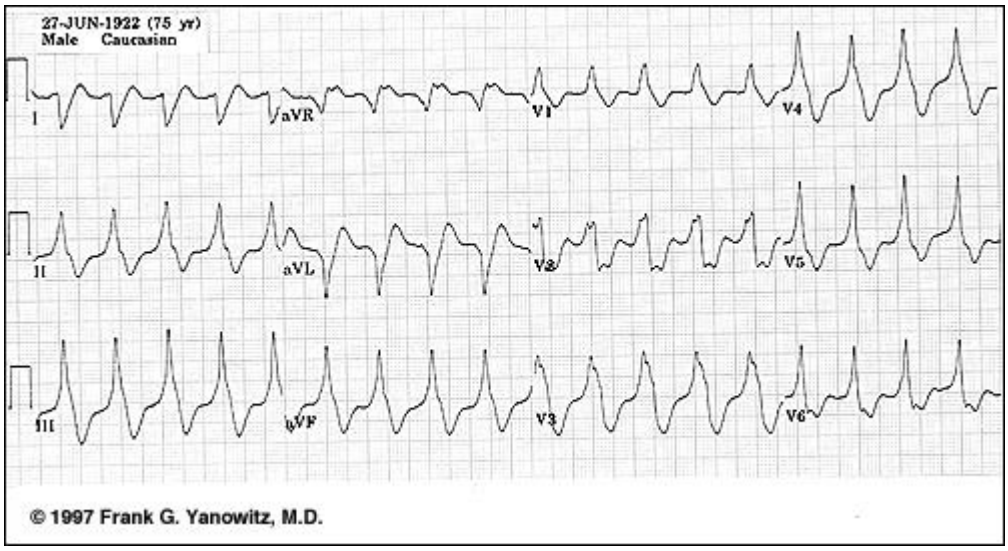

Figure 2.9 Ventricular tachycardia.

2.5.3 Ventricular Fibrillation

Ventricular fibrillation results when multiple sites in the ventricles fire impulses very rapidly in an uncoordinated fashion as shown in figure 2.10. This condition causes an entirely uncoordinated, ineffective contraction that is best regarded as a tremor rather than beat. Although this general scheme appears correct, the exact mechanisms of ventricular fibrillation remain unknown. The ventricles quiver and cease to pump blood effectively, thereby stopping the circulation of blood. Death follows within a

few minutes, unless a normal rhythm is restored with emergency treatment. VF depicts the final sign for death in most patients experiencing out of hospital cardiac arrest.

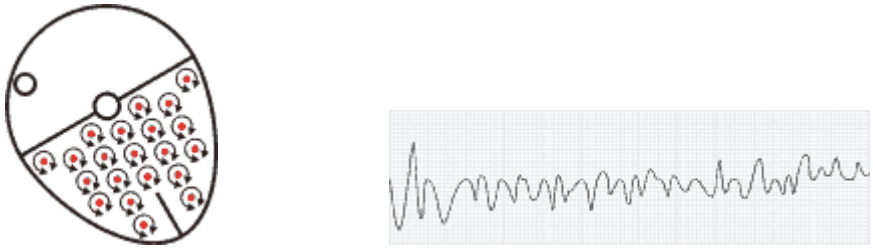

Figure 2.10 Conduction mechanism and the corresponding ECG of VF.

Ventricular fibrillation (VF) occurs when electrical activity in the ventricles is fractioned. The ventricular myocardial fibers do not contract in any coordinated fashion, but fibrillate or quiver ineffectively and asynchronously. Therefore, blood is not pumped to the body, and a patient will become unconscious and collapse within 10 to 20 seconds. Causes of ventricular tachycardia and/or ventricular fibrillation include heart disease, aging, effects of medications, metabolic imbalances and other medical problems. Defibrillation is the only therapy and is required immediately, before any damage to the brain cells and the body. VF is one of three sources of cardiac arrest and the primary cause of sudden cardiac death. Defibrillation must be performed within seconds to save the patient's life. Long-term treatment includes medications or implanted defibrillators.

The ventricular fibrillation is a multifocal arrhythmia, and can appear coarse or fine on the electrocardiogram. P, QRS and T waves are not detectable. The ECG in VF is an undulating pattern without discrete P waves or QRS complexes. The waveform may be either coarse or fine (Fine fibrillation waveform is below 0.2 mV), An example of VF is shown in figure 2.11. Usually, VF has larger waveform at the onset. The rate varies from 150 to 500 beats per minute.

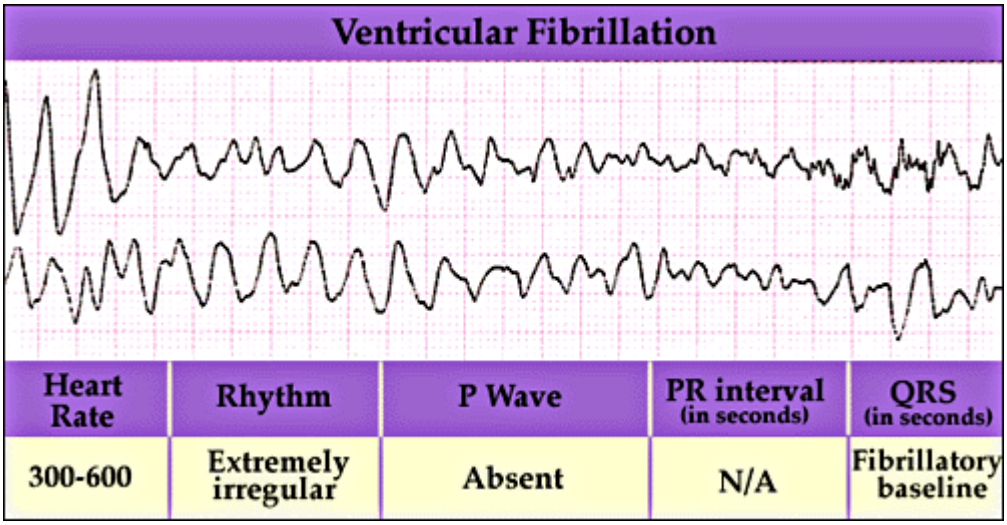

Figure 2.11 Ventricular fibrillation.

Chapter 3

Nonlinear Dynamical Systems (Chaos Theory)

3.1 Introduction

"Dynamical systems" is the branch of mathematics that attempts to understand processes in motion. Such processes occur in all branches of science, for example the world's weather. Since a dynamical system is moving or changing in time, the main interest in studying dynamical systems is to predict where the system is heading, where it will ultimately go. Clearly, some dynamical systems are predictable, whereas others are not. For example, predicting the weather a month from now seems impossible. This can be simply because too many variables present in the system (e.g. the meteorological system); that is a true but incomplete explanation. One of the remarkable discoveries of the twentieth-century mathematics is that very simple systems, even systems depending on only one variable, may behave just as unpredictably as any wild disturbances in the weather. Mathematicians have called the reason of this unpredictable behavior "chaos" [Devaney 1992].

In the last two decades, there has been an increasing interest in applying techniques from the domains of chaos theory in studying biological systems [Abarbanel *et al.* 1998]. For example, more complete dynamical information can be obtained when we model the heart as a multivariate, non-linear pumping system that sometimes exhibits unpredictable (chaotic) ECG pattern. This is a direct consequence of the complex nature of ECG signals, which involves interactions between several physiological variables including autonomic and central nervous regulation, hemodynamic effects, and other extrinsic variables.

In the field of chaotic dynamical system theory, several features can be used to describe system dynamics including correlation dimension (D_2), Lyapunov exponents (λ_k), approximate entropy, etc. These features have been used to explain ECG signal behavior by several studies (c.f., [Casaleggio *et al.* 1997, Govindan *et al.* 1998]). Nevertheless, these studies applied such techniques only to a few sample ECG signals. Due to the stochastic of such signals, such studies did not allow the extraction of a general statistical description of the dynamics of different arrhythmia types. Moreover, the details of implementation of feature extraction techniques were not discussed.

In this work, we address the problem of characterizing the nonlinear dynamics of the ECG signal and its variation with different arrhythmia types. In this chapter, the implementation details to automatically compute two important chaotic system parameters namely, the correlation dimension and largest Lyapunov exponent, are discussed. In particular, the algorithms used, parameter values in each technique, and their selection criteria are given and explained.

The first step is to get the state space trajectory of the ECG time series. Then, extract the parameters (D_2 and λ_1) that represent that trajectory for each signal type.

3.2 State Space Reconstruction

The mathematical description of a dynamical system consists of two parts: the *state* which is a snapshot of the process at a given instant in time, and the *dynamics* which is the set of rules by which the states evolve over time. In the case of the heart as a dynamical system, the available information about the system is a set of ECG measurements from skin-mounted sensors. There is no mathematical description of the underlying dynamics of the heart. That is, we deal only with observables whose both mathematical formulation and total number of state variables are not known. Therefore, to study the dynamics of such system, we first need to reconstruct the state space trajectory. The most common method to do this is using delay time embedding theorem.

3.2.1 Delay Time Embedding

Consider a single-variable sampled ECG time series v(.) (in units of voltage), that consists of N data points evenly-spaced in time: v(1), v(2), v(3), … , v(t), …, v(N). To create a larger dimensional geometric object out of these observables, the time series is embedded into a larger m-dimensional embedding space. The rows of the matrix X of reconstructed state vectors of length m is defined as follows [Pritchard & Duke 1995],

$$
\begin{aligned}
&x(1)= [v(1),v(1+L),\ldots,v(1+(m-1)L)]\\
&x(2)= [v(1+J),v(1+J+L),\ldots,v(1+J+(m-1)L)]\\
&x(k)=[v(1+(k-1)J),v(1+(k-1)J+L),\ldots, v(1+(k-1)J+(m-1)L)],\\
&k=1,2,\ldots,((N-(m-1)L-1)/J)+1. \qquad (3.1)
\end{aligned}
$$

Here, x(k) is a vector that constitutes the k^{th} row in the matrix X, m is the embedding dimension, L is the lag time that is equal to the number of data points between components of each vector, J is the number of data points between vectors (e.g., if vectors are formed from each data point, then J=1), and ((N-(m-1)L-1)/J)+1 is the number of vectors that could be formed from N data points. For chaotic systems, the reconstructed attractor is referred to as a "strange attractor."

The value of m must be large enough for delay time embedding to work. When a suitable m value is used, the orbits of the system do not cross each other. This condition is tested using the false nearest neighbor (FNN) algorithm [Abarbanel *et al.* 1998]. The dimension m in which false neighbors disappear is the smallest dimension that can be used for the given data.

Various algorithms for estimating a suitable time lag (L) for the reconstruction procedure have been proposed. For example, L can be chosen to be the value at which the autocorrelation function reaches zero, 1/e, 0.5, or 0.1 [Pritchard & Duke 1995]. It can also be selected as the value at which the first minimum of the mutual information function occurs [Abarbanel *et al.* 1998]. We used another approach where we used the time window length to calculate L [Albano *et al.* 1988, Kugiumtzis 1996]. The time window length (W) is defined by the time spanned by each embedding vector,

$$W = (m-1)L\,. \tag{3.2}$$

To choose the best embedding dimension m, the FNN criterion was applied and the first zero has been observed at m=8. Then, the optimal time window length (W) is selected to give the best plateau in the plot of the slope S_1 of the log C(r)-log(r) curve vs. log (r) (G-P method). The optimal window length was found to be around 583 ms (i.e., 210 samples at 360 samples per sec). Consequently, the time lag (L) was estimated to be 83 ms using equation 3.2. Examples of reconstructed state spaces of the arrhythmias under study are shown in figure 3.1.

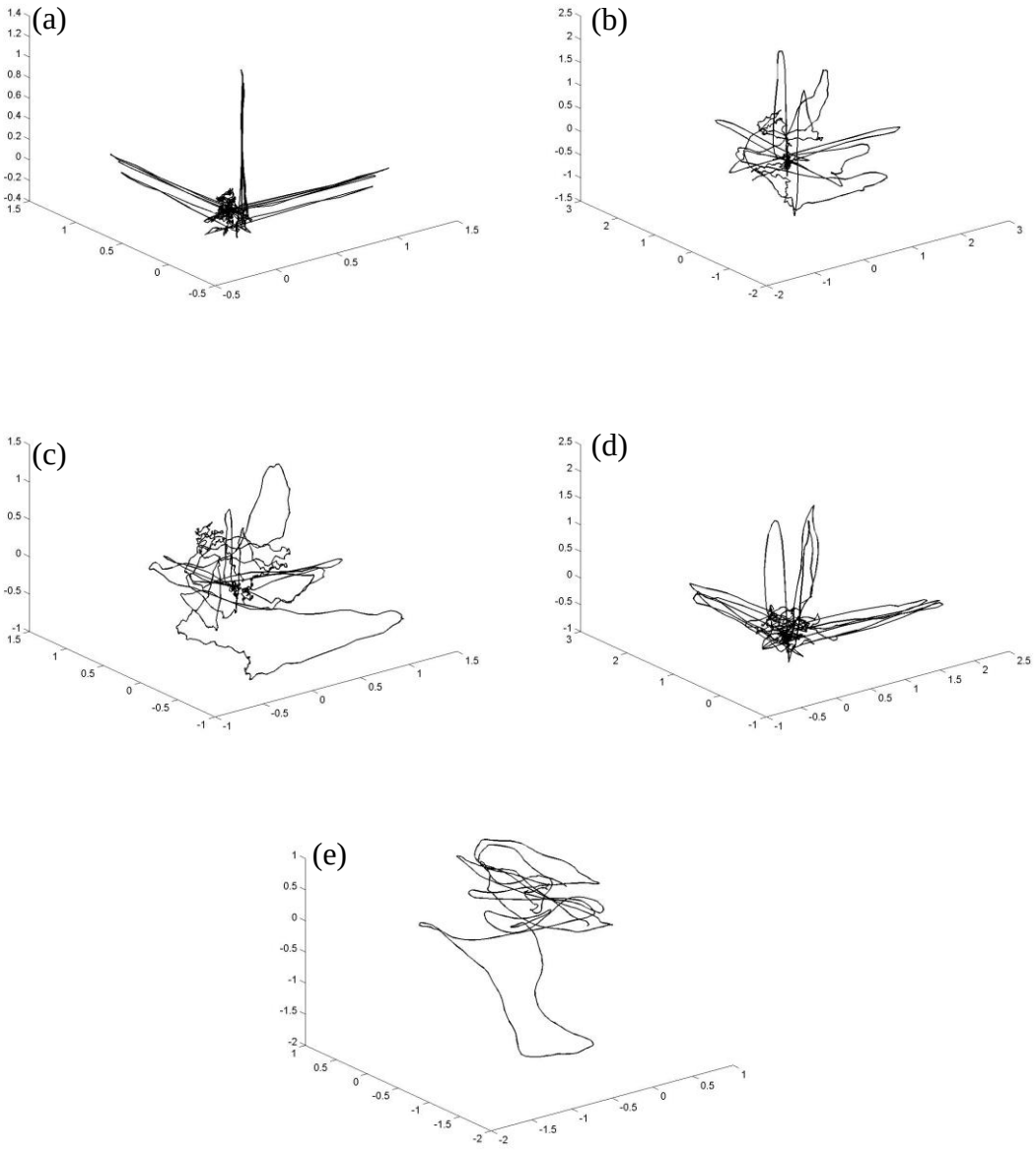

Figure 3.1 3-D plots of the first 3 vectors of the reconstructed phase space of (a) Normal, (b) Ventricular Couplet, (c) Ventricular Tachycardia, (d) Ventricular Bigeminy, and (e) Ventricular Fibrillation.

3.2.1.1 Theiler's Plausibility Argument

Theiler's time embedding algorithm for reconstruction of the state space can be justified by the means of smooth transformation as illustrated in this subsection. Assume a hypothetical system whose dynamics is given by two coupled differential equations as follows:

$$\frac{dx}{dt} = f(x, y) \quad \frac{dy}{dt} = g(x, y) \tag{3.3}$$

Both f and g are assumed to be smooth functions. The state space of the system is illustrated in figure 3.2 (a). A series of smooth transformations will be shown to lead to the delay-time coordinates. Since geometric shapes that can be smoothly transformed one into the other are identical from a topological standpoint, a state-space attractor reconstructed by delay-time embedding will have the same dimension as the system's actual attractor [Pritchard & Duke 1992] (topology is the branch of mathematics dealing with shapes, specifically, what properties remain invariant when shapes are smoothly transformed, with dimension being one of the properties; for example, a circular surface of dimension 2 can be smoothly transformed into a square surface of also dimension 2).

Since f and g are smooth functions, transformation of the y-axis from y(t) to f[x(t),y(t)] will be a smooth transformation as illustrated in figure 3.2 (b). Now we have x(t) on the x-axis and its derivative on the y-axis. Any first derivative can be written:

$$\lim_{L \to 0} [x(t + \Delta t) - x(t)]/L \tag{3.4}$$

Since x(t) is a smooth function of t, it is plausible to assume that this limit is a smooth function of Δt. Therefore, making y-axis time increments finite will be a smooth transformation of the system presented in figure 3.2 (b), where y-axis time is continuous. This transformation is illustrated in figure 3.2 (c). Assuming in this case that Δt equals the time interval between the data points of an experimental time series

(i.e. $\Delta t = L$), then it is a smooth linear transformation to change equation 3.4, the y-axis of figure 3.2 (c), into x(t+L) as seen in figure 3.2 (d). Thus the original state space seen in figure 3.2 (a) has now been transformed in a generically smooth manner into the delay-time coordinates of figure 3.2 (d).

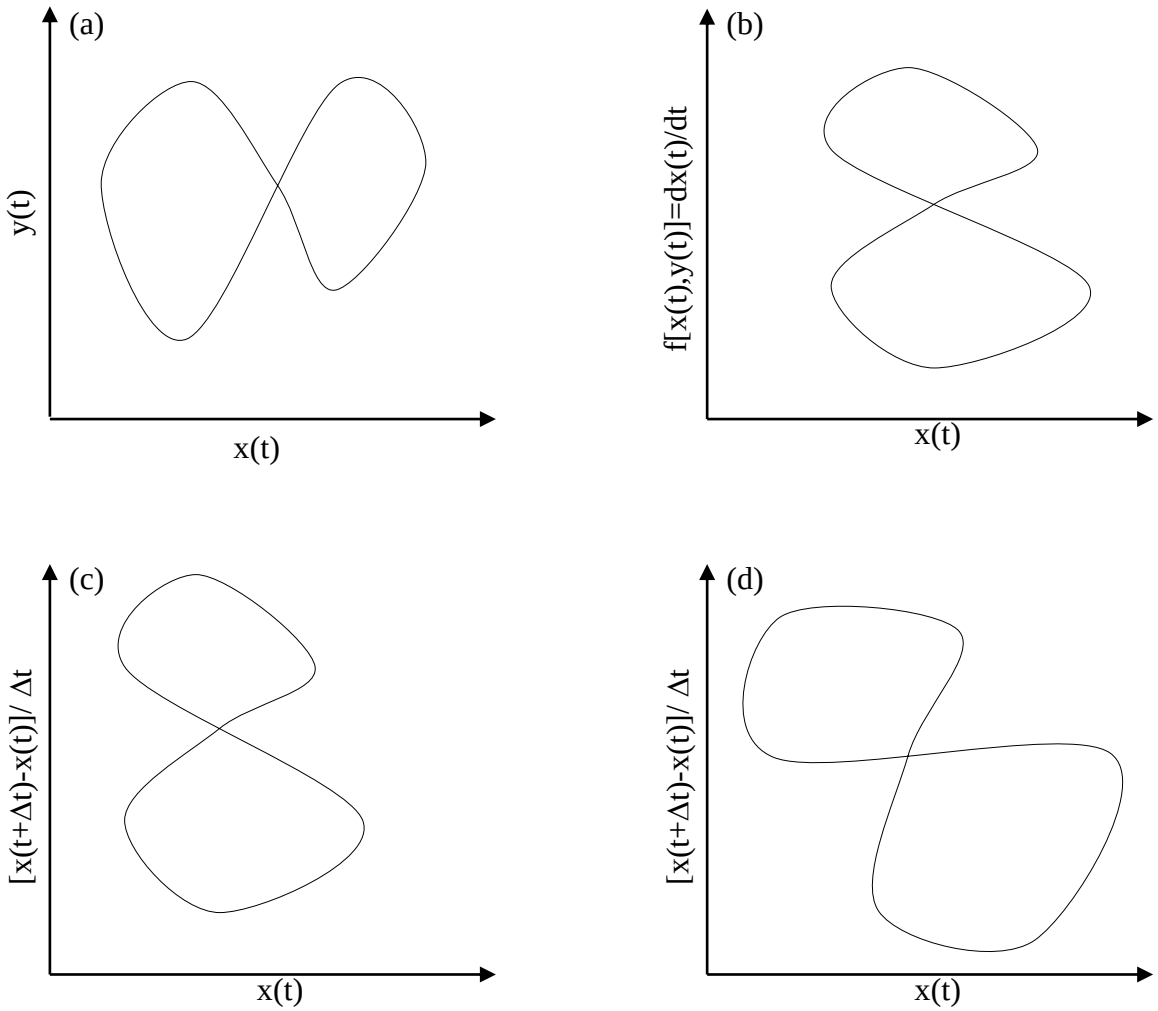

Figure 3.2 Sequence of smooth transformations of the state space of a hypothetical dynamical system (a) yields the delay-time embedding state space (d)

3.3 Estimation of Correlation Dimension

So, after the reconstruction of the state space described in the previous section, we have the state space trajectory of our dynamical system (the ECG time series). In this section we will talk about estimation the correlation dimension as a fractal dimension featuring the ECG signal. First, we will introduce the relation of fractals and the analysis of dynamical systems. Then, Grassberger-Procaccia (G-P) Algorithm is introduced as the most common method of computing the correlation dimension of the reconstructed system trajectory [Grassberger & Procaccia 1983].

3.3.1 Dynamical Systems and Fractals

Technically there is no connection between dynamical systems and fractals. Fractals are geometric objects that are static images, whereas dynamics is the study of objects in motion. However, the importance of fractals in the study of dynamical systems is that, most of the strange attractors are fractals.

3.3.1.1 Fractals

A fractal is a subset of R^n which is self-similar and whose fractal dimension exceeds its topological dimension. Self-similarity means, if we examine small portions of the set under a microscope, the image we see resemble the original. Actually, what distinguishing fractals is not only their self-similarity (lines, squares, and cubes are self-similar), but also, their dimension. Fractals have non-integer dimensions, however lines, squares, and cubes have integer dimensions of 1, 2, and 3 respectively. Naively speaking the reason is that there is only one direction to move along a line (forwards and backwards), for squares we have two directions to move in (x-direction and y-direction), while for cubes we have 3 directions for movement (x, y & z directions).

3.3.1.2 Fractal Dimension

Suppose the affined self-similar set S may be subdivided into K similar pieces, each of which may be magnified by a factor of M to yield the whole set S. Then the fractal

dimension D of S is:

$$D = \frac{\log(K)}{\log(M)} \tag{3.5}$$

The presence of a fractal dimension is an indication for diagnosing the system as "chaotic system".

There are several fractal dimensions, which address different statistical aspects of a set. Fractal dimension can be classified into two main classes:

1. Metric-based as: Hausdoroff, and capacity (D_o) dimensions.

2. Measure-based as: information (D_1), correlation (D_2), pointwise, and Lyapunov dimensions.

Generally speaking all the above are fractal dimensions, measured by different methods. By example, when measuring the fractal dimension by covering the attractor by hypercubes of similar length we got the Hausdoroff dimension, and when covering the attractor by hypercubes of different lengths we got the Capacity dimension. Also, the Correlation and pointwise dimension are very similar. On the other hand, when measuring the dimension using Grassberger-Procaccia algorithm (the method we utilize in this work) we got the Correlation dimension, and when measuring the dimension by another method called pointwise method, very near in concept with Grassberger-Procaccia, we got the pointwise dimension, and so the information and Lyapunov dimensions.

The following subsection illustrates fractal dimensions of two common fractals.

3.3.1.3 Examples of Fractals

3.3.1.3.1 The Cantor Middle-Third

The basic and most famous fractal is the Cantor set. Starting with a line with a unit length. We obtain the Cantor by applying a replacement rule several times. This rule is simply "Remove the middle third of the line each time," as clear in figure 3.3. Applying this rule for a first time, we obtain two pieces of lines, each one has third the original length (i.e. when magnified by a factor of 3 each piece yields the original line). Applying the replacement rule again, each piece yields 2 pieces, of length equals third the original, and so on.

The Cantor has a fractal dimension = Log (2) / Log (3) = 0.63

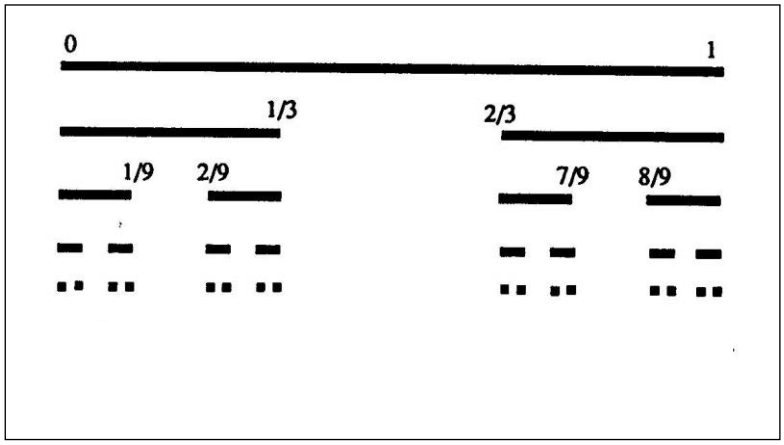

Figure 3.3 Construction of the Cantor middle-thirds set.

3.3.1.3.2 The Sierpinski Triangle

Starting with black triangle, then applying the replacement rule of "Remove from the middle a triangle whose dimensions are half that of the original triangle" as clear in figure 3.4. Applying this rule for a first time, we obtain three smaller triangles (i.e. three pieces) each triangle when magnified by a factor of two resembles the original one. After thousands of replacement rule applications, the Sierpinski triangle will not be a triangle surface of dimension 2 (as the first triangle), and of course it is more

complicated than a typical line or curve of dimension equal 1. Actually, Sierpinski triangle has a dimension somewhere in between 1 & 2.

The Sierpinski triangle has a fractal dimension = Log (3) / Log (2) = 1.58

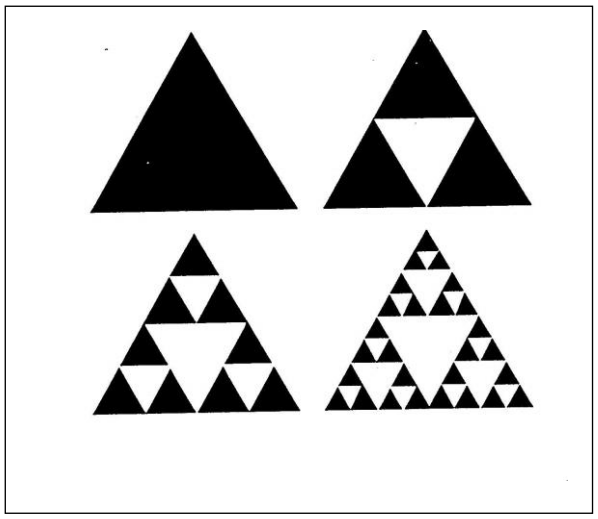

Figure 3.4 Constructing the Sierpinski triangle.

3.3.2 Dimension Estimation using Grassberger-Procaccia (G-P) Algorithm.

The simplest way to think about the dimension D of an object is that it represents the exponent that scales the bulk b of an object with linear distance r (i.e., $b \alpha r^D$). The G-P algorithm uses the correlation integral C(r) to represent the bulk b which is defined as the average number of neighbors each point has within a given distance r. That is,

$$C(r) = \frac{1}{N_p} \sum_{i,j} \theta\left[r - \|x(i) - x(j)\|\right] \tag{3.6}$$

Here, $\|\cdot\|$ symbolizes the Euclidean distance (2-norm) between reconstructed state vectors x(i) and x(j), $N_p = k(k-1)/2$ is the number of distinct pairs of reconstructed state vectors, θ symbolizes the Heaviside unit step function (i.e., $\theta(x)=0$ when $x < 0$ and $\theta(x)=1$ when $x \geq 0$). The correlation dimension D_2 is defined as the slope of the

linear region of the plot of log(C(r)) versus log(r) for small values of r. That is,

$$D_2 = \lim_{r \to 0} \frac{\log[C(r)]}{\log(r)} \tag{3.7}$$

Examples of the plot of log(C(r)) versus log(r) are shown in figure 3.5.

Unlike the calculation of C(r), the determination of the linear scaling region is not an easy task in practice. Because of the presence of noise, it may not be practical to compute the slope for very small values of r. Several different regions may appear visually to be equally valid, which makes this determination not repeatable for manual computation. In our implementation, we tried this approach combined with computerized regression and the results were not satisfactory. Then, we improved our implementation using a second order regression for the whole curve. The linear regression was then obtained for the part of this curve that appeared linear by vision. More consistent values for D_2 were obtained. Finally, we developed an automatic algorithm to determine the linear region to eliminate the need for human interaction. The algorithm can be summarized as follows:

1. Calculate the first derivative of the curve S_1 using the following approximation of differentiation,

$$S_1 = \frac{\partial \log C(r)}{\partial \log r} \cong \frac{\Delta \log C(r)}{\Delta \log r} \tag{3.8}$$

2. Differentiate S_1 once again to obtain the second derivative S_2. The linear region of the curve appears as a number of consecutive points with very small values in S_2.

3. Threshold the absolute value of S_2 to determine the extent of linear region using a small threshold of 0.1.

4. Examine the resultant contiguous linear regions. Discard short linear regions

composed of a sequence of 5 points or less on the curve.

5. If more than one linear region satisfies the above conditions, select the one that yields the maximum D_2 value (i.e., smaller values of r as per its definition in equation 3.6).

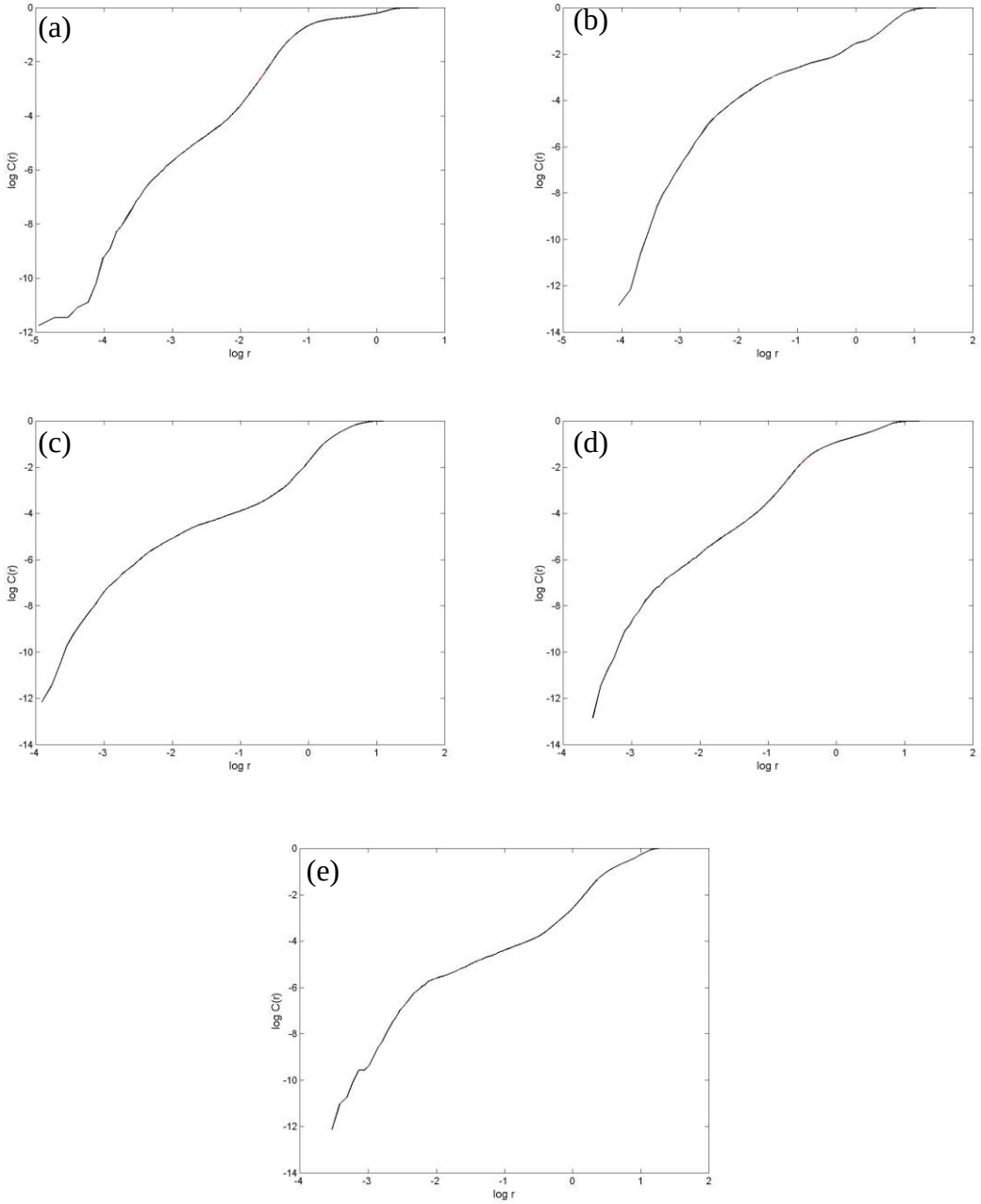

Figure 3.5 log C(r) vs. log r curves of the five ECG signals: (a) Normal, (b) Ventricular Couplet, (c) Ventricular Tachycardia, (d) Ventricular Bigeminy, (e) Ventricular Fibrillation.

3.4 Lyapunov Exponents

Lyapunov exponents quantify the sensitivity of the system to initial conditions, which is an important feature of chaotic systems. Sensitivity to initial conditions means that small changes in the state of a system will grow at an exponential rate and eventually dominate the behavior. Lyapunov exponents are defined as the long time average exponential rates of divergence of nearby states. If a system has at least one positive Lyapunov exponent, then the system is chaotic. The larger the positive exponent, the more chaotic the system becomes (i.e., the shorter the time scale of system predictability). Lyapunov exponents will be arranged such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, where λ_1 and λ_n correspond to the most rapidly expanding and contracting principal axes, respectively. Therefore, λ_1 may be regarded as an estimator of the dominant chaotic behavior of a system.

The presence of a positive exponent is sufficient for diagnosing chaos and represents local instability in a particular direction. It is important to notice that for the existence of an attractor (a stable regime), the overall dynamics must be dissipative (i.e., globally stable) and the total rate of contraction must dominate the total rate of expansion.

Now, consider the case of n-dimensional space where n is the number of state variables used to describe the system. A small n-dimensional hyper-sphere of initial conditions evolves into a hyper-ellipsoid as time progresses. In particular, its principal axes expand (or contract) at rates given by the Lyapunov exponents. Measuring the separation of nearby initial conditions is done along the Lyapunov directions that correspond to those principal axes. The Lyapunov directions depend on the system flow and are defined using the Jacobian matrix (i.e., the tangent map) at each point of interest along the flow.

3.4.1 Calculation of the largest Lyapunov exponent using Wolf's algorithm

In this work, the largest Lyapunov exponent, λ_1, is calculated as a measure of the chaotic behavior of the system using the Wolf algorithm [Wolf]. Consider two trajectories with nearby initial conditions on an attracting manifold. When the attractor is chaotic, the trajectories diverge, on the average, at an exponential rate characterized by the largest Lyapunov exponent λ_1. The algorithm used is as follows,

1. Compute the distance d_0 of two, very close, points in the reconstructed phase space orbit.

2. Follow both points as they travel a short distance along the orbit. The distance d_1 between them is calculated.

3. If d_1 becomes too large, one of the points is kept and an appropriate replacement for the other point is chosen.

4. The two points are now allowed to evolve again following steps 1-3.

5. After s propagation steps, the largest Lyapunov exponent λ_1 is estimated as,

$$\lambda_1 = \frac{1}{t_s - t_0} \sum_{k=1}^{s} \log_2 \left(\frac{d_1(t_k)}{d_0(t_{k-1})} \right) \tag{3.9}$$

3.4.2 Practical Implementation

We used a software implementation of Wolf's algorithm [Wolf]. This software is divided into two programs: BASGEN and FET. BASGEN stands for "dataBASe GENerator" and is considered as a preprocessing step for the main program FET. It generates a database that is used by FET to determine the closest points to any specific point. FET stands for "Fixed Evolution Time," it does the main job of calculating the average exponential rate of divergence of short segments of the reconstructed orbit.

There are a lot of parameters that need to be defined for the two programs. The used parameters in our work are presented in the following two subsections.

3.4.2.1 BASGEN Parameters

- Embedding dimension m=4. It should be noted that for Lyapunov exponent calculations, the embedding dimension m was chosen as D_2 rounded to the next highest integer [Govindan *et al.* 1998].
- Time delay L=60 for sampling rate of 360 Hz and L=40 for sampling rate of 250 Hz.
- Grid resolution ires= 20. The grid resolution refers to the fact that BASGEN places the reconstructed data into a grid of dimension m, with a resolution of ires cells per side. This grid will be used later by FET to efficiently find nearest neighbors to any point.

3.4.2.2 FET Parameters

- The time step was chosen as the sampling period
  - 2.777 msec for the all types that sampled at 360 Hz (except ventricular fibrillation).
  - 4 msec for the ventricular fibrillation signals (sampled at 250 Hz).

- The evolution time (*evolve*) was chosen as 25. That is, each pair of points is followed through the phase space for this number of steps at which the local contribution to orbital divergence is calculated, and a replacement is attempted if necessary.

- The minimum separation at replacement (*dismin*) was selected to be 0.01. When a replacement is decided, points whose distance from the kept point is less than *dismin* are rejected.

- The maximum separation at replacement (*dismax*) determines the distance

between the pair of points beyond which a replacement is decided and was chosen as 15% of the data range.

- The maximum orientation error (*thmax*) is selected to be 30 to define the maximum allowed angular deviation from the identical orientation between two points when a replacement is decided.

Chapter 4

Blind Source Separation Techniques (PCA and ICA)

4.1 Introduction

A “source” means here an original signal, “Blind” means that we know very little, if any, of the mixing matrix, and make general assumptions on the source signals. Independent component analysis (ICA) is one method, perhaps the most widely used, for performing blind source separation. ICA can be recognized as the generalization of principal component analysis (PCA). ICA extracts higher-order relationships among data, whilst PCA only looks at second order statistics as shown in this chapter.

Blind source separation assumes that the acquired signal is composed of a weighted sum of a number of basic components corresponding to a number of limited sources. In this work, we pose the problem of ECG signal diagnosis in the form of a blind source separation problem. In particular, ECG signals undergo two of the most commonly used blind source separation techniques; namely, principal component

analysis (PCA) and independent component analysis (ICA), in order to identify the basic components underlying this complex signal. Given that such techniques are sensitive to signal shift, we utilize a simple transformation that computes the magnitude of the Fourier transformation of ECG signals. This allows the phase components corresponding to such shifts to be removed. This issue is discussed in more details in section 4.4. But first, let us investigate PCA and ICA in the following two sections; 4.2 and 4.3.

4.2 Principal Component Analysis (PCA)

Principal component analysis is analogous to Fourier analysis in that the data is described in terms of the coefficients of a predetermined orthogonal set. But rather than using sines and cosines, the orthogonal set is chosen so that it describes the samples from the study population most efficiently, with the smallest number of terms. Principal component analysis is an efficient technique for dimensionality reduction in multivariate statistical analysis. Multivariate statistical analysis deals with the analysis of the data that consist of measurements on a number of individuals or objects [Gerbrands1981]. PCA derives the direction of a set of orthogonal vectors that point into the direction of the highest variance of the data set. The principal components (PCs) are calculated as the eigenvectors of the covariance matrix of the data set. The eigenvalues denote the variance that corresponding PCs (i.e. eigenvectors) account for.

Furthermore, PCA allows capturing the most important variations in the data set by keeping only a few of the high-variance principal components. By such unsupervised learning, hidden linear relations among the original set of measured variables are discovered. Conventionally learning problems are divided in supervised and unsupervised learning. Supervised learning concerns the identification of functional relationships between two or more variables as in, e.g., linear regression. The objective of PCA and other unsupervised learning schemes is to capture statistical relationships, i.e., the structure of the underlying data distribution.

4.2.1 Geometrical Meaning of Principal Components

Geometrical interpretation of principal components represents a good introduction to the derivations of principal components. We start this analysis by a simple three-dimensional problem and then it will be generalized to n-dimensions. Imagine that a sample of N trivariate observations has the scatter plot shown in fig. 4.1, where the origin has been taken at the sample means. The swarm of points seems to have a generally ellipsoidal shape, with a major axis Y_1 and less well defined minor axes Y_2 and Y_3. Let us confine our attention for the moment to the major axis and

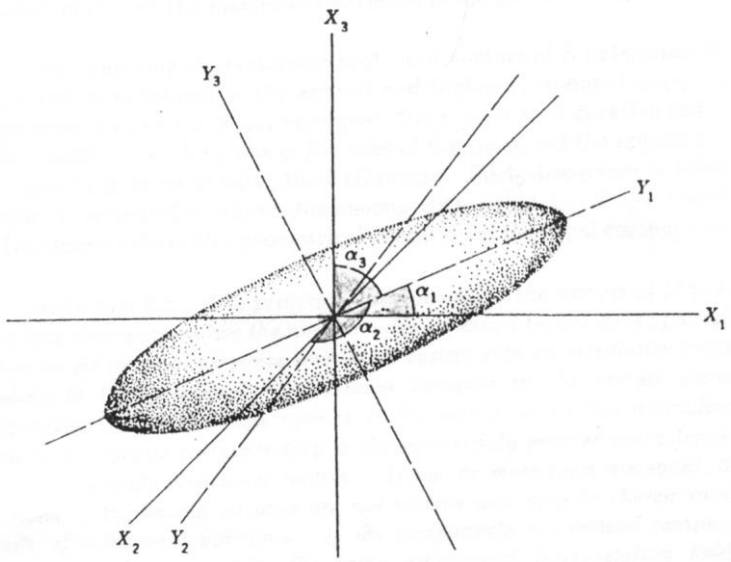

Figure 4.1 Principal axes of trivariate observations.

denote its angles with the original axes as α_1 α_2 α_3. If Y_1 passes through the sample mean point, its orientation is completely determined by the direction cosines

$$a_{11} = \cos\alpha_1, a_{21} = \cos\alpha_2, a_{31} = \cos\alpha_3, a_{11}^2 + a_{21}^2 + a_{31}^2 = 1 \qquad (4.1)$$

It is known from the analytic geometry that the value of the observation $[x_{i1},x_{i2},x_{i3}]$ on the new coordinate axis Y_1 will be

$$y_{i1} = a_{11}(x_{i1} - \bar{x}_1) + a_{21}(x_{i2} - \bar{x}_2) + a_{31}(x_{i3} - \bar{x}_3) \qquad (4.2)$$

Note that the mean of the Y_1 variate is

$$\bar{y}_1 = \frac{1}{N}\sum_{i=1}^{N} y_{i1} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{3} a_{j1}(x_{ij} - \bar{x}_j) = 0 \tag{4.3}$$

Let us define the major axis of a scatter of points as passing through the direction of maximum variance in the points. In the present case of three variates the variance is

$$s_{Y_1}^2 = \frac{1}{N\text{-}1}\sum_{i=1}^{N} y_{i1}^2 = \frac{1}{N\text{-}1}\sum_{i=1}^{N}\left[\sum_{j=1}^{3} a_{j1}(x_{ij} - \bar{x}_j)\right]^2 \tag{4.4}$$

and the angles of Y_1 would be found by differentiating this expression with respect to the a_{j1} (with suitable provision for the constraint) and solving for the a_{j1} which make the derivatives zero. The solution would be the characteristic vector of the greatest root of the sample covariance matrix of the x_{ij}, and Y_1 would be the first principal component of the system.

4.2.2 Principal Component Analysis in p-dimensional system

Let us prove the statement of the previous section for the general case of p variates. Write the direction cosines of the first principal component axis as $a_1^T = [a_{11}, ..., a_{p1}]$, Where the constraint

$$a_1^T a_1 = 1 \tag{4.5}$$

must always be satisfied. The variance of the projections on the Y_1 axis is

$$s_{Y_1}^2 = \frac{1}{N-1}\sum_{i=1}^{N} y_{i1}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[\sum_{j=1}^{p} a_{j1}(x_{ij} - \bar{x}_j)\right]^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[(x_i - \bar{x})^T a_1\right]^2$$

$$= \frac{1}{N-1}\sum_{i=1}^{N} a_1^T (x_i - \bar{x})(x_i - \bar{x})^T a_1 = a_1^T S a_1 \tag{4.6}$$

Introduce the constraint (4.5) by the Lagrange multiplier l_1, then differentiate with respect to a_1

$$\frac{\partial}{\partial a_1}[\mathrm{s}^2_{\mathrm{Y}_1} + l_1(1 - a_1^T a_1)] = \frac{\partial}{\partial a_1}[a_1^T \mathrm{S} a_1 + l_1(1 - a_1^T a_1)] = 2(\mathrm{S} - l_1\mathrm{I})a_1 \qquad (4.7)$$

The coefficients must satisfy the p simultaneous linear equations

$$(\mathrm{S} - l_1\mathrm{I})a_1 = 0 \qquad (4.8)$$

If the solution to these equations is to be other the null vector, the value of l_1 must be chosen so that

$$|\mathrm{S} - l_1\mathrm{I}| = 0 \qquad (4.9)$$

l_1 is thus a characteristic root of the covariance matrix, and a_1 is the associated characteristic root [Watkins 1991]. To determine which of the p roots should be used, premultiply the system of equations (4.8) by a_1^T . Since $a_1^T a_1 = 1$, it follows that

$$l_1 = a_1^T \mathrm{S} a_1 = \mathrm{s}^2_{\mathrm{Y}_1} \qquad (4.10)$$

But the coefficient vector, a_1, was chosen to maximize this variance, $\mathrm{s}^2_{\mathrm{Y}_1}$. So, l_1 must be the greatest characteristic root of S.

The second principal component is that linear compound

$$\mathrm{Y}_2 = \mathrm{a}_{12}\mathrm{X}_1 + \cdots + \mathrm{a}_{\mathrm{p}2}\mathrm{X}_\mathrm{p} \qquad (4.11)$$

whose coefficients have been chosen, subject to the constraints

$$a_2^T a_2 = 1, \quad a_1^T a_2 = 0 \qquad (4.12)$$

so that the variance of Y_2 is a maximum. The first constraint is merely a scaling to assure the uniqueness of the coefficients, while the second requires that a_1 and a_2 be orthogonal. The immediate consequence of the orthogonality is that the variances of the successive components sum to the total variance of the system.

The coefficients of the second component are found by introducing the constraints (4.12) by the Lagrange multipliers l_2 and μ and differentiating with respect to a_2:

$$\frac{\partial}{\partial a_2}[a_2^T \mathrm{S} a_2 + l_2(1 - a_2^T a_2) + \mu a_1^T a_2] = 2(\mathrm{S} - l_2 \mathrm{I})a_2 + \mu \mathrm{a}_1 \tag{4.13}$$

If the right-hand side is set equal to 0 and premultiplied by a_1^T, it follows from the normalization and orthogonality conditions that

$$2a_1^T \mathrm{S} a_2 + \mu \tag{4.14}$$

Similar premultiplication of the equations (4.8) by a_2^T implies that

$$a_1^T \mathrm{S} a_2 = 0 \tag{4.15}$$

and hence $\mu = 0$. The second vector must satisfy

$$(\mathrm{S} - l_2 \mathrm{I})a_2 = 0 \tag{4.16}$$

And it follows that the coefficients of the second component are thus the elements of the characteristic vector corresponding to the second greatest characteristic root. The remaining principal components are found in their turn from the other characteristic vectors. Let us summarize the process:

The j^{th} principal component of the sample of p-variate observations is the linear compound

$$Y_j = a_{1j}X_1 + ... + a_{pj}X_p \tag{4.17}$$

whose coefficients are the elements of the characteristic vector of the sample covariance matrix S corresponding to the jth largest characteristic root (l_j). The sample variance of the jth component is l_j and the total system variance is thus

$$l_1 + \cdots + l_p = \text{tr } S \tag{4.18}$$

The importance of the jth component is $\frac{l_j}{\text{tr } S}$. More discussion of the "importance" of principal components is presented in the subsection 4.2.4.

4.2.3 Implementation

Given a n×m data matrix X of m samples and a centered matrix Z [$Z = (X - E\{X\})$] where E{X} is the matrix of mean vectors, the principal component transform of X is defined in

$$Y = B^T Z = B^T (X - E\{X\}) \tag{4.19}$$

with B defined by

$$K_z = B\Lambda B^T, \quad \Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}. \tag{4.20}$$

where Λ is a diagonal matrix of the eigenvalues, $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_n$, of the covariance matrix K_z and the columns of B are the corresponding eigenvectors.

However, since

$$K_z = E\{ZZ^T\} = E\{(X - E\{X\})(X - E\{X\})^T\} = K_x \tag{4.21}$$

Finally, the principal component analysis is performed by the following transform

$$Y = B^T(X - E\{X\}) \tag{4.22}$$

$$K_x = B\Lambda B^T \tag{4.23}$$

The output of the PCA transform is uncorrelated vectors. The covariance matrix of the output Y is

$$K_y = E\{YY^T\} = \Lambda \tag{4.24}$$

To summarize, the methodology consists of two distinct stages, that of defining the principal components (PCs) of the study population, and that of calculating the coefficients of the PCs for each sample:

1. Calculate the sample mean vector SMR

$$SMR_i = \sum_{n=1}^{N} x_{in} / N \tag{4.25}$$

   where SMR_i is the i^{th} element of the SMR, x_{in} is the i^{th} element of the n^{th} waveform (observation), and N is the total number of waveforms in the sample.

2. Calculate the estimate of the population covariance matrix. Each element of the covariance matrix can be written:

$$K_{ij} = \sum_{n=1}^{N} (x_{in} - SMR_i)(x_{jn} - SMR_j) / (N-1) \tag{4.26}$$

3. Get the eigenvalues and the corresponding eigenvectors of the covariance matrix. The first principal component is the eigenvector corresponding to the largest eigenvalue and so on.

4. Choose only first “q” principal components (see subsection 4.2.4).

5. Once the PCs have been defined, the calculation of their coefficients for each test waveform is straightforward. If y_k is the coefficient of the k_{th} PC, then:

$$y_k = \sum_{i=1}^{I} (x_i - SMR_i)(b_{ki}) \tag{4.27}$$

where x_i is the i^{th} element of the test waveform and b_{ki} is the i^{th} element of the k^{th} PC.

4.2.4 Dimensionality Reduction

A question can be raised is what is the utility of this artificial variates constructed from the original data (principal components)? In the extreme case of Z of rank one, the first principal component would explain all the variation in the multivariate system. In the more usual case of the data matrix of full rank, the importance and usefulness of the component would be measured by the proportion of the total variance attributable to it. For example, if 85% of the variation in a system of six variables could be accounted for by a simple weighted average of the response values, it would appear that almost all the variation could be expressed along a single continuum rather than in six-dimensional space. The coefficients of the six responses would indicate the relative importance of each original variate in the new derived component.

Due to orthogonality of the matrix B, equation (4.19) can be rewritten as,

$$Z = BY \tag{4.28}$$

where the matrices are n×m, n×n, and n×m, respectively. In general, principal components analysis results in a dimensionality reduction, to be described as

$$Z' = B'Y' \tag{4.29}$$

where the matrices are n×m, n×q, and q×m, respectively and where the approximation Z' contains as much of the 'variability' of the original data as needed. The efficiency

of this approximation can be estimated by the ratio between the chosen variance to the total system variance

$$E = \sum_{i=1}^{q} \lambda_i \Big/ \sum_{i=1}^{n} \lambda_i \tag{4.30}$$

For feature extraction purposes, each centered sample is represented by its projections on the q PCs.

4.3 Independent Component Analysis (ICA)

Independent component analysis (ICA) is a method for finding underlying factors or components from multivariate (multidimensional) statistical data [Hyvärinen *et al.* 2001]. Independent component analysis (ICA) is an extension of PCA; in ICA higher order statistics rather than second order moments are used to determine base vectors that are statistically as independent as possible. What also distinguishes ICA from other methods is that it looks for components that are both statistically independent and gaussian. In other words, Independent component analysis is a signal processing technique whose goal is to express a set of random variables as linear combinations of statistically independent component variables [Hyvärinen 1999, Hyvärinen & Oja 2000]. ICA has applications in different fields, for example, among many others, face recognition [Bartlett *et al.* 1998].

4.3.1 Definition of Independent Component Analysis

Independent Component Analysis (ICA) defines a model for the multivariate data. The measured n random variables (x_i, i=1,…,n) are modeled as linear combinations of m random variables (s_j, j=1,…,m)

$$x_i = a_{i1}s_1 + a_{i2}s_2 + ... + a_{im}s_m, i = 1, ..., n \tag{4.31}$$

where the a_{ij}, i,j = 1, … ,n are some real coefficients. By definition, the s_i are statistically mutually independent. One restriction for the problem is naturally $m \leq n$; usually m is assumed known and often m = n.

This is the basic ICA model. The ICA model is a generative model, which means that it describes how the observed data are generated by a process of mixing the components s_j. The independent components s_j (ICs) are latent variables, meaning that they cannot be directly observed. Also the mixing coefficients a_{ij} are assumed to be unknown. All we observe are the random variables x_i, and we must estimate both the mixing coefficients a_{ij} and the ICs s_i using x_i. This must be done under as general assumptions as possible.

It is usually more convenient to use vector-matrix notation instead of the sums as equation (4.31). Let us denote by x the random vector whose elements are the mixtures x_1, …. , x_n, and s the random vector with elements s_1, …. , s_n. Let us denote by A the matrix with elements a_{ij}. So, the mixing model is rewritten as

$$x = As \tag{4.32}$$

4.3.2 Assumptions in ICA

To make sure that the basic ICA model can be estimated, some assumptions have to be taken care of. First, the independent components are assumed statistically independent. This is the principle on which ICA rests. Basically, random variables y_1, y_2, … , y_n are said to be independent if information on the value of y_i does not give any information on the value of y_j for $i \neq j$. Technically, independence can be defined by the probability densities. Let us denote by $p(y_1, y_2, \dots, y_n)$ the joint probability density function (pdf) of the y_i, and by $p_i(y_i)$ the marginal pdf of y_i, i.e., the pdf of y_i when it is considered alone. Then the y_i are independent if and only if the joined pdf is factorizable:

$$p(y_1, y_2, \dots, y_n) = p_1(y_1)\, p_2(y_2) \dots p_n(y_n) \tag{4.33}$$

Nongaussianity is another assumption essential for independent component analysis. The independent components must have nongaussian distributions. The higher-order information is essential for estimation of the ICA model, but it is well known that higher-order cumulants are zero for gaussian distributions. More details in this issue will be seen later in this chapter.

4.3.3 Preprocessing

4.3.3.1 Centering

The first preprocessing step is centering the observable variables, i.e., subtracting their sample mean. This means that the original mixtures, say x' are preprocessed by

$$x = x' - E\{x'\} \tag{4.34}$$

before doing ICA. Thus the independent components are made zero mean as well, since

$$E\{s\} = A^{-1}E\{x\} \tag{4.35}$$

The mixing matrix, on the other hand, remains the same after this preprocessing step. So we can do this without affecting the estimation of the mixing matrix. After estimating the mixing matrix and the independent components for the zero-mean data, the subtracted mean can be simply reconstructed by adding $A^{-1}E\{x\}$ to the zero-mean independent components.

4.3.3.2 Whitening

The ICA problem is greatly simplified if the observed mixture vectors are first whitened or sphered. A zero random vector $z = (z_1 ... z_n)^T$ is said to be white if its elements z_i are uncorrelated and have unit variances. In terms of the covariance matrix, this means that $E\{zz^T\} = I$.

Because whitening is essentially decorrelation followed by scaling, the technique of PCA can be used. Given a random vector x with n elements, apply whitening by finding a linear transformation V such that the vector z is white:

$$z = Vx = VAs = \tilde{A}s \tag{4.36}$$

The problem has a straightforward solution in terms of the PCA expansion. Let $E = (e_1 ... e_n)$ be the matrix whose columns are the unit-norm eigenvectors of the covariance matrix $C_x = E\{xx^T\}$. Let $D = \mathrm{diag}(d_1 ... d_n)$ be the diagonal matrix of the eigenvalues of C. Then a linear transform is given by

$$V = D^{-1/2}E^T \tag{4.37}$$

The matrix V is indeed a whitening transformation as explained next. The covariance matrix C_x can be written in terms of the eigenvector and eigenvalue matrices E and D as $C_x = EDE^T$, with E an orthogonal matrix satisfying $EE^T = E^TE = I$, it holds:

$$E\{zz^T\} = VE\{xx^T\}V^T = D^{-1/2}E^TEDE^TED^{-1/2} = I \tag{4.38}$$

The covariance of z is the unit matrix, hence z is white.

The utility of whitening resides in the fact that the new mixing matrix $\tilde{A} = VA$ is orthogonal. This can be seen from

$$E\{zz^T\} = \tilde{A}E\{ss^T\}\tilde{A}^T = \tilde{A}\tilde{A}^T = I \tag{4.39}$$

The problem of finding an arbitrary full-rank matrix A is reduced to the simpler problem of finding an orthogonal matrix.

On the other hand, applying PCA as a processor prior to ICA has another advantage. It is useful in reducing the dimension of the data, which is considered denoising to the data.

4.3.4 Nongaussianity

Usefulness of nongaussianity in ICA is motivated by the central limit theorem, which says that the distribution of a sum of independent and identically distributed random variables tends toward a gaussian distribution. A sum of two independent random variables usually has a distribution that is closer to gaussian than any of the original random variables.

The data vector is distributed according to the ICA model ($x = As$), i.e. it is a mixture of independent components. All the independent components are assumed to have identical distributions. The inversion of the ICA transform is $s = A^{-1}x$. Thus, to estimate one of the independent components, we can consider a linear combination of x_i. Let us denote this by $y = b^T x = \sum_i b_i x_i$, where b is a vector to be determined. Note that we also have $y = b^T As$. Thus y is a certain linear combination of the s_i, with coefficients given by $b^T A$. Let us denote this vector by q. Then we have

$$y = b^T x = q^T s = \sum_i q_i s_i \tag{4.40}$$

If b were one of the rows of the inverse of A, this linear combination $b^T x$ would actually equal one of the independent components. In that case, the corresponding q would be such that just one of its elements is 1 and all the others are zero.

So, how could the central limit theorem be used to determine b so that it would equal one of the rows A^{-1}? In practice, such a b cannot be determined exactly, because there is no knowledge of matrix A, but we can find an estimator that gives a good approximation. The fundamental idea used is that since a sum of even two independent random variables is more gaussian than the original variables, $y = q^T s$ is usually more gaussian than any of the s_i and becomes least gaussian when it in fact equals one of the s_i. In this case, obviously only one of the elements q_i of q is nonzero. We do not in practice know the values of q, but we do not need to, because $q^T s = b^T x$ by the definition of q. We only need to tune b and look at the distribution of $b^T x$.

Therefore, we could take the vector b that maximizes the nongaussianity of $b^T x$. Such a vector would necessarily correspond to a $q = b^T A$, which has only one nonzero element. This means that $y = b^T x = q^T s$ equals one of the independent components. Thus, maximizing the nongaussianity of $b^T x (q^T s)$ gives one of the independent components.

To summarize, ICA estimation is formulated as the search for directions that are maximally nongaussian; each local maximum gives one independent component. So we need a measure of nongaussianity and a technique to compute the values of b that (locally) maximize such a measure of nongaussianity. That will be explained in the following sections.

4.3.5 Kurtosis; A Measure of Nongaussianity

Nongaussianity is measured by many ways; most suggested solutions to the ICA problem use the fourth-order cumulant or kurtosis [Hyvärinen & Oja 1997]. The main reason is its simplicity, both computational and theoretical.

The kurtosis of a random variable y, denoted by kurt(y), is defined by

$$\text{kurt}(y) = E\{y^4\} - 3(E\{y^2\})^2 \tag{4.41}$$

If y is normalized so that its variance is equal to one: $E\{y^2\}=1$. Then the right hand side of equation (4.41) simplifies to $E\{y^4\}-3$. So, computationally, kurtosis can be estimated simply by using the fourth moment of the sample data. Theoretical analysis is simplified because of the following linearity property: if x_1 and x_2 are two independent random variables, it holds:

$$\text{kurt}(x_1 + x_2) = \text{kurt}(x_1) + \text{kurt}(x_2) \tag{4.42}$$

$$\text{kurt}(\alpha x_1) = \alpha^4 \text{kurt}(x_1) \tag{4.43}$$

where α is a constant.

For a gaussian variable y, the fourth moment ($E\{y^4\}$) equals $3(E\{y^2\})^2$. Thus, kurtosis is zero for gaussian random variables; for densities peaked at zero, it is positive, and negative for flatter densities. Typically nongaussianity is measured by the absolute value of kurtosis. The square can also be used. These measures are zero for a gaussian variable, and greater than zero for most nongaussian random variables.

4.3.6 Fast Fixed-Point Algorithm (FastICA)

As shown in the previous two sections, the procedure of estimating ICA is to maximize the absolute value of the kurtosis function of $y = b^T x = q^T s$ so that y will be one of the independent components.

The whitening-preprocessing step simplifies the ICA problem by estimating an orthogonal matrix $\tilde{A}$ instead of a full-rank matrix A (section 4.3.3.2). So, with the whitened data z, y will be $y = w^T z$ ($q = (VA)^T w$, and $b = V^T w$).

$$y = w^T z = w^T V x = w^T V A s = w^T \tilde{A} s \tag{4.44}$$

When reaching a solution, the columns of the mixing matrix $\tilde{A}$ are obtained as solutions for w, and the linear combination y itself will be one of the independent components.

The kurtosis of $w^T z$ is:

$$\text{kurt}(w^T z) = E\{(w^T z)^4\} - 3(E\{(w^T z)^2\})^2$$

$$= E\{(w^T z)^4\} - 3\|w\|^4 \tag{4.45}$$

where $E\{(w^T z)^2\} = \|w\|^2$ for whitened data.

So, the gradient of the absolute value of $\mathrm{kurt}(w^T z)$ is:

$$\frac{\partial \left|\mathrm{kurt}(w^T z)\right|}{\partial w} = 4\,\mathrm{sign}(\mathrm{kurt}(w^T z))\left[E\{z(w^T z)^3\} - 3w\|w\|^2\right] \tag{4.46}$$

under the constraint of $\|w\| = 1$.

To solve this problem, we followed a fast fixed-point algorithm (FastICA) that is used for finding the local extrema of the kurtosis function [Hyvärinen & Oja 1997]. FastICA is based on that at a stable point of the gradient algorithm, the gradient must point in the direction of w, i.e. the gradient must be equal to w multiplied by some scalar constant

$$w \alpha [E\{z(w^T z)^3\} - 3\|w\|^2 w] \tag{4.47}$$

This equation suggests a fixed-point algorithm where the iteration will be

$$w \leftarrow [E\{z(w^T z)^3\} - 3w] \tag{4.48}$$

After each iteration, w is divided by its norm to remain applying the constraint ($\|w\| = 1$). The final vector w gives one of the independent components as the linear combination $w^T z$. Note that convergence of the fixed-point iterations means that the old and new values of w point in the same direction.

FastICA algorithm for estimating one independent component can be summarized in:

1. Take a random initial vector w(0) of norm 1. Let k=1.
2. Let $w(k) = E\{z(w(k-1)^T z)^3\}$-3w(k-1).
3. Divide w(k) by its norm.
4. If $|w(k)^T w(k-1)|$ is not close enough to 1, let k=k+1 and go back to step 2. Otherwise, output the vector w(k).

Example of one-unit FastICA: consider a mixture of two uniformly distributed independent components. The mixtures are whitened; see the figure. Initialize w as $w(0) = (1,0)^T$. Running the FastICA iteration just two times, convergence is reached. In fig. 4.2, the obtained vectors of each iteration are shown. The dashed line gives the direction of w after the first iteration, and the solid line gives the direction of w after the second iteration. The third iteration did not significantly change the direction of w, which means that the algorithm converged. The figure shows that w may change drastically during the iteration, because the value w and –w are considered as equivalent. This is because the sign of the vector cannot be determined in the ICA model. The kurtoses of the projections w^Tz obtained in the iterations are plotted in Fig. 4.3, as a function of iteration count. It shows that the algorithm steadily increased the absolute value of the kurtosis of the projection, until it reached convergence at the third iteration.

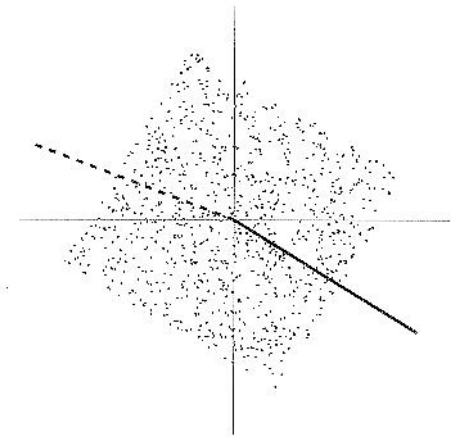

Figure 4.2 Result of FastICA using kurtosis, for ICs with uniform distributions. Dashed line: w after the first iteration. Solid line: w after the second iteration.

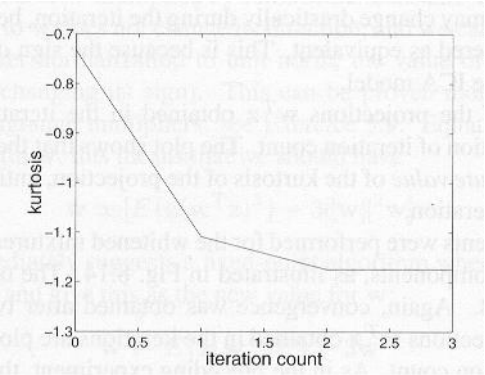

Figure 4.3 The convergence of FastICA using kurtosis, for ICs with uniform distributions. The value of kurtosis is shown as function of iteration count.

To estimate m independent components, simply run the algorithm m times. To ensure that a different independent component is estimated each time, an orthogonalizing projection is added to the iteration at step number 3.

$$w(k) = w(k) - \overline{A}\,\overline{A}^{T} w(k) \tag{4.49}$$

where $\overline{A}$ is a matrix whose columns are the previously found w's. At the end, the matrix $\tilde{A}$ is constructed from all w's and s is estimated by $s = \tilde{A}^{T} z$ where z is the whitened data. Modified step 3 of the algorithm to be:

3. Let $w(k) = w(k) - \overline{A}\,\overline{A}^{T} w(k)$. Divide w(k) by its norm.Let us now see an example of FastICA running. Figure 4.4 shows original source signals. Notice the dependence of the fifth and the sixth signals on the first one and of the seventh signal on the second one. The output of mixing them with an arbitrary matrix is shown in figure 4.5. Applying FastICA on those mixed signals gives the estimation of the underlying independent components shown in figure 4.6.

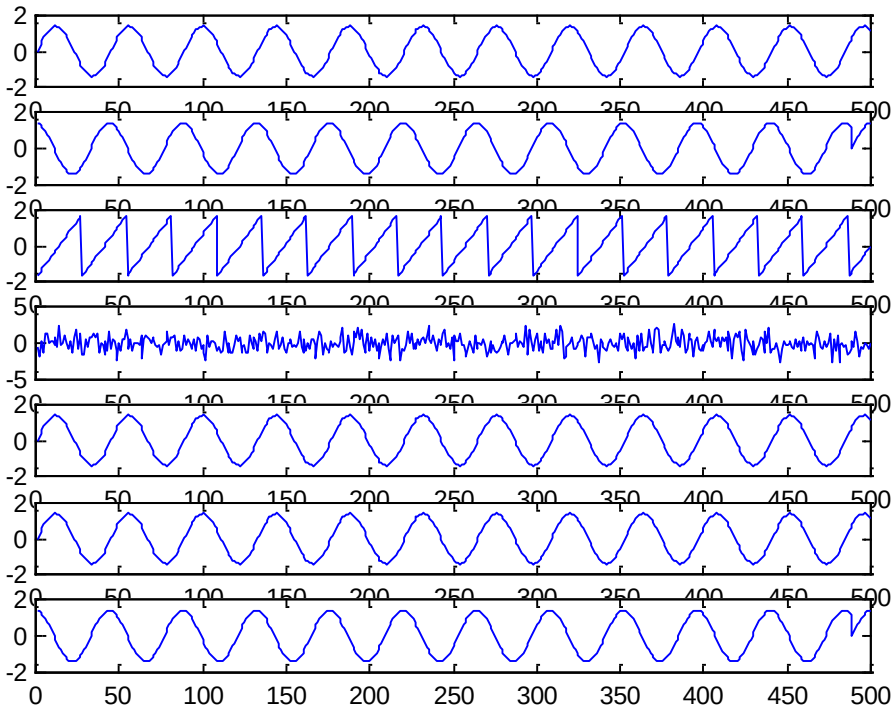

Figure 4.4 Original seven source signals.

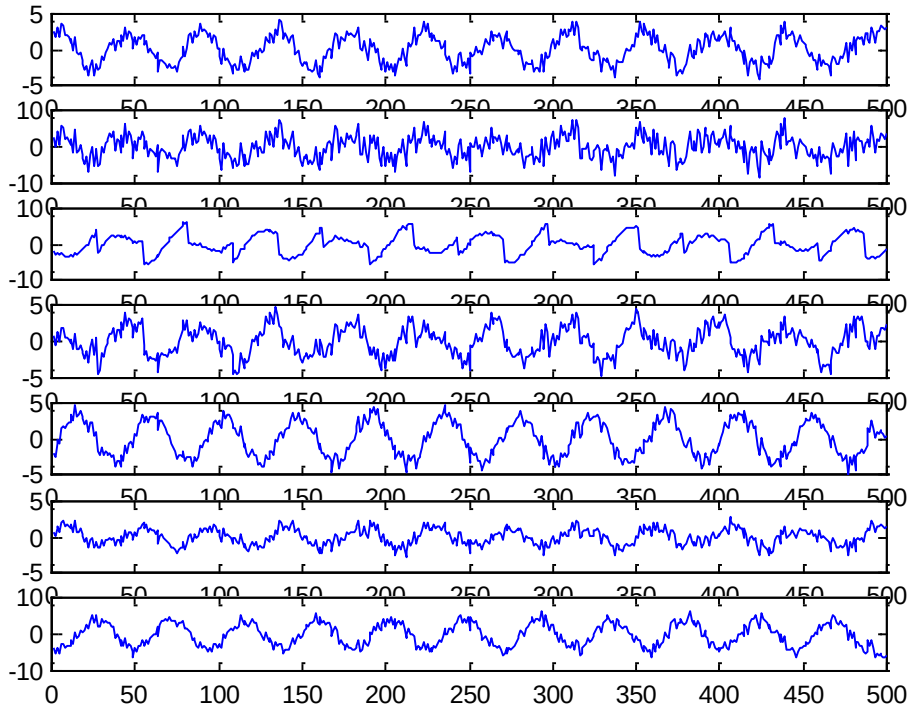

Figure 4.5 Input of FastICA; 7 mixtures of the original source signals.

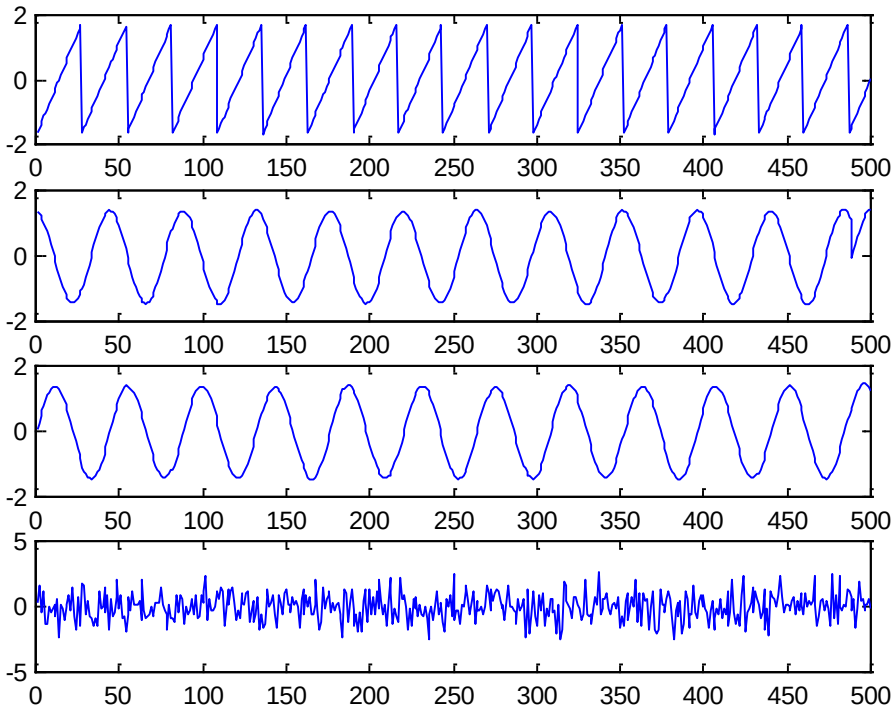

Figure 4.6 Output of FastICA; four independent components.

4.4 Shift Invariance Transformation

The detection of ECG arrhythmia type relies on observing changes in the ECG signal characteristics as computed from a short window of the signal. This window is generally taken as a moving window that covers a number of seconds of signal starting from the current sample and back in time. This results in unknown phase shift to the ECG signal within the sample window. Moreover, even if the signal window is synchronized to the onset of an R point, the heart rate variability prevents the direct comparison of windows obtained from different patients or at different times for the same patient. Given the sensitivity of PCA/ICA based techniques to the presence of such practical conditions, their direct use in the analysis of ECG signals has been limited. In order to overcome this limitation, we propose to apply a simple transformation whereby the magnitude of the Fourier transformation of the signal is used instead of the time domain signal. Since the relative delay between samples are

manifested as linear phase in the Frequency domain, the proposed transformation yields the same result for all circular shifts of a given signal. Given the periodic nature of the ECG signal, windows with different starting points generally approximate circular shifts of the same signal. As a result, such windows provide similar outputs after this transformation. This alleviates the need for a reference point and allows flexible choice of sample windows.

Chapter 5

Significance Test

5.1 Introduction

The reasoning of significance tests has appealed to researchers in many fields, so that the tests are widely used to report research results. In this setting H_a is a "research hypothesis" asserting that some effect or difference is present. The "null hypothesis" H_0 says that there is no effect or no difference [Moore & McCabe 1993].

We use significance test in our work to assess the use of the new parameters from nonlinear dynamical modeling (D_2 and λ_1) for discriminating between the different ECG signal types. This is considered a two-sample problem, which is defined as comparing random samples separately selected from two populations (for the same measured variable in the two samples) by comparing the two population means through testing the hypothesis of no difference, H_0: $\mu_1 = \mu_2$.

An overview of hypothesis testing is presented in section 5.2. In sections 5.3 and 5.4 t-test is presented, which we use in our results. T-test is considered as the most commonly used method to evaluate the differences in means between two groups.

5.2 Hypothesis Testing

Hypothesis tests address the uncertainty of the sample estimate [Papoulis 1991]. A hypothesis test attempts to refute a specific claim about a population parameter based on the sample data. In our case, the hypothesis is: the means from two populations are equal.

To reject a hypothesis is to conclude that it is false. However, to accept a hypothesis does not mean that it is true, only that we do not have evidence to believe otherwise. Thus hypothesis tests are usually stated in terms of both a condition that is doubted (null hypothesis) and a condition that is believed (alternative hypothesis).

A common format for a hypothesis test is:

- H_0: A statement of the null hypothesis, e.g., two population means are equal.

- H_a: A statement of the alternative hypothesis, e.g., two population means are not equal.

- Significance Level: The significance level, α, defines the sensitivity of the test. A value of $\alpha = 0.05$ means that we reject the null hypothesis 5% of the time when it is in fact true. The choice of α is somewhat arbitrary, although in practice values of 0.1, 0.05, and 0.01 are commonly used.

- Critical Region: The critical region encompasses those values of the test statistic that lead to a rejection of the null hypothesis. Based on the distribution of the test statistic and the significance level, a cut-off value for the test statistic is computed. Values either above or below or both

(depending on the direction of the test). This cut-off defines the critical region.

5.3 T-Test

Student t statistic is the most common used test in the context of hypothesis testing. A t-test is any of a number of tests based on the t distribution. The t distribution is used instead of the normal distribution whenever the standard deviation is estimated (from a random sample of the population). The t distribution has relatively more scores in its tails than does the normal distribution. The shape of the t distribution depends on the degrees of freedom (df) that went into the estimate of the standard deviation. With 100 or more degrees of freedom, the t distribution is almost indistinguishable from the normal distribution. As the degrees of freedom increases, the t distribution approaches the normal distribution. Estimates of parameters can be based upon different amounts of information. The number of independent pieces of information that go into the estimate of a parameter is called the degrees of freedom (df). In general, the degrees of freedom of an estimate is equal to the number of independent scores that go into the estimate minus the number of parameters estimated as intermediate steps in the estimation of the parameter itself. For example, if the variance, σ^2, is to be estimated from a random sample of N independent scores, then the degrees of freedom is equal to the number of independent scores (N) minus the number of parameters estimated as intermediate steps (one, μ is estimated by M) and is therefore equal to N-1.

For example, to contain 95% of the t distribution with 4 df, the interval must extend 2.78 estimated standard deviations from the mean in both directions. Compare this to the normal distribution for which the interval need only extend 1.96 standard deviations in both directions. Figure 5.1 shows t distributions with 1, 4, and 15 degrees of freedom. Areas greater than +2 and less than -2 are shaded. This figure shows that the t distribution with 1 df has the least area in the middle of the distribution and the greatest area in the tails.

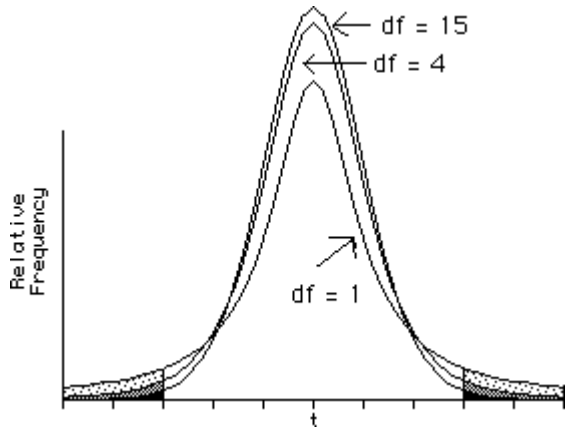

Figure 5.1 Example of a t distribution.

The most common t-test is a test for a difference between two means. The two-sample t-test is used to determine if two population means are equal. There are several variations on this test.

1. The data may either be paired or not paired. By paired, we mean that there is a one to one correspondence between the values in the two samples. That is, if X_1, X_2, ..., X_n and Y_1, Y_2, ... , and Y_n are the two samples, then X_i corresponds to Y_i. For paired samples, the difference X_i - Y_i is usually calculated. For unpaired samples, the sample sizes for the two samples may or may not be equal. The formulas for paired data are somewhat simpler than the formulas for unpaired data.

2. The variances of the two samples may be assumed to be equal or unequal. Equal variances yields somewhat simpler formulas, although with computers this is no longer a significant issue.

3. The null hypothesis is that the two population means are not equal ($\mu_1 \neq \mu_2$).

5.3.1 Two-sample pooled t-test

Assuming that the two normal population distributions have the same standard deviation σ, the two-sample t-test is called pooled t-test. That is what we use in our results and can be summarized in the following steps:

1. For the two samples, A and B, of sizes of N_1 and N_2 respectively, calculate the sample means $\overline{Y}_1$ and $\overline{Y}_2$, and the sample variances s_1^2 and s_2^2.

2. Estimate the variance of the source population as

$$s_p^2 = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \tag{5.1}$$

   The "source population" is the population of measures that the null hypothesis assumes to have been the common source of the measures in both groups.

3. Both sample variances s_1^2 and s_2^2 estimate σ^2. The best way to combine these two estimates is to average them with weights equal to their degrees of freedom. This gives more weight to the information from the larger sample, which is reasonable [Moore & McCabe 1993].

   So, the standard deviation of the sampling distribution of sample-mean differences (the "standard error" of $\overline{Y}_1 - \overline{Y}_2$) is estimated as

$$\sigma_{\overline{Y}_1 - \overline{Y}_2} = s_p \sqrt{1/N_1 + 1/N_2} \tag{5.2}$$

4. To test the hypothesis H_0: $\mu_1 = \mu_2$, compute the pooled two-sample t statistic as

$$T = \frac{\overline{Y}_1 - \overline{Y}_2}{\sigma_{\overline{Y}_1 - \overline{Y}_2}} \tag{5.3}$$

5. Refer the calculated value of t to the table of critical values of t, with df=(N_1-1)+(N_2-1).

6. Given the significance Level α, the null hypothesis that the two means are equal is rejected if $T < -t(\alpha/2, df)$, $T > t(\alpha/2, df)$ where $t(\alpha/2, df)$ is the critical value of the t distribution with df degrees of freedom where df = (N_1-1)+(N_2-1).

Chapter 6

Statistical Classifiers

6.1 Introduction

In order to investigate the performance of the proposed dynamical model parameters in classifying different arrhythmia types, we attempt to implement some of the most commonly used classifiers and use them to perform this task. Generally speaking, statistical classifiers can be divided into parametric and non-parametric techniques. Parametric statistical pattern recognition uses given or assumed information about the prior probabilities to obtain the classification. For example, in the Bayes minimum-error classifier, one can assume a Gaussian *a priori* conditional probability densities for the classification process, and estimate the parameters of the Gaussian probability density function (i.e. the mean, covariance, and correlation coefficient) from the given samples [Fukunaga 1990]. On the other hand, the non-parametric approach does not require any *a priori* information about the probability distributions of the data in each class; classification is performed based on the provided data samples with known class membership. For example, in the voting k-nearest neighbor technique, classification is

made using the labeled design set only [Kadah *et al.* 1996]. Hence, this approach is less subjected to the bias effects due to incorrect assumptions or parameter changes than the former. In both techniques, the implementation of the classifiers requires the division of the available data into two subsets: a training (design) subset and a test subset. These subsets must be independent in order to eliminate any bias effects on the results. In this work, the available data was divided equally between the two subsets.

Before we describe the classifiers that were implemented in this study, we will define a set of statistics that will be referred to frequently in describing the results.

False-negative rate The probability that the classification result indicates a normal rhythm while the true diagnosis is indeed an abnormal rhythm (arrhythmia). This case should be completely avoided since it represents a danger to the patient.

False-positive rate The probability that the classification result indicates an arrhythmia while the true diagnosis is indeed a normal rhythm (i.e., negative). This case can be tolerated, but should be as infrequent as possible.

Sensitivity The conditional probability of detecting an abnormal rhythm while there is in fact an arrhythmia. By definition, the sensitivity is equal to 1 minus the false negative rate.

Specificity The conditional probability of detecting a normal rhythm while it is indeed normal. By definition, the specificity is equal to 1 minus the false positive rate.

6.2 Minimum Distance Classifier

This method assumes that the classes are similar in distribution and are linearly separable. Hence, the decision lines are allocated halfway between the centers of clusters of different classes. The algorithm to implement this method is as follows:

1. Group the design set into five supervised clusters according to their labels, representing the five types of interest.

2. Estimate the sample mean for each class by averaging the parameter set of the class.

3. Test samples are classified by assigning each to the class that has the nearest mean vector from its feature vector.

4. Error rate is estimated by the percentage of misclassified samples.

6.3 Bayes Minimum-Error Classifier

The Bayes decision rule classifies an observation (i.e., a test sample) to the class that has the highest *a posteriori* probability among the five classes [Kadah *et al.* 1996]. In this study, the data set is assumed to have a Gaussian conditional density function and the *a priori* probabilities are assumed to be equal for the five types. That is, the conditional density function is,

$$f_X(x \mid \omega_i) = \frac{1}{(2\pi)^{N/2}\left|\Sigma_i\right|^{1/2}} \cdot \exp\left[-\frac{1}{2}(x - M_i)^T \Sigma_i^{-1}(x - M_i)\right] \qquad (6.1)$$

and the a priori probabilities are assumed equal such that,

$$P(\omega_i) = \frac{1}{5}, i \in \{1,2,3,4,5\} \qquad (6.2)$$

Here, x is a 2×1 data sample from the random vector X, M_i is a 2×1 vector representing the sample mean of class i, Σ_i is a 2×2 matrix representing the covariance matrix of class i, and ω_i refers to class i. The Bayes decision rule is: choose class $j \in \{1,2,3,4,5\}$ if

$$f_X(x \mid \omega_i)P(\omega_j) = \max\{f_X(x \mid \omega_i)P(\omega_i) \mid i = 1,2,3,4,5\} \qquad (6.3)$$

6.4 Voting k-Nearest Neighbor (k-NN) Classifier

This technique is non-parametric and assigns a test sample to the class of the majority of its k-neighbors [Kadah *et al.* 1996]. That is, assuming the number of voting neighbors to be $k = \sum_{i=1}^{5} k_i$ (where k_i is the number of samples from class i in the k-sample neighborhood of the test sample), the test sample is assigned to class m if $k_m = \max\{k_i, i=1,2,3,4,5\}$. The algorithm used to implement this technique is described as follows:

1. Obtain and store the distances between the parameter sets of the test sample and all the samples in the design set.

2. Sort the obtained distance values in ascending order.

3. Consider the subset of the first k distances in the sorted array; i.e., the k nearest neighbors. Knowing the class membership of each of these samples, assign the test sample to the majority class in this subset if it exists, otherwise the result is considered inconclusive.

4. Estimate the error rate by comparing the classification results with actual class membership. Treat the special case of inconclusive decisions individually as a separate entity (i.e., neither an error nor a correct decision) and obtain its rate of occurrence.

Chapter 7

Results and Discussion

7.1 Data Collection

The lead-II ECG signals used in this work were obtained from the MIT-BIH arrhythmia database [MIT-BIH 1997]. The MIT-BIH Arrhythmia Database was the first generally available set of standard test material for evaluation of arrhythmia detectors, and it has been used for that purpose as well as for basic research into cardiac dynamics at about 500 sites worldwide since 1980 [Moody & Mark 2001].

Ventricular Fibrillation (VF) time series are sampled at 250 Hz while the other signals are sampled at 360 Hz. We extracted from the MIT-BIH database 3-sec signal windows. The signal window length can be arbitrarily chosen provided that it is less than 10 sec. This is to satisfy the ANSI/AAMI EC13-1992 standard, which requires alarms for abnormal ECG signals to be activated within 10 seconds of their onset. The use of two sampling rate variation of the number of points within this duration was not found to be crucial as long as the ECG signal is sufficiently sampled.

The initial step in statistical classifiers is to divide the data set into two independent subsets, design and test sets. The data set used in this research consists of 480 samples divided to:

1. 320 for design (learning) subset: 64 samples for each arrhythmia type (NR, VC, VT, VB, and VF).
2. 160 for testing subset: 32 samples for each arrhythmia type.

Results of our proposed features extracted from the ECG signals are presented in the following two sections. The results of applying the three statistical classifiers are presented for the two problems of detection arrhythmia (only normal and abnormal) and classification of the type of arrhythmia. In section 7.2, results of features based on nonlinear dynamical modeling (D_2 and λ_1) are demonstrated and in section 7.3 results of features based on blind source separation (PCA and ICA) are reported. The implementation was mainly using Matlab® [Hanselman & Littlefield 1998, Burrus *et al.* 1994].

7.2 Nonlinear Dynamical Modeling

The proposed features extracted from the ECG signals based on nonlinear dynamical modeling are correlation dimension (D_2) and the largest Lyapunov exponent (λ_1). The statistical mean and standard deviation of the computed parameters (D_2 and λ_1) for each type are computed as shown in table 7.1. From table 7.1, we observe non-integer correlation dimension D_2 values for all types indicating the presence of strange attractor. Also, the positive sign of λ_1 confirms the chaotic behavior of the ECG signal. The results generally support the hypothesis that cardiac electrical activity reflects a low-dimensional dynamic system behavior.

The resultant values are compared to detect statistically significant differences among different arrhythmia types. Then, three statistical classifiers are applied; minimum distance classifier, Bayes minimum-error classifier, and voting k-nearest

neighbor (k-NN) classifier. The results suggest the potential and robustness of using such features in ECG arrhythmia detection.

Type	Parameter	
	D_2	λ_1
NR	3.27 ± 0.42	8.18 ± 3.63
VC	2.54 ± 0.39	17.36 ± 3.68
VT	3.07 ± 0.52	13.55 ± 7.24
VB	2.71 ± 0.40	12.11 ± 5.08
VF	2.93 ± 0.71	13.20 ± 4.45

Table 7.1 Computed values for dynamical system features (mean ± standard deviation).

7.2.1 T-test

P-values of the t-test based on D_2 are shown in table 7.2. The P-values of the t-test based on λ_1 are shown in table 7.3. As indicated from Tables 7.2 and 7.3, the results confirm that normal ECG signals can be statistically differentiated from abnormal by both dynamical system features. The very low p-values suggest the rejection of the null hypothesis and hence the presence of a significant difference. For example, the first row of these tables show that normal ECG signals can be differentiated from all other arrhythmia types. On the other hand, these measures are not successful in discriminating between some of the abnormal signals. In particular, when using D_2, there is significant difference between all pairs at 5% level except between VB and VF, which are significant at the 10% level. Moreover, there was no statistically significant difference between VT and VF (shown in boldface inside the table). This may somewhat be explained by the presence of similarities in dynamics between these types. Given the common nature of VT and VF of producing higher heart rate, this might explain the similarity between them in their underlying dynamics. This is particularly apparent in their λ_1 values. Similarly for λ_1, it is not possible to find statistically significant difference between VT, VF, and VB (shown in boldface inside the table). The lack of separation between VB and both VT and VF in their λ_1 values can also be explained by the clinical observation that VB can lead to VT in some conditions [Wagner 2001]. Given that λ_1 values describe the sensitivity to the initial

condition, this explains the observed similarity in this domain. That is, VB eventually leads to the same chaotic behavior as VT and VF. These statistically insignificant differences represent the fundamental limitations of the dynamical features in differentiating between abnormal arrhythmia types.

Type	VC	VT	VB	VF
NR	<1.0e-16	0.0071	1.7e-14	0.0006
VC		1.96e-9	0.0148	0.0002
VT			3.01e-5	**0.2201**
VB				0.0309

Table 7.2 P-values of t-test for D_2.

Type	VC	VT	VB	VF
NR	<1.0e-16	4.7e-7	1.16e-6	1.38e-10
VC		2.7e-4	6.0e-10	5.82e-8
VT			**0.1929**	**0.7396**
VB				**0.1976**

Table 7.3 P-values of t-test for λ_1.

7.2.2 Classification Results

The classification results for only normal versus abnormal ECG are shown in Tables 7.4-7.7. The results of applying the three classifiers to classify the 5 different ECG types are listed in Tables 7.8-7.11. We have three different sets of results corresponding to applying the classifiers on three different feature vectors; each one of the parameters D2 and λ_1 at a time and then both of them are used together in the feature vector.

Even though the ECG signal classes are shown to be statistically different (with the exception of VT and VF), the classification results indicate that their distribution have significant overlap. Moreover, the classification rates for individual abnormalities appear very low. This suggests the practicality of using chaos-based features to detect the presence of abnormality rather than to specify the type of abnormality.

	D_2	λ_1	D_2 & λ_1
Specificity (%)	50.00	81.25	81.25
Sensitivity (%)	72.66	50.78	50.78

Table 7.4 Minimum distance classifier results for detection problem.

	D_2	λ_1	D_2 & λ_1
Specificity (%)	56.25	90.63	65.63
Sensitivity (%)	67.97	43.75	67.97

Table 7.5 Bayes minimum-error classifier results for detection problem.

	D_2		λ_1		D_2 & λ_1	
k	**Spec. (%)**	**Sens. (%)**	**Spec. (%)**	**Sens. (%)**	**Spec. (%)**	**Sens. (%)**
1	21.88	82.81	34.38	71.88	34.38	75.00
2	0.0	97.67	23.08	87.83	20.00	86.67
3	6.25	84.38	31.25	75.78	28.13	78.91
4	3.70	95.46	18.18	88.17	26.32	83.96
5	9.38	91.41	28.13	81.25	40.63	80.45
6	6.67	96.52	28.08	84.31	36.36	85.05
7	6.25	94.53	31.25	79.69	34.38	80.47
8	6.25	95.16	19.23	89.19	33.33	84.21
9	6.25	95.31	25.00	83.59	37.50	81.25
10	6.45	95.94	12.00	85.83	30.77	84.96
11	9.38	93.75	18.75	82.03	34.38	81.25
12	6.45	96.64	11.11	91.89	34.62	85.84

Table 7.6 Voting k-NN classifier results for detection problem.

	D_2		λ_1		D_2 & λ_1	
k	**NR (%)**	**AbNR (%)**	**NR (%)**	**AbNR (%)**	**NR (%)**	**AbNR (%)**
1	0.0	0.0	0.0	0.0	0.0	0.0
2	31.25	32.81	59.38	42.18	37.5	29.69
3	0.0	0.0	0.0	0.0	0.0	0.0
4	15.63	14.06	31.25	27.34	40.63	17.19
5	0.0	0.0	0.0	0.0	0.0	0.0
6	6.25	10.16	18.75	20.31	31.25	16.41
7	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	3.13	18.75	13.28	25.00	10.94
9	0.0	0.0	0.0	0.0	0.0	0.0
10	3.13	3.91	21.88	6.25	18.75	11.72
11	0.0	0.0	0.0	0.0	0.0	0.0
12	3.13	7.03	15.63	13.28	18.75	11.72

Table 7.7 Voting k-NN inconclusive rates for detection problem.

	D_2	λ_1	D_2 & λ_1
Specificity (%)	50.00	81.25	81.25
Sensitivity for VC (%)	62.50	9.38	9.38
Sensitivity for VT (%)	6.25	6.25	6.25
Sensitivity for VB (%)	31.25	15.63	15.63
Sensitivity for VF (%)	12.50	9.38	6.25

Table 7.8 Minimum distance classifier results for classification problem.

	D_2	λ_1	D_2 & λ_1
Specificity (%)	56.25	90.63	65.63
Sensitivity for VC (%)	65.63	15.63	25.00
Sensitivity for VT (%)	0.0	0.0	0.0
Sensitivity for VB (%)	56.25	0.0	31.25
Sensitivity for VF (%)	6.25	21.88	28.13

Table 7.9 Bayes minimum-error classifier results for classification problem.

	k	1	2	3	4	5	6	7	8	9	10	11	12
D_2	Spec. (%)	21.88	0.0	10.00	14.82	23.81	20.00	20.00	25.93	19.05	38.10	43.75	32.00
	Sens; VC (%)	25.00	66.67	31.25	34.62	45.00	42.11	50.00	56.52	52.00	60.00	54.55	53.57
	Sens; VT (%)	25.00	33.33	40.00	39.13	37.04	27.79	40.91	25.00	19.23	13.04	19.23	18.52
	Sens; VB (%)	28.13	40.00	41.67	38.46	32.00	41.18	30.00	19.05	34.78	30.00	32.00	38.10
	Sens; VF (%)	15.63	14.29	15.79	21.74	26.09	10.00	21.43	4.35	4.76	0.00	4.35	3.85
λ_1	Spec. (%)	34.38	100.0	47.62	45.83	56.52	45.83	68.00	62.5	60.87	68.00	65.39	59.26
	Sens; VC (%)	25.00	50.00	18.18	20.83	14.29	12.5	13.64	18.52	20.00	20.83	22.22	16.00
	Sens; VT (%)	9.38	0.0	4.55	8.00	8.33	0.0	0.0	11.54	8.00	10.00	7.69	7.14
	Sens; VB (%)	18.75	33.33	20.00	19.23	17.39	21.05	18.52	20.00	20.00	17.39	17.86	20.00
	Sens; VF (%)	9.38	0.0	8.70	5.00	5.00	4.17	4.00	12.5	4.76	3.85	4.77	14.29
D_2 & λ_1	Spec. (%)	34.38	33.33	42.86	52.00	53.85	59.26	50.00	57.14	53.57	53.57	61.54	62.96
	Sens; VC (%)	18.75	27.27	20.00	20.00	19.23	26.32	25.00	26.09	18.18	16.00	17.39	18.52
	Sens; VT (%)	25.00	0.0	0.0	0.0	0.0	0.0	3.57	3.85	3.33	3.85	3.45	6.67
	Sens; VB (%)	40.63	41.67	42.11	34.62	40.00	27.27	26.09	30.77	28.57	35.71	29.63	32.14
	Sens; VF (%)	34.38	18.18	25.00	20.00	21.74	13.64	17.39	16.67	17.86	19.23	15.39	14.82

Table 7.10 Voting k-NN classifier results for classification problem.

	k	1	2	3	4	5	6	7	8	9	10	11	12
D_2	Normal (%)	0.0	78.13	37.5	15.63	34.38	37.5	21.88	15.63	34.38	34.38	50.00	21.88
	VC (%)	0.0	81.25	50.00	18.75	37.5	40.63	18.75	28.13	21.88	37.50	31.25	12.50
	VT (%)	0.0	81.25	37.5	28.13	15.63	43.75	31.25	25.00	18.75	28.13	18.75	15.63
	VB (%)	0.0	68.75	25.00	18.75	21.88	46.88	37.50	34.38	28.13	37.50	21.88	34.38
	VF (%)	0.0	78.13	40.63	28.13	28.13	37.50	12.50	28.13	34.38	15.63	28.13	18.75
λ_1	Normal (%)	0.0	90.63	34.38	25.00	28.13	25.00	21.88	25.00	28.13	21.88	18.75	15.63
	VC (%)	0.0	81.25	31.25	25.00	34.38	25.00	31.25	15.63	21.88	25.00	15.63	21.88
	VT (%)	0.0	71.88	31.25	21.88	25.00	25.00	12.50	18.75	21.88	6.25	18.75	12.50
	VB (%)	0.0	81.25	37.5	18.75	28.13	40.63	15.63	21.88	21.88	28.13	12.50	21.88
	VF (%)	0.0	84.38	28.13	37.50	37.50	25.00	21.88	25.00	34.38	18.75	34.38	34.38
D_2 & λ_1	Normal (%)	0.0	62.5	34.38	21.88	18.75	15.63	18.75	12.5	12.50	12.5	18.75	15.63
	VC (%)	0.0	65.63	37.50	21.88	18.75	40.63	25.00	28.13	31.25	21.88	28.13	15.63
	VT (%)	0.0	78.13	31.25	21.88	12.50	28.13	12.50	18.75	6.250	18.75	9.38	6.25
	VB (%)	0.0	62.50	40.63	18.75	21.88	31.25	28.13	18.75	12.50	12.50	15.63	12.5
	VF (%)	0.0	65.63	25.00	21.88	28.13	31.25	28.13	25.00	12.50	18.75	18.75	15.63

Table 7.11 Voting k-NN inconclusive rates for classification problem.

7.3 Blind Source Separation (PCA & ICA)

The proposed BSS feature estimation techniques (PCA and ICA) were applied to the 3-second long signals in the training (design) data set to generate basis sets spanning the feature space. Each sample window in the testing data set (the testing subset consists of another 32 independent signals of the same length from each type) is projected on to this basis set and the magnitudes of these projections are used for classification purposes.

Given that such techniques (PCA and ICA) are sensitive to signal shift, we utilize a simple transformation that computes the magnitude of the Fourier transformation of ECG signals. This allows the phase components corresponding to such shifts to be removed. In our implementation, given the fact that ECG signals are real, the magnitude of the Fourier transformation of the sample windows is symmetric. Therefore, assuming the sample window length is N, the length of transformed window used in subsequent analysis is taken as N/2 after redundant samples are excluded. So, only the first 375 frequency components are used in the subsequent analysis (in order to obtain the same size for all signals). This is achieved by simply truncating the number of frequency components beyond 375 when the sampling rate is 360 samples/sec (in this case N = 3 sec × 360 samples/sec = 1080 and N/2 = 540). Given that the sampling window length in time is the same, the frequency bin size is the same between the different sampling rates and the above procedure is correct. The PCA and ICA procedures are applied to compute a total of 320 principal and independent components since the available data matrix will be 320×375 given the number of training vectors and the length of the feature vector.

7.3.1 Principal Component Analysis (PCA)

Table 7.12 demonstrates the energy packing property of the PCA, for each q (number of PCs), whereby a small percentage of all principal components account for most of the energy. Energy compaction, which leads to dimensionality reduction, is one of the main features of PCA. For feature extraction, each centered sample is represented by

its projections on the q principal components computed as the inner product between the centered sample and each of the computed eigenvectors. The results for both the detection and classification problems using PCA are shown in Tables 7.13 - 7.22 using different numbers of principal components to demonstrate the efficiency of the technique.

q	320	100	80	60	40	30	20	15	10	5
Efficiency	100	99.97	99.92	99.74	99.08	98.24	95.98	93.32	88.27	78.35

Table 7.12 Efficiency of reducing dimensionality using PCA

q	**Specificity(%)**	**Sensitivity(%)**
5	68.75	95.31
10	71.88	94.53
15 to 320	71.88	94.31

Table 7.13 Minimum distance classifier results for detection problem (PCA)

q	**Specificity(%)**	**Sensitivity(%)**
5	65.63	91.41
10	56.25	96.88
15	40.63	98.44
20	46.88	99.22
30	34.38	100
40	31.25	100

Table 7.14 Bayes minimum-error classifier results for detection problem (PCA)

q	**Specificity(%)**	**Sensitivity(%)**
5	59.38	94.53
10	75.00	92.97
15	90.63	96.09
20	93.75	96.09
30	96.88	96.09
40	96.88	96.88
60 to 320	96.88	97.66

Table 7.15 NN classifier (k-NN with k=1) results for detection problem (PCA).

q	k	1	2	3	4	5	6	7	8	9	10	11	12
5	Specificity (%)	59.38	64.00	62.50	69.23	71.88	74.19	71.88	76.67	75.00	75.00	75.00	75.00
	Sensitivity (%)	94.53	98.35	95.31	96.06	95.31	95.83	91.41	92.80	90.63	91.34	90.63	92.06
10	Specificity (%)	75.00	71.43	71.88	73.33	71.88	76.67	75.00	80.00	78.13	77.42	78.13	78.13
	Sensitivity (%)	92.97	98.32	92.97	94.35	93.75	94.44	93.75	95.12	92.97	93.70	92.97	93.65
15	Specificity (%)	90.63	93.10	90.63	93.10	87.50	87.50	87.50	87.10	87.50	86.21	81.25	81.25
	Sensitivity (%)	96.09	97.46	92.19	92.86	92.19	95.93	94.53	95.93	93.75	94.31	92.97	94.40
20	Specificity (%)	93.75	100.0	93.75	96.55	90.63	90.00	84.38	86.67	87.50	90.00	87.50	89.66
	Sensitivity (%)	96.09	97.52	93.75	94.40	94.53	96.00	95.31	95.16	92.97	94.35	92.19	93.60
30	Specificity (%)	96.88	100.0	96.88	96.55	87.50	93.33	87.50	90.00	90.63	93.10	90.63	92.86
	Sensitivity (%)	96.09	97.50	95.31	95.12	94.53	96.72	94.53	95.93	92.97	93.60	92.19	93.55
40	Specificity (%)	96.88	100.0	93.75	93.55	90.63	93.33	87.50	90.00	90.63	93.33	93.75	93.10
	Sensitivity (%)	96.88	97.48	94.53	95.16	93.75	96.75	95.31	95.97	92.97	92.91	92.97	93.55
60	Specificity (%)	96.88	100.0	93.75	93.55	90.63	90.32	87.50	87.10	87.50	90.32	90.63	93.10
	Sensitivity (%)	97.66	97.54	95.31	95.16	93.75	96.75	95.31	95.97	92.97	93.65	92.97	93.60
80 to 320	Specificity (%)	96.88	100.0	93.75	93.55	90.63	90.32	87.50	90.00	87.50	93.33	90.63	93.10
	Sensitivity (%)	97.66	97.54	95.31	95.16	93.75	96.75	95.31	95.97	92.97	93.70	92.97	93.60

Table 7.16 Voting k-NN classifier results for detection problem (PCA).

q	k	1	2	3	4	5	6	7	8	9	10	11	12
5	Normal (%)	0.0	21.88	0.0	18.75	0.0	3.13	0.0	6.25	0.0	0.00	0.0	0.00
	Abnormal (%)	0.0	5.47	0.0	0.78	0.0	6.25	0.0	2.34	0.0	0.78	0.0	1.56
10	Normal (%)	0.0	12.50	0.0	6.25	0.0	6.25	0.0	6.25	0.0	3.13	0.0	0.00
	Abnormal (%)	0.0	7.03	0.0	3.13	0.0	1.56	0.0	3.91	0.0	0.78	0.0	1.56
15	Normal (%)	0.0	9.38	0.0	9.38	0.0	0.0	0.0	3.13	0.0	9.38	0.0	0.00
	Abnormal (%)	0.0	7.81	0.0	1.56	0.0	3.91	0.0	3.91	0.0	3.91	0.0	2.34
20	Normal (%)	0.0	9.38	0.0	9.38	0.0	6.25	0.0	6.25	0.0	6.25	0.0	9.38
	Abnormal (%)	0.0	5.47	0.0	2.34	0.0	2.34	0.0	3.13	0.0	3.13	0.0	2.34
30	Normal (%)	0.0	15.63	0.0	9.38	0.0	6.25	0.0	6.25	0.0	9.38	0.0	12.50
	Abnormal (%)	0.0	6.25	0.0	3.91	0.0	4.69	0.0	3.91	0.0	2.34	0.0	3.13
40	Normal (%)	0.0	12.50	0.0	3.13	0.0	6.25	0.0	6.25	0.0	6.25	0.0	9.38
	Abnormal (%)	0.0	7.03	0.0	3.13	0.0	3.91	0.0	3.13	0.0	0.78	0.0	3.13
60	Normal (%)	0.0	9.38	0.0	3.13	0.0	3.13	0.0	3.13	0.0	3.13	0.0	9.38
	Abnormal (%)	0.0	4.69	0.0	3.13	0.0	3.91	0.0	3.13	0.0	1.56	0.0	2.34
80 to 320	Normal (%)	0.0	9.38	0.0	3.13	0.0	3.13	0.0	3.13	0.0	6.25	0.0	9.38
	Abnormal (%)	0.0	4.69	0.0	3.13	0.0	3.91	0.0	3.91	0.0	3.13	0.0	2.34

Table 7.17 Voting k-NN inconclusive rates for detection problem (PCA).

q	NR	VC	VT	VB	VF
5	68.75	68.75	53.13	21.88	78.13
10	71.88	71.88	56.25	25.00	81.25
15 to 320	71.88	75.00	56.25	25.00	81.25

Table 7.18 Minimum distance classifier sensitivity (%) results for classification problem (PCA).

q	NR	VC	VT	VB	VF
5	65.63	78.13	62.50	46.88	71.88
10	56.25	65.63	71.88	68.75	78.13
15	40.63	65.63	81.25	81.25	75.00
20	46.88	65.63	84.38	62.50	87.50
30	34.38	53.13	81.25	68.75	90.63
40	31.25	40.63	84.38	71.88	87.50

Table 7.19 Bayes minimum-error classifier sensitivity (%) results for classification problem (PCA).

q	NR	VC	VT	VB	VF
5	59.38	53.13	65.63	56.25	71.88
10	75.00	71.88	71.88	71.88	78.13
15	90.63	75.00	68.75	84.38	81.25
20	93.75	78.13	68.75	84.38	87.50
30	96.88	75.00	71.88	84.38	87.50
40	96.88	71.88	75.00	87.50	87.50
60	96.88	68.75	71.88	84.38	90.63
80	96.88	68.75	71.88	81.25	87.50
100 to 320	96.88	68.75	71.88	81.25	90.63

Table 7.20 NN classifier (k-NN with k=1) sensitivity (%) results for classification problem (PCA).

q	k	1	2	3	4	5	6	7	8	9	10	11	12
5	**NR**	59.38	84.21	71.43	82.61	80.00	83.33	83.33	78.13	78.13	80.65	77.42	78.13
	VC	53.13	63.16	61.29	60.00	59.26	53.33	55.17	66.67	59.38	62.07	56.67	56.67
	VT	65.63	86.36	76.67	75.86	72.41	67.74	68.75	66.67	65.63	65.63	67.74	64.52
	VB	56.25	66.67	65.52	60.71	63.33	58.06	61.29	59.38	58.06	61.29	56.67	54.84
	VF	71.88	88.00	80.00	85.71	79.31	73.33	81.48	75.86	79.31	76.67	75.00	77.42
10	**NR**	75.00	90.91	79.31	82.76	79.31	80.65	81.25	83.87	83.87	86.67	80.65	83.33
	VC	71.88	78.95	66.67	75.00	68.97	70.00	63.33	70.37	61.29	61.29	61.29	59.38
	VT	71.88	87.50	73.33	76.00	71.43	65.52	67.86	64.29	63.33	61.29	60.00	59.38
	VB	71.88	80.95	64.52	67.86	67.86	69.23	62.50	62.96	54.84	51.61	51.61	51.61
	VF	78.13	95.83	86.67	88.89	74.07	75.00	72.41	78.57	73.33	76.67	77.42	76.67
15	**NR**	90.63	100.0	96.67	93.55	93.33	96.55	90.32	90.32	93.33	90.32	90.63	96.67
	VC	75.00	82.61	74.19	81.48	79.31	80.77	75.00	78.26	66.67	60.00	62.07	62.07
	VT	68.75	82.61	64.52	66.67	70.00	63.33	62.50	62.07	56.25	58.06	54.84	54.85
	VB	84.38	95.00	76.67	72.41	66.67	64.29	66.67	67.86	63.33	59.38	61.29	60.00
	VF	81.25	86.36	75.86	73.33	70.00	78.57	75.86	78.57	74.19	75.00	75.00	77.42
20	**NR**	93.75	100.0	100.0	96.77	96.67	96.67	90.32	93.33	93.55	93.55	93.75	93.75
	VC	78.13	83.33	78.57	83.33	81.48	80.77	71.88	80.00	72.41	70.97	67.74	67.74
	VT	68.75	83.33	70.97	77.78	70.97	68.97	61.29	65.52	62.50	61.29	61.29	59.38
	VB	84.38	94.74	68.75	72.41	72.41	66.67	68.97	68.97	63.33	66.67	62.50	64.29
	VF	87.50	87.50	80.00	77.78	80.00	79.31	75.00	79.31	81.25	78.57	75.86	74.19
30	**NR**	96.88	100.0	100.0	96.88	96.55	96.77	93.55	90.63	96.67	96.77	93.75	96.77
	VC	75.00	90.00	81.48	76.67	80.00	78.57	80.00	80.77	74.19	70.00	75.86	70.97
	VT	71.88	80.00	73.33	72.41	70.00	72.41	68.75	68.97	63.33	63.33	61.29	59.38
	VB	84.38	94.44	72.41	72.41	74.07	65.52	68.97	62.07	61.29	58.62	56.25	56.25
	VF	87.50	86.96	82.14	80.77	79.31	79.31	80.65	79.31	70.97	76.67	76.67	74.19
40	**NR**	96.88	100.0	100.0	93.75	96.67	96.77	96.67	90.63	93.55	96.77	93.75	96.77
	VC	71.88	90.00	79.31	79.31	80.77	79.31	76.67	83.33	68.75	70.00	76.67	70.97
	VT	71.88	80.00	70.00	72.41	70.00	66.67	67.74	68.97	61.29	62.50	61.29	59.38
	VB	87.50	100.0	73.33	75.86	74.07	63.33	64.52	62.07	61.29	58.62	56.67	56.67
	VF	87.50	86.96	82.17	81.48	80.00	79.31	80.65	79.31	70.97	76.67	76.67	71.88
60	**NR**	96.88	100.0	100.0	93.75	93.55	93.55	93.55	90.32	96.67	96.67	93.75	96.77
	VC	68.75	81.82	80.00	79.31	81.48	84.62	78.13	84.00	68.75	70.00	76.67	71.88
	VT	75.00	80.00	70.00	68.97	70.00	66.67	67.74	67.86	64.52	64.52	61.29	61.29
	VB	84.38	100.0	76.67	71.43	74.07	65.52	64.52	62.07	65.52	58.62	60.00	56.67
	VF	90.63	86.96	83.33	81.48	80.00	82.14	80.65	79.31	71.88	76.67	74.19	71.88
80	**NR**	96.88	100.0	100.0	93.75	93.55	93.55	93.55	90.63	96.67	96.77	93.75	96.77
	VC	68.75	85.71	80.00	79.31	81.48	84.62	77.42	84.00	68.75	70.00	76.67	71.88
	VT	71.88	80.00	70.00	66.67	70.00	66.67	67.74	67.86	64.52	64.52	61.29	59.38
	VB	81.25	100.0	76.67	71.43	71.43	63.33	64.52	62.07	61.29	60.00	60.00	56.67
	VF	87.50	86.96	86.67	81.48	80.00	82.14	80.65	79.31	71.88	74.19	74.19	71.88
100 to 320	**NR**	96.88	100.0	100.0	93.75	93.55	93.55	93.55	90.63	96.67	96.77	93.75	96.77
	VC	68.75	85.71	80.00	79.31	81.48	84.62	77.42	84.00	68.75	70.00	76.67	71.88
	VT	71.88	80.00	70.00	66.67	70.00	66.67	67.74	67.86	64.52	64.52	61.29	59.38
	VB	81.25	100.0	76.67	71.43	71.43	65.52	64.52	64.29	61.29	60.00	60.00	56.67
	VF	90.63	86.96	86.67	81.48	80.00	82.14	80.65	79.31	71.88	76.67	74.19	71.88

Table 7.21 Voting k-NN classifier sensitivity results for classification problem (PCA).

q	k	1	2	3	4	5	6	7	8	9	10	11	12
5	**NR**	0.00	40.63	12.50	28.13	6.25	6.25	6.25	0.00	0.00	3.13	3.13	0.00
	VC	0.00	40.63	3.13	21.88	15.63	6.25	9.38	15.63	0.00	9.38	6.25	6.25
	VT	0.00	31.25	6.25	9.38	9.38	3.13	0.00	6.25	0.00	0.00	3.13	3.13
	VB	0.00	53.13	9.38	12.50	6.25	3.13	3.13	0.00	3.13	3.13	6.25	3.13
	VF	0.00	21.88	6.25	12.50	9.38	6.25	15.63	9.38	9.38	6.25	0.00	3.13
10	**NR**	0.00	31.25	9.38	9.38	9.38	3.13	0.00	3.13	3.13	6.25	3.13	6.25
	VC	0.00	40.63	6.25	12.50	9.38	6.25	6.25	15.63	3.13	3.13	3.13	0.00
	VT	0.00	25.00	6.25	21.88	12.50	9.38	12.50	12.50	6.25	3.13	6.25	0.00
	VB	0.00	34.38	3.13	12.50	12.50	18.75	0.00	15.63	3.13	3.13	3.13	3.13
	VF	0.00	25.00	6.25	15.63	15.63	12.50	9.38	12.50	6.25	6.25	3.13	6.25
15	**NR**	0.00	15.63	6.25	3.13	6.25	9.38	3.13	3.13	6.25	3.13	0.00	6.25
	VC	0.00	28.13	3.13	15.63	9.38	18.75	12.50	28.13	6.25	6.25	9.38	9.38
	VT	0.00	28.13	3.13	6.25	6.25	6.25	0.00	9.38	0.00	3.13	3.13	3.13
	VB	0.00	37.50	6.25	9.38	6.25	12.50	6.25	12.50	6.25	0.00	3.13	6.25
	VF	0.00	31.25	9.38	6.25	6.25	12.50	9.38	12.50	3.13	12.50	0.00	3.13
20	**NR**	0.00	9.38	6.25	3.13	6.25	6.25	3.13	6.25	3.13	3.13	0.00	0.00
	VC	0.00	25.00	12.50	6.25	15.63	18.75	0.00	21.88	9.38	3.13	3.13	3.13
	VT	0.00	25.00	3.13	15.63	3.13	9.38	3.13	9.38	0.00	3.13	3.13	0.00
	VB	0.00	40.63	0.00	9.38	9.38	6.25	9.38	9.38	6.25	6.25	0.00	12.50
	VF	0.00	25.00	6.25	15.63	6.25	9.38	0.00	9.38	0.00	12.50	9.38	3.13
30	**NR**	0.00	15.63	3.13	0.00	9.38	3.13	3.13	0.00	6.25	3.13	0.00	3.13
	VC	0.00	37.50	15.63	6.25	21.88	12.50	6.25	18.75	3.13	6.25	9.38	3.13
	VT	0.00	21.88	6.25	9.38	6.25	9.38	0.00	9.38	6.25	6.25	3.13	0.00
	VB	0.00	43.75	9.38	9.38	15.63	9.38	9.38	9.38	3.13	9.38	0.00	0.00
	VF	0.00	28.13	12.50	18.75	9.38	9.38	3.13	9.38	3.13	6.25	6.25	3.13
40	**NR**	0.00	12.50	6.25	0.00	6.25	3.13	6.25	0.00	3.13	3.13	0.00	3.13
	VC	0.00	37.50	9.38	9.38	18.75	9.38	6.25	25.00	0.00	6.25	6.25	3.13
	VT	0.00	21.88	6.25	9.38	6.25	6.25	3.13	9.38	3.13	0.00	3.13	0.00
	VB	0.00	46.88	6.25	9.38	15.63	6.25	3.13	9.38	3.13	9.38	6.25	6.25
	VF	0.00	28.13	12.50	15.63	6.25	9.38	3.13	9.38	3.13	6.25	6.25	0.00
60	**NR**	0.00	9.38	6.25	0.00	3.13	3.13	3.13	3.13	6.25	6.25	0.00	3.13
	VC	0.00	31.25	6.25	9.38	15.63	18.75	0.00	21.88	0.00	6.25	6.25	0.00
	VT	0.00	21.88	6.25	9.38	6.25	6.25	3.13	12.50	3.13	3.13	3.13	3.13
	VB	0.00	50.00	6.25	12.50	15.63	9.38	3.13	9.38	9.38	9.38	6.25	6.25
	VF	0.00	28.13	6.25	15.63	6.25	12.50	3.13	9.38	0.00	6.25	3.13	0.00
80	**NR**	0.00	9.38	6.25	0.00	3.13	3.13	3.13	0.00	6.25	3.13	0.00	3.13
	VC	0.00	34.38	6.25	9.38	15.63	18.75	3.13	21.88	0.00	6.25	6.25	0.00
	VT	0.00	21.88	6.25	6.25	6.25	6.25	3.13	12.50	3.13	3.13	3.13	0.00
	VB	0.00	50.00	6.25	12.50	12.50	6.25	3.13	9.38	3.13	6.25	6.25	6.25
	VF	0.00	28.13	6.25	15.63	6.25	12.50	3.13	9.38	0.00	6.25	3.13	0.00
100 to 320	**NR**	0.00	9.38	6.25	0.00	3.13	3.13	3.13	0.00	6.25	3.13	0.00	3.13
	VC	0.00	34.38	6.25	9.38	15.63	18.75	3.13	21.88	0.00	6.25	6.25	0.00
	VT	0.00	21.88	6.25	6.25	6.25	6.25	3.13	12.50	3.13	3.13	3.13	0.00
	VB	0.00	46.88	6.25	12.50	12.50	9.38	3.13	12.50	3.13	6.25	6.25	6.25
	VF	0.00	28.13	6.25	15.63	6.25	12.50	3.13	9.38	0.00	6.25	3.13	0.00

Table 7.22 Voting k-NN inconclusive rates for classification problem (PCA).

7.3.2 Independent Component Analysis (ICA)

For feature extraction purposes, each centered sample is represented by its projections on selected q independent components in a similar fashion to the PCA. Nevertheless, there is no direct way to order the independent components based on their contribution, unlike the case of principal components. In order to identify those independent components that provide the best discrimination between pathologies, we sort the independent components based on the following approach:

1. In the first step, a table of p-values of the standard two-sample t-test between all possible pairs of classes within the five classes of interest is constructed based on the training data samples and a particular independent component. This results in a table containing ten distinct values.

2. The second step involves the calculation of the number of p-values in this table above 5% (i.e., not statistically significant at the 5% level), which we will call here the discrimination index (DI).

3. The independent components are then sorted according to this number. This ensures that the selected number of independent components will contain the ones that are most discriminating.

4. The number of independent components is selected based on the discrimination index whereby the first group contains those independent components with one or less p-values more than 5%, the second group contains those with two or less, and so on.

The reason for this approach is to eliminate any bias in the classification outcome as a result of the arbitrary selection of independent components within the group having the same discrimination index. Number of chosen ICs for each DI is shown in table 7.23.

DI	1	2	3	4	5	6	7	8	9	10
q	13	99	162	219	249	265	269	279	290	320

Table 7.23 Results for sorting the ICs.

The detection and classification results using ICA are shown in Tables 7.24 – 7.33. The number of independent components used is a direct function of the discrimination index (DI) described above.

DI	**Specificity(%)**	**Sensitivity(%)**
1	71.88	85.16
2	84.38	80.47
3	75.00	96.88
4	78.13	96.88
5 – 10	78.13	96.09

Table 7.24 Minimum distance classifier results for detection problem (ICA).

DI	**Specificity(%)**	**Sensitivity(%)**
1	53.13	92.97

Table 7.25 Bayes minimum-error classifier results for detection problem (ICA).

DI	**Specificity(%)**	**Sensitivity(%)**
1	71.88	95.31
2	90.63	96.88
3	96.88	97.66
4	100	97.66
5	100	98.44
6 - 10	100	97.66

Table 7.26 NN classifier (k-NN with k=1) results for detection problem (ICA).

DI	k	1	2	3	4	5	6	7	8	9	10	11	12
1	Specificity (%)	71.88	81.82	59.38	65.52	59.38	61.29	62.50	57.14	53.13	51.61	53.13	46.43
	Sensitivity (%)	95.31	97.48	95.31	95.93	94.53	96.72	93.75	94.40	94.53	95.16	92.97	93.70
2	Specificity (%)	90.63	92.00	75.00	76.67	71.88	74.19	71.88	73.33	68.75	68.75	68.75	68.75
	Sensitivity (%)	96.88	97.50	95.31	96.00	95.31	97.58	96.09	97.54	97.66	97.58	96.88	96.77
3	Specificity (%)	96.88	96.30	93.75	93.55	90.63	90.63	90.63	93.10	87.50	90.32	87.50	90.00
	Sensitivity (%)	97.66	97.56	96.09	96.72	96.09	96.09	94.53	94.35	92.97	94.31	91.41	92.74
4	Specificity (%)	100.0	100.0	93.75	96.67	93.75	93.33	93.75	93.33	90.63	93.55	90.63	90.00
	Sensitivity (%)	97.66	98.33	95.31	95.87	95.31	95.96	95.31	95.08	93.75	93.50	90.63	91.80
5	Specificity (%)	100.0	100.0	96.88	96.77	93.75	93.10	93.75	93.10	90.63	93.10	93.75	93.33
	Sensitivity (%)	98.44	98.36	95.31	95.16	94.53	96.75	96.09	95.90	92.97	93.60	91.41	92.86
6 - 10	Specificity (%)	100.0	100.0	96.88	96.67	93.75	93.33	93.75	93.10	93.75	93.33	90.63	93.33
	Sensitivity (%)	97.66	98.35	95.31	95.12	94.53	96.75	96.09	95.90	92.97	93.60	92.19	92.86

Table 7.27 Voting k-NN classifier results for detection problem (ICA).

DI	k	1	2	3	4	5	6	7	8	9	10	11	12
1	Normal (%)	0.0	31.25	0.0	9.38	0.0	3.13	0.0	12.50	0.0	3.13	0.0	12.50
	Abnormal (%)	0.0	7.03	0.0	3.91	0.0	4.69	0.0	2.34	0.0	3.13	0.0	0.78
2	Normal (%)	0.0	21.88	0.0	6.25	0.0	3.13	0.0	6.25	0.0	0.0	0.0	0.0
	Abnormal (%)	0.0	6.25	0.0	2.34	0.0	3.13	0.0	4.69	0.0	3.13	0.0	3.13
3	Normal (%)	0.0	15.63	0.0	3.13	0.0	0.0	0.0	9.38	0.0	3.13	0.0	6.25
	Abnormal (%)	0.0	3.91	0.0	4.69	0.0	5.47	0.0	3.13	0.0	3.91	0.0	3.13
4	Normal (%)	0.0	18.75	0.0	6.25	0.0	6.25	0.0	6.25	0.0	3.13	0.0	6.25
	Abnormal (%)	0.0	6.25	0.0	5.47	0.0	3.91	0.0	4.69	0.0	3.91	0.0	4.69
5	Normal (%)	0.0	15.63	0.0	3.13	0.0	9.38	0.0	9.38	0.0	9.38	0.0	6.25
	Abnormal (%)	0.0	4.69	0.0	3.13	0.0	3.91	0.0	4.69	0.0	2.34	0.0	1.56
6 – 10	Normal (%)	0.0	15.63	0.0	6.25	0.0	6.25	0.0	9.38	0.0	6.25	0.0	6.25
	Abnormal (%)	0.0	5.47	0.0	3.91	0.0	3.91	0.0	4.69	0.0	2.34	0.0	1.56

Table 7.28 Voting k-NN inconclusive rates for detection problem (ICA).

DI	NR	VC	VT	VB	VF
1	71.88	68.75	59.38	37.50	53.13
2	84.38	62.50	62.50	53.13	46.88
3	75.00	78.13	53.13	21.88	84.38
4	75.00	75.00	56.25	18.75	84.38
5-10	78.13	78.13	56.25	21.88	87.50

Table 7.29 Minimum distance classifier sensitivity (%) results for classification problem (ICA)

DI	NR	VC	VT	VB	VF
1	53.13	65.63	68.75	59.38	65.63

Table 7.30 Bayes minimum-error classifier sensitivity (%) results for classification problem (ICA)

DI	NR	VC	VT	VB	VF
1	71.88	59.38	59.38	59.38	65.63
2	90.63	56.25	65.63	71.88	87.50
3	96.88	68.75	68.75	71.88	87.50
4	100	68.75	68.75	84.38	87.50
5	100	68.75	71.88	81.25	90.63
6 - 10	100	65.63	71.88	81.25	90.63

Table 7.31 NN classifier (k-NN with k=1) sensitivity (%) results for classification problem (ICA).

DI	k	1	2	3	4	5	6	7	8	9	10	11	12
1	**NR**	71.88	85.71	76.00	73.33	67.86	62.50	64.52	62.50	64.52	64.52	65.63	62.07
	VC	59.38	73.68	64.29	61.54	56.67	70.83	68.97	64.00	67.86	67.86	65.52	65.52
	VT	59.38	70.00	65.52	64.00	69.23	66.67	67.86	67.86	65.63	66.67	71.43	72.41
	VB	59.38	70.00	57.69	64.00	66.67	66.67	65.52	60.00	60.71	65.52	61.29	61.29
	VF	65.63	73.68	73.08	63.33	67.86	71.43	64.52	66.67	67.74	66.67	68.97	65.52
2	**NR**	90.63	100.0	96.00	89.29	85.71	82.76	80.65	80.00	83.33	86.67	86.67	83.87
	VC	56.25	78.95	80.00	70.37	61.54	56.67	60.71	59.26	64.29	56.67	62.96	64.29
	VT	65.63	80.00	76.67	75.00	74.19	75.00	71.43	73.33	70.97	68.97	65.52	59.38
	VB	71.88	85.71	75.00	66.67	76.92	78.57	80.77	72.41	72.41	73.08	65.52	65.52
	VF	87.50	95.65	85.71	89.66	85.19	80.00	80.77	74.07	84.00	77.78	74.07	74.07
3	**NR**	96.88	100.0	100.0	93.75	100.0	93.55	96.67	93.75	93.55	100.0	93.55	93.55
	VC	68.75	85.00	74.07	82.14	88.89	85.19	85.71	86.96	75.00	77.78	75.86	77.42
	VT	68.75	86.36	75.00	68.97	74.07	60.71	51.61	58.62	56.67	54.84	56.67	51.72
	VB	71.88	84.62	68.00	69.23	76.92	68.00	61.29	63.33	63.33	62.50	62.50	62.50
	VF	87.50	90.00	89.29	84.62	80.00	81.48	80.00	77.42	80.65	78.13	78.13	78.13
4	**NR**	100.0	100.0	100.0	96.88	100.0	100.0	96.77	93.75	93.55	96.77	93.55	93.55
	VC	68.75	85.00	77.78	73.33	85.19	76.67	77.42	80.77	73.33	68.97	70.97	70.00
	VT	68.75	90.48	71.43	70.00	78.57	68.97	65.52	64.52	63.33	59.38	56.67	56.25
	VB	84.38	89.47	67.86	71.43	68.97	62.96	65.52	61.29	62.07	58.62	59.38	58.62
	VF	87.50	90.91	85.19	78.57	73.33	80.00	78.13	77.42	78.13	78.13	78.13	78.13
5	**NR**	100.0	100.0	100.0	96.88	100.0	100.0	96.77	93.75	93.55	96.77	93.75	96.77
	VC	68.75	89.47	82.14	73.33	80.77	80.77	83.33	80.00	64.52	72.41	75.86	68.75
	VT	71.88	84.00	70.97	70.00	70.97	68.97	65.63	66.67	61.29	63.33	63.33	61.29
	VB	81.25	94.12	75.86	68.97	68.97	65.52	66.67	60.00	56.25	56.67	60.00	56.67
	VF	90.63	91.30	82.76	82.14	82.76	88.89	80.65	80.00	75.00	77.42	77.42	78.13
6 - 10	**NR**	100.0	100.0	100.0	96.88	100.0	100.0	100.0	93.75	96.77	96.77	93.55	96.77
	VC	65.63	80.95	79.31	75.00	80.77	84.62	80.65	83.33	64.52	72.41	75.00	68.75
	VT	71.88	80.77	70.97	70.00	70.97	66.67	65.63	66.67	61.29	64.52	61.29	65.52
	VB	81.25	94.12	73.33	71.43	71.43	65.52	66.67	62.07	60.00	60.00	60.00	56.67
	VF	90.63	91.67	82.76	83.33	86.67	88.89	80.65	80.00	75.00	77.42	78.13	76.67

Table 7.32 Voting k-NN classifier sensitivity (%) results for classification problem (ICA).

DI	k	1	2	3	4	5	6	7	8	9	10	11	12
1	**NR**	0.0	34.38	21.88	6.25	12.50	0.00	3.13	0.00	3.13	3.13	0.00	9.38
	VC	0.0	40.63	12.50	18.75	6.25	25.00	9.38	21.88	12.50	12.50	9.38	9.38
	VT	0.0	37.50	9.38	21.88	18.75	6.25	12.50	12.50	0.00	6.25	12.50	9.38
	VB	0.0	37.50	18.75	21.88	15.63	15.63	9.38	6.25	12.50	9.38	3.13	3.13
	VF	0.0	40.63	18.75	6.25	12.50	12.50	3.13	6.25	3.13	6.25	9.38	9.38
2	**NR**	0.0	28.13	21.88	12.50	12.50	9.38	3.13	6.25	6.25	6.25	6.25	3.13
	VC	0.0	40.63	21.88	15.63	18.75	6.25	12.50	15.63	12.50	6.25	15.63	12.50
	VT	0.0	37.50	6.25	25.00	3.13	12.50	12.50	6.25	3.13	9.38	9.38	0.00
	VB	0.0	34.38	12.50	6.25	18.75	12.50	18.75	9.38	9.38	18.75	9.38	9.38
	VF	0.0	28.13	12.50	9.38	15.63	6.25	18.75	15.63	21.88	15.63	15.63	15.63
3	**NR**	0.0	18.75	6.25	0.00	9.38	3.13	6.25	0.00	3.13	6.25	3.13	3.13
	VC	0.0	37.50	15.63	12.50	15.63	15.63	12.50	28.13	12.50	15.63	9.38	3.13
	VT	0.0	31.25	12.50	9.38	15.63	12.50	3.13	9.38	6.25	3.13	6.25	9.38
	VB	0.0	59.38	21.88	18.75	18.75	21.88	3.13	6.25	6.25	0.0	0.0	0.0
	VF	0.0	37.50	12.50	18.75	6.25	15.63	6.25	3.13	3.13	0.0	0.0	0.0
4	**NR**	0.0	18.75	6.25	0.0	6.25	6.25	3.13	0.0	3.13	3.13	3.13	3.13
	VC	0.0	37.50	15.63	6.25	15.63	6.25	3.13	18.75	6.25	9.38	3.13	6.25
	VT	0.0	34.38	12.50	6.25	12.50	9.38	9.38	3.13	6.25	0.0	6.25	0.0
	VB	0.0	40.63	12.50	12.50	9.38	15.63	9.38	3.13	9.38	9.38	0.0	9.38
	VF	0.0	31.25	15.63	12.50	6.25	6.25	0.0	3.13	0.0	0.0	0.0	0.0
5	**NR**	0.0	15.63	3.13	0.0	6.25	6.25	3.13	0.0	3.13	3.13	0.0	3.13
	VC	0.0	40.63	12.50	6.25	18.75	18.75	6.25	21.88	3.13	9.38	9.38	0.0
	VT	0.0	21.88	3.13	6.25	3.13	9.38	0.0	6.25	3.13	6.25	6.25	3.13
	VB	0.0	46.88	9.38	9.38	9.38	9.38	6.25	6.25	0.0	6.25	6.25	6.25
	VF	0.0	28.13	9.38	12.50	9.38	15.63	3.13	6.25	0.0	3.13	3.13	0.0
6 - 10	**NR**	0.0	15.63	3.13	0.0	6.25	6.25	6.25	0.0	3.13	3.13	3.13	3.13
	VC	0.0	34.38	9.38	12.50	18.75	18.75	3.13	25.00	3.13	9.38	12.50	0.0
	VT	0.0	18.75	3.13	6.25	3.13	6.25	0.0	6.25	3.13	3.13	3.13	9.38
	VB	0.0	46.88	6.25	12.50	12.50	9.38	6.25	9.38	6.25	6.25	6.25	6.25
	VF	0.0	25.00	9.38	6.25	6.25	15.63	3.13	6.25	0.0	3.13	0.0	6.25

Table 7.33 Voting k-NN inconclusive rates for classification problem (ICA).

7.3.3 General Discussion of BSS techniques

It is noticed that the Bayes minimum-error classifier results contain one entry in the case of ICA results (tables 7.25 & 7.30) and it does not exist for some q values in the case of PCA results (tables 7.14 & 7.19). It can be interpreted by the following; the Bayes minimum-error classifier was computed up to the value of components that allowed the covariance matrix to have a stable inverse.

As can be observed from the results, the k-nearest neighbor classifier seems to provide the best results followed by the minimum distance classifier. On the other hand, the results of Bayes minimum-error classifier were rather poor and suggest that the distribution of the clusters may not be Gaussian as this technique assumes.

The first observation on the results is that the accuracy of the results gets better as the number of principal/independent components gets larger up to a certain value. After that, the accuracy may slightly deteriorate for some cases (see for example Tables 7.15, 7.20, 7.26, and 7.31). This is particularly evident in the case of PCA where the additional principal components above a certain limit are associated with very small eigenvalues. This means that these components do not contribute much to the signal and can be ignored in the detection/classification process. This is apparent from the fact that the results seem to saturate beyond a certain number of components.

The results of PCA and ICA appear to become closer as the number of components included in the feature vector increase. This is a direct result of the fact that both PCA and ICA provide a complete basis set of vectors to describe the space of ECG signals. The differences between the two techniques are only apparent in the specific directions of each of these vectors. These differences tend to make the spanned subspaces obtained using the two techniques rather different when a small number of vectors are used. In the extreme case when all vectors (in this case 375) are used, one can find an orthogonal transformation to transform the feature vector based on PCA to that based on ICA. Given that the classification techniques used here rely primarily on the Euclidean distance in assigning class membership and since orthogonal transformations preserve Euclidean distance, it is not surprising to see that the classification results match in this special case and to realize the convergence of the two sets of results to the same solution.

The three classifiers implemented in this work appear to provide substantially different receiver operating characteristics demonstrating the compromise between detection rates and false alarm rates. The optimization of this classification is left for future work. Nevertheless, the results of these classifiers provide a general conclusion

about the classification accuracy and the upper limits in the sensitivity and specificity values obtainable using the proposed features.

It should be noted that the proposed methods classify ECG samples based on the presence and strength of its frequency components ignoring their relative delays. This logic is particularly justified in those cases in which the heart rate becomes higher as in tachycardia and fibrillation where the frequency components become significantly different. Even though the intuitive justification in the other arrhythmia types might not be obvious, the results demonstrate the presence of significant differences between these types using only the magnitude information.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

- The use of ECG signal features from nonlinear dynamical modeling was studied. The detailed implementation of automatic feature extraction algorithms was demonstrated (chapter 3). The results of applying this program on a large data set of actual ECG signals from five different classes were presented (chapter 7). The statistical analysis of the results suggests that the use of such features can be advantageous to ECG arrhythmia detection. They also illustrate the limitations of such features in classifying the type of ECG abnormality.

- Two blind source separation techniques were used to derive ECG signal features for arrhythmia detection and classification. A large database of ECG signals was used to compute a set of basic signal components that compose any ECG signal using PCA and ICA. A set of features was obtained by

projecting a given ECG signal onto the subspace of those basic signals, which were subsequently used for arrhythmia detection and classification using conventional statistical methods. The results indicate the value of such features for practical use in clinical settings.

8.2 Future Work

- Taking into consideration other leads than lead II, i.e. combining extracted features from different leads.

- In order to improve the overall classification accuracy, using sequential hypothesis testing by applying the statistical classification to multiple independent sample windows that may or may not overlap.

- The possibility of augmenting the feature vector with other classical ECG signal features like the R-R interval and statistical features the signal to improve the results.

- The results of PCA and ICA suggest the value of using either PCA or ICA for noise suppression and dimensionality reduction prior to classification with any other technique. It is observed that the accuracy of the results gets better as the number of principal/independent components gets larger up to a certain value after that, the accuracy may slightly deteriorate for some cases. This indicates that there is indeed a part of the signal that contributes random noise into the classification problem.

- The questions of whether the inclusion of phase information in the application of PCA and ICA would provide better discrimination.

- The use of PCA in compression of ECG signals. The results show that the use of PCA demonstrates a significant degree of energy compaction of the training samples. However, the proposed transformation cannot be used for such applications as data compression for ECG signals since the phase part of

the signals was ignored. A more useful transformation for this application would be the discrete cosine transformation whereby a (2N-1)-point symmetric signal is composed using the N-point ECG sample window.

References

[Abarbanel *et al.* 1998] H.D.I. Abarbanel, T.W. Frison, and L. Tsimring, "Obtaining order in a world of chaos, time domain analysis of nonlinear and chaotic signals," *IEEE Sig. Proc. Mag.*, pp. 49-65, May 1998.

[Afonso & Tompkins 1995] V.X. Afonso, and W. J. Tompkins, "Detecting ventricular fibrillation, selecting the appropriate time-frequency analysis tool for the application," *IEEE Engineering in Medicine and Biology Mag.*, vol. 14, pp. 152-159, March 1995.

[Albano *et al.* 1988] A.M. Albano, J. Munech, and C. Schwartz, "Singular-value decomposition and the Grassberger-Procaccia algorithm," *Phys. Rev. A*, vol. 38, no. 6, pp.3017-3026, September 1988.

[Al-Fahoum & Howitt 1999] A.S. Al-Fahoum, I. Howitt, "Combined wavelet transformation and radial basis neural networks for classifying life-threatening cardiac arrhythmias," *Medical & Biological Engineering & Computing*, vol. 37, pp 566-573, September 1999.

[Bartlett *et al.* 1998] M. S. Bartlett, H. M. Lades, and T. J. Sejnowski, "Independent component representations for face recognition," *Proceedings of the SPIE Symosium on Electronic Imaging: Science and Technology; Conference on Human Vision and Electronic Imaging III, San Jose, California,* January 1998.

[Biel *et al.* 2001] L. Biel, O. Pettersson, L. Philipson, P. Wide, "ECG analysis: a new approach in human identification," *IEEE Trans. Instrumentation and Measurement*, vol. 50, no. 3, pp. 808 – 812, June 2001.

[Burrus *et al.* 1994] C. S. Burrus, J. H. McClellon, A. V. Oppenheim, T. W. Parks, R. W. Schafer, and H. W. Schuessler, *Computer-Based Exercises for Signal Processing Using Matlab®*, Prentice-Hall, New Jersy, 1994.

[Casaleggio *et al.* 1997] A. Casaleggio, and S. Braiotta, "Estimation of Lyapunov exponents of ECG time series – the influence of parameters," *Chaos, Solitons & Fractals,* vol. 8, no. 10, pp. 1591-1599, 1997.

[Devaney 1992] R. L. Devaney, *A First Course in Chaotic Dynamical Systems: Theory and Experiment,* Addison-Wesely, 1992.

[Fukunaga 1990] K. Fukunaga, *Introduction to Statistical Pattern Recogntion*, 2nd ed., Academic Press, New York, 1990.

[Gerbrands 1981] J.J. Gerbrands, "On the relationships between SVD, KLT, and PCA," *Pattern Recognition*, vol. 14, Nos. 1-6, pp. 375-381, 1981.

[Goldberger 1999] A. L. Goldberger, *Clinical Electrocardiography, A Simplified Approach*, 6th ed., Mosby, 1999.

[Govindan *et al.* 1998] R.B. Govindan, K. Narayanan, and M.S. Gopinathan, "On the evidence of deterministic chaos in ECG: surrogate and predictability analysis," *Chaos,* vol. 8, no. 2, pp.495-502, 1998.

[Grassberger & Procaccia 1983] P. Grassberger and I. Procaccia, “Measuring the strangeness of strange attractors,” Physica D, vol. 9, pp. 189-208, 1983.

[Guillén *et al.* 1989] S.G. Guillén, M.T. Arredondo, G. Martin, and J.M.F. Corral, “Ventricular fibrillation detection by autocorrelation function peak analysis,” *Journal of Electrocardiology*, vol. 22 supplement, pp. 253-262, 1989.

[Hampton 1999] J. R. Hampton, *The ECG Made Easy*, 5th ed., Churchill Livingstone, 1999.

[Hanselman & Littlefield 1998] D. Hanselman and B. Littlefield, *Mastering Matlab 5: A Comprehensive Tutorial and Reference,* Prentice Hall, New Jersy, 1998.

[Hobbie 1988] R. K. Hobbie, *Intermediate Physics for Medicine and Biology*, 2nd ed., John Wiley & Sons, 1988.

[Hyvärinen & Oja 1997] A. Hyvärinen, and E. Oja, “A fast fixed-point algorithm for independent component analysis,” *Neural Computation*, 9(7): 1483-1492, 1997.

[Hyvärinen 1999] A. Hyvärinen, “Survey on independent component analysis,” *Neural Computing Surveys*, vol. 2, pp. 94-128, 1999.

[Hyvärinen & Oja 2000] A. Hyvärinen and E. Oja, “Independent Component Analysis: Algorithms and Applications,” *Neural Networks*, vol. 13, no. (4-5), pp. 411-430, 2000.

[Hyvärinen *et al.* 2001] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.

[Ifeachor & Jervis 1993] E. C. Ifeachor and B. W. Jervis, *Digital Signal Processing: A Practical Approach*, Addison-Wesely, 1993.

[Kadah *et al.* 1996] Y.M. Kadah, A.A. Farag, J.M. Zurada, A.M. Badawi, and A.M. Youssef, "Classification algorithms for quantitative tissue characterization of diffuse

liver disease from ultrasound images," *IEEE Trans. Medical Imaging*, vol. 15, no. 4, pp. 466-478, August 1996.

[Khadra *et al.* 1997] L. Khadra, A.S. Al-Fahoum, and H. Al-Nashash, "Detection of life-threatening cardiac arrhythmias using the wavelet transformation," *Medical & Biological Engineering & Computing,* vol. 35, pp 626-632, November 1997.

[Kugiumtzis 1996] D. Kugiumtzis, "State space reconstruction parameters in the analysis of chaotic time series - the role of the time window length," *Physica D*, vol. 95, pp. 13 - 28, 1996.

[Lathauwer *et al.* 2000] De Lathauwer L, De Moor B, Vandewalle J., "Fetal electrocardiogram extraction by blind source subspace separation," *IEEE Trans. Biomedical Engineering*, vol. 47, no. 5, pp. 567-572, May 2000.

[Minami *et al.* 1999] K. Minami, H. Nakajima, and T. Toyoshima, "Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 2, pp. 179-185, February 1999.

[MIT-BIH 1997] The MIT-BIH Arrhythmia Database, *3rd ed*., Harvard-MIT Division of Health Sciences and Technology, May 1997.

[Moody & Mark 2001] G. B. Moody, and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," *IEEE Engineering in Medicine and Biology Magazine,* vol. 20, no. 3, pp. 45-50, May 2001.

[Moore & McCabe 1993] D.S. Moore, and G.P. McCabe, *Introduction to the Practice of Statistics, 2nd ed.,* W.H. Freeman and Company, New York, 1993.

[Oppenheim & Schafer 1989] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing,* Prentice Hall, New Jersy, 1989.

[Panerai *et al.* 1988] R.B. Panerai, A. Ferriera, O.F. Brum, "Principal component analysis of multiple noninvasive blood flow derived signals," *IEEE Trans. Biomedical Engineering*, vol. 35, no. 7, pp. 533 – 538, July 1988.

[Papoulis 1991] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed., WCB McGraw-Hill, 1991.

[Pritchard & Duke 1992] W. S. Pritchard, and D. W. Duke, "Measuring chaos in the brain: a tutorial review of nonlinear dynamical EEG analysis," *Internatoinal Journal of Neuroscience*, vol 67, pp. 31-80, 1992.

[Pritchard & Duke 1995] W. S. Pritchard, and D. W. Duke, "Measuring chaos in the brain: a tutorial review of EEG dimension estimation," *Brain and Cognition,* vol. 27, no. 3, pp. 353-397, 1995.

[Richter *et al.* 1998] M. Richter, T. Schreiber, and D. T. Kaplan, "Fetal ECG Extraction with Nonlinear State-Space Projections," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 1, pp. 133-137, January 1998.

[Stamkopoulos *et al.* 1998] T. Stamkopoulos, K. Diamantaras, N. Maglaveras, M. Strintzis, "ECG analysis using nonlinear PCA neural networks for ischemia detection," *IEEE Trans. Signal Processing*, vol. 46, no. 11, pp. 3058 – 3067, Nov. 1998.

[Thakor & Zhu 1991] N.V. Thakor, and Y. Zhu, "Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection," *IEEE Transactions on Biomedical Engineering*, vol. 38, no. 8, pp. 785-794, August 1991.

[Thakor & Pan 1990] N.V. Thakor and K. Pan, "Detection of tachycardia and fibrillation: a sequential time domain approach," *IEEE Engineering in Medicine and Biology Magazine,* vol. 9, pp. 21-24, March 1990.

[Thakor *et al.* 1990] N.V. Thakor, Y. Zhu, and K. Pan, "Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 9, pp. 837-843, September 1990.

[Thakor *et al.* 1994] N.V. Thakor, A. Natarajan, and G.F. Tomaselli, "Multiway sequential hypothesis testing for tachyarrhythmia discrimination," *IEEE Transactions on Biomedical Engineering*, vol. 41, no. 5, pp. 480-487, May 1994.

[Wagner 2001] G.S. Wagner, *Marriott's Practical Electrocardiography*, *10th ed*., Lippincott Williams & Wilkins, Philadelphia, 2001.

[Watkins 1991] D. S. Watkins, *Fundamentals of Matrix Computations,* John Wiley & Sons, 1991.

[Wolf] Http://www.users.iterport.net/~wolf.

[Zarzoso *et al.* 1997] V. Zarzoso, A. K. Nandi and E. Bacharakis, "Maternal and Foetal ECG Separation using Blind Source Separation Methods," *IMA Journal of Mathematics Applied in Medicine & Biology*, vol. 14, pp. 207-225, 1997.

ملخص الرسالة

يعتبر القلب أهم عضوفي جسم الإنسان حيث أنه المسئول عن إمداد الدم إلى الجسم كله. يتم هذا عن طريق تنشيط عضلة القلب بواسطة إشارات كهربية تبدأ من عقدة خلايا موجودة في أعلى الأذين الأيمن تسمى العقدة الجيبية الأذينية ثم تنتقل إلى الأذين الأيسر حيث يؤدي هذا إلى تنشيط كلا الأذينين لينقبضوا و يضخوا الدم إلى البطينين. ثم بعد ذلك تنتقل الإشارة الكهربية للبطينين عن طريق الوصلة الأذينيبطينية التي تتكون من عقدة و حزمة ألياف. وتنتشر الإشارة الكهربية خلال عضلة البطين عن طريق ألياف متخصصة تسمى ألياف بوركينجي. انتشار الإشارة الكهربية خلال البطينين يؤدي إلى انقباضها و ضخها للدم إلى الرئتين وإلى الدورة الدموية العامة.

أي تغير في هذا المسار الطبيعي للإشارة الكهربية خلال عضلة القلب يسمى "لا نظمية القلب" و يظهر هذا على صورة اسقاط نبضة أو أن ينبض القلب بصورة غير منتظمة بمعدل أبطئ أو أسرع من المعدل الطبيعي. بالرغم من أن لا نظمية القلب تعتبر إشارة مهمة لوجود مرض في القلب، فإنها من الممكن أن تنتج عن أسباب طبيعية مثل الإرهاق الزائد أو تناول الكافيين و السجائر أو بعض الأدوية.

تقسم لا نظمية القلب حسب مصدرها بطينية أو غير بطينية و أيضاً حسب معدل ضربات القلب. عادةً ما تكون اللانظمية الصادرة من البطين أكثر خطورة و مهددة لحياة المريض. و لذلك فالتعرف على اللانظميات الصادرة من البطين تكون ضرورية لحياة المريض.

تقدم الرسالة دراسة حل لمشكلة التعرف على وجود لانظمية بطينية و تحديد نوعها. تشتمل الدراسة على أربعة أنواع للانظمية القلب: 1- زوج النبضات البطينية المبكرة، 2- الإزدواجية البطينية (واحدة طبيعية وأخرى بطينية مبكرة)، 3- إسراع قلبي بطيني، 4- إنقباض بطيني غير منتظم. تنقسم هذه المشكلة إلى قسمين الأول هو استخلاص ملامح مميزة لإشارات رسم القلب و الثاني هو إستخدام هذه الملامح أو الصفات للتعرف على وجود اللانظمية و تصنيفها.

تقدم الرسالة طرق جديدة لتحليل و معالجة إشارات رسم القلب وهي: 1- نمذجة القلب على أنه نظام ديتاميكي غير خطي، 2- تحليل المركبات الأساسية لإشارة رسم القلب، 3- تحليل المركبات المستقلة لإشارة رسم القلب. ثم يتم تصنيف نوع اللا نظمية عن طريق تطبيق طرق فرز إحصائية: 1- المسافة الصغرى، 2- طريقة بايز لنسبة الخطأ الصغرى، 3- طريقة أقرب جار.

طـرق جـديـدة للتعـرف عـلـى لا نظمية القـلـب

إعـــــــــــداد

المهندس / محمد إبـراهـيم إسـمـاعـيل عـويـس

رسـالـة مقـدمـة إلي كليـة الهندسـة – جـامعـة القـاهـرة

كجـزء مـن متطـلبـات الحصـول على درجة الدكتوراة

فى الهندسـة الحيـوية الـطبيـة والمنظـومـات

كلية الهندسة – جامعة القاهرة

الجيزة- جمهورية مصر العربية

نوفمبر 2001

طـرق جـديـدة للتعـرف عـلـى لا نظمية القـلب

إعـــــــــــداد

المهندس / محمد إبـراهـيم إسـمـاعـيل عـويـس

رسـالـة مقـدمـة إلي كليـة الهندسـة – جـامعـة القـاهـرة

كجـزء مـن متطـلبـات الحصـول على درجة الدكتوراة

فى الهندسـة الحيـوية الـطبيـة والمنظـومـات

تحـت إشـراف

أ.د/ أبـوبكر محمـد عـبدالـفتـاح يـوسـف

د/ ياسـر مصطـفىإبـراهـيم قـدح

كلية الهندسة – جامعة القاهرة

الجيزة- جمهورية مصر العربية

نوفمبر 2001

طرق جديدة للتعرف على لا نظمية القلب

إعـــــــــــداد

المهندس / محمد إبراهيم إسماعيل عويس

رسالة مقدمة إلي كلية الهندسة – جامعة القاهرة

كجزء من متطلبات الحصول على درجة الدكتوراة

فى الهندسة الحيوية الطبية والمنظومات

يعتمد من لجنة الممتحنين:

أ.د/ عبدالمنعم عبدالظاهر وهدان (عضو)

أ.د/ محمد عماد موسي رسمي (عضو)

أ.د/ أبوبكر محمد عبدالفتاح يوسف (المشرف الرئيسي)

كلية الهندسة – جامعة القاهرة

الجيزة- جمهورية مصر العربية

نوفمبر 2001