Identification of Cardiac Signals in Ambulatory ECG data

Mahdi Torabi

Supervisor: Dr Kevin Paulson

School of Engineering and Computer Science

University of Hull

This dissertation is submitted for the degree of

Doctor of Philosophy

December 2018


بسم الله الرحمن الرحيم

This thesis is dedicated to the memory of:

my dearest father, Haj Khosro Torabi;

and also dedicated to my dearest mother Eradat Mohammadi.

Acknowledgements

This thesis would not have been possible without the help of many people.

I would like to show my gratitude and give my thanks to Dr Kevin Paulson, who was my supervisor during the PhD and MSc courses. This research would not have been possible without his constant support. Kevin has been far beyond a supervisor for me and I hope to carry forward the lessons which I have learned from him in my life. I would like to thank my second supervisor, Dr Anthony Wilkinson, for his advice during my study and Dr Philip Langley who let me use his Holter device during this project.

Many members of my family have contributed considerable support during this project. In particular, I would like to thank my beloved wife Dr Armaghan Naghili, my father-in-law Prof Behrouz Naghili and my mother-in-law Tahereh Choupani, for all their love and patience during the last year of my study, which I lost my father (May he rest in peace). Also, my oldest brother Firooz and his family who have always supported me in life. I also give special thanks to my oldest sister Dr Azam Torabi, my brother-in-law Prof Hossein Panahi and my lovely niece and nephew: Mona and Majid Panahi. They have provided housing, support and encouragement throughout my UK studies. My other siblings and their families: Javad, Mohammad, Akram, Amir Reza, Dr Effat, Ashraf, Dr Behrang, Ali Reza, Amin, Dr Arman and my numerous nieces and nephews; have also provided considerable support and encouragement throughout my life.

I sincerely thank to my committee members: my examiners Dr. Mansour Youseffi and Dr Sergey Rybchenko for their insightful feedbacks and Dr Anthony Bateson who was chair in my viva.

My friends have also been vital to this project. I shared many of the ideas in this research with Dr Yashar Baradaran who gave me invaluable feedback.

I also would like to thank Dr Reza Sameni from the University of Shiraz- Iran and Dr Hamed Danandeh Hesar from University of Khaje Nasir Toosi - Iran for their advice and the help they provided, permitting me to use their thesis and software in comparisons.

My sincere thanks also goes to my friends and colleagues who have helped me stay positive through these years. Their support and care helped me overcome setbacks and stay focused on my studies. We have shared excellent moments together. I greatly value their friendship and I deeply appreciate their belief in me.

Abstract

Identification of Cardiac Signals in Ambulatory ECG Data

Mahdi Torabi

Thesis submitted for the degree of Doctor of Philosophy

School of engineering and computer science

September 2018

The Electrocardiogram (ECG) is the primary tool for monitoring heart function. ECG signals contain vital information about the heart which informs diagnosis and treatment of cardiac conditions. The diagnosis of many cardiac arrhythmias require long term and continuous ECG data, often while the participant engages in activity. Wearable ambulatory ECG (AECG) systems, such as the common Holter system, allow heart monitoring for hours or days. The technological trajectory of AECG systems aims towards continuous monitoring during a wide range of activities with data processed locally in real time and transmitted to a monitoring centre for further analysis. Furthermore, hierarchical decision systems will allow wearable systems to produce alerts or even interventions. These functions could be integrated into smartphones.

A fundamental limitation of this technology is the ability to identify heart signal characteristics in ECG signals contaminated with high amplitude and non-stationary noise. Noise processing become more severe as activity levels increase, and this is also when many heart problems are present.

This thesis focuses on the identification of heart signals in AECG data recorded during participant activity. In particular, it explored ECG filters to identify major heart conditions in noisy AECG data. Gold standard methods use Extended Kalman filters with extrapolation based on sum of Gaussian models. New methods are developed using linear Kalman filtering and extrapolation based on a sum of Principal Component basis signals. Unlike the gold standard methods, extrapolation is heartcycle by heartcycle. Several variants are explored where basic signals span one or two heartcycles, and applied to single or multi-channel ECG data.

The proposed methods are extensively tested against standard databases or normal and abnormal ECG data and the performance is compared to gold standard methods. Two performance metrics are used: improvement in signal to noise ratio and the observability of clinically important features in the heart signal. In all tests the proposed method performs better, and often significantly better, than the gold standard methods. It is demonstrated that abnormal ECG signals can be identified in noisy AECG data.

Keywords: ECG signal processing, principal component analysis, Kalman filter, ambulatory ECG.

Table of contents

List of figures

List of tables

Abbreviations

AB Apnea Bradycardia

AEMN Additive Electrode Movement Noise

AF Atrial Fibrillation

AEMN Additive Muscle Noise

APGN Additive Pink Gaussian Noise

AV-node Atrioventricular node

AWGN Additive White Gaussian Noise

BHF British Heart Foundation

BW Baseline Wander

CPR Cardiopulmonary resuscitation

CVD Cardiovascular disease

dB decibles

DB double beat

EA Ensemble Averaging

ECG Electrocardiogram

EDM ECG Dynamical Model

EKF Extended Kalman Filter

EKS Extended Kalman Smoother

EM Electrode Motion

EMN Electromyography Artifact

GE General Electric

GoF Goodness of Fit

HMM Hidden Markov Models

ICA Independant Component Analysis

ICU Intensive care units

KF Kalman Filter

KLT Karhunen Loève Transform

MA Electromyography Artifact

NN Neural Networks

PCA Principal Component Analysis

PF Particle Filter

PM pacemaker

PNI Power Noise Interference

PVC Premature Ventricular Contractions

SA-node sinoatrial node

SB Single beat

SKF Switching Kalman Filter

SNR Signal-to-Noise Ratio

SVM Support Vector Machines

TWA T wave Alternans

UK United Kingdom

WD Wavelet denoising

WF Wiener Filtering

WPRD Weighted Percentage Root-mean-square Difference

Chapter 1

Introduction

1.1 Motivation

Over the centuries, contagious and infectious diseases such as the tuberculosis, cholera and plague have been the major causes of human mortality. However, from the twentieth century the effects of these diseases dramatically reduced due to the significant advancements in medicine, principally the understanding of contagion and the development of a variety of vaccines and medicines. Since then, conditions of the heart or brain and different types of cancers, have become the major factors in human mortality. Primary prevention, control and risk factor reduction are the most cost-effective steps to reduce the incidence of such chronic and acute diseases; and to increase life expectancy (Jousilahti et al., 2016).

In the twenty first century, one of the most common health problems globally is Cardiovascular disease (CVD), which affect millions of people. CVDs take the lives of 17.7 million people each year which is an estimated 31% of all mortalities worldwide (World Health Organazasion, 2018). In England the mortality due to the CVDs was 152,465 people aged less than 75 years old in 2017, corresponding to about 200 people dying every day. The annual statistical report of British Heart Foundation (BHF) in 2018, estimates that more than 7 million are living after heart failure in the United Kingdom (UK) (British Heart Foundation,

2018). The number of young people suffering heart attacks, sometimes fatal, has dramatically increased in recent years (Stanhope et al., 2015).

Monitoring people who are at risk of cardiac problems can save their life. Using an Electrocardiogram (ECG) is a safe, cheap, non-invasive, reliable way to check the heart status, and is frequently used by the Health Service. It is important to extract medically useful information from ECG signals while the subject engages in a range of physical activities, as some conditions only manifest during specific conditions. However, ECG data collected during activity suffers from large amplitude and non-stationary noise that often obscures the heart signal. Many signal processing techniques have been developed to identify and reduce a range of typical noise processes in ECG, leaving a cleaner signal on which to diagnose heart disease. The early recognition of heart problems using ECG reduces the fatalities and increase the chance of successful treatment.

A range of instruments have been available for many decades for the monitoring of ECG signals during activity. Historically, Holter devices have been worn by participants for 24 hours at a time. These book-sized instruments are often worn on the belt and record ECG signals. Generally, the instrument is returned to a hospital for the data to be downloaded, processed and interpreted (Shafqat et al., 2004). More modern smart phone based devices allow signals to be processed in real-time and transmitted to another site for archiving or interpretation. Despite the improvements of technologies in the field of health monitoring, the ECG signal in ambulatory monitoring systems is difficult to extract from the noisy contaminated signal. This limits the amount of usable data produced by ambulatory systems and also limits the amount of decision making that can be devolved to the mobile device. For instance, if anomalous heart signals can be identified in noisy signals, then these can be instantly flagged to a remote clinician and samples of data can be sent for further scrutiny.

1.2 Aim

The main aim of the thesis is to provide an effective filtering method to remove noise from Ambulatory Electrocardiogram (AECG). ECG signals recorded over periods of hours or days are contaminated by a wide range of noise sources, both biological and non-biological, from within the equipment or from other sources. Typically these noises include 50 Hz electric field interference which can disrupt the recorded heart signal (Youseffi and Achilleos, 2015). The different types of noise on the AECG recordings are discussed in detail in Section 2.3.1. It is not unusual for portions of the recorded AECG data to be so contaminated by noise that no information related to the heart signal can be extracted. It is important that the ECG signal is estimated but also that its reliability is quantified.

In this thesis, both normal and abnormal cardiac signals are considered. Several different abnormalities are considered including Premature ventricular contractions (PVC), Atrial Fibrillation (AF) and T wave Alternans (TWA). Normal signals provide the initial test of proposed filtering methods. If the methods distort normal signals then they are unlikely to be of use when identifying abnormal signals. However, the significant clinical applications involve the identification of abnormal signals, such as AF and TWA. These often occur only for short periods or under specific circumstances, with long periods of normal signal between incidences. Therefore, signal filters should not distort normal heart signals and also separate abnormal signals from noise generated by a wide range of processes, some produce signals with many of the characteristics of abnormal heart signals. It is important that filtering and interpretation methods are able to deal with different types of arrhythmias.

The majority of ECG systems record the heart signal from multiple electrodes simultaneously. Each electrode (channel) provides a different view of the temporal changes in the heart electric field. However, theoretical consistencies between channels allow some discrimination between heart signals and noise.

1.3 Overview of the Thesis and Contributions

This thesis has six chapters, plus a glossary to technical terms and appendix. In figure 1.1, the conceptual links between different chapters of this thesis are depicted. Chapter 2 provides the necessary background of cardiovascular system, signal processing and noise on ECG signals. Chapter 3 includes a literature review of ECG signal processing identified a family of Extended Kalman Filter (EKF) and Extended Kalman Smoother (EKS) algorithms as the current gold-standards: EKF2/EKS2 and the variants with extensions 6 and 17. Chapter 4 suggests a new denoising framework called PCAKF. The modified version of PCAKF algorithm is provided in chapter 5. Conclusions and future works are presented in chapter 6. Each chapter is ended with a summary of the work in that chapter.Below is a summary of each chapter:

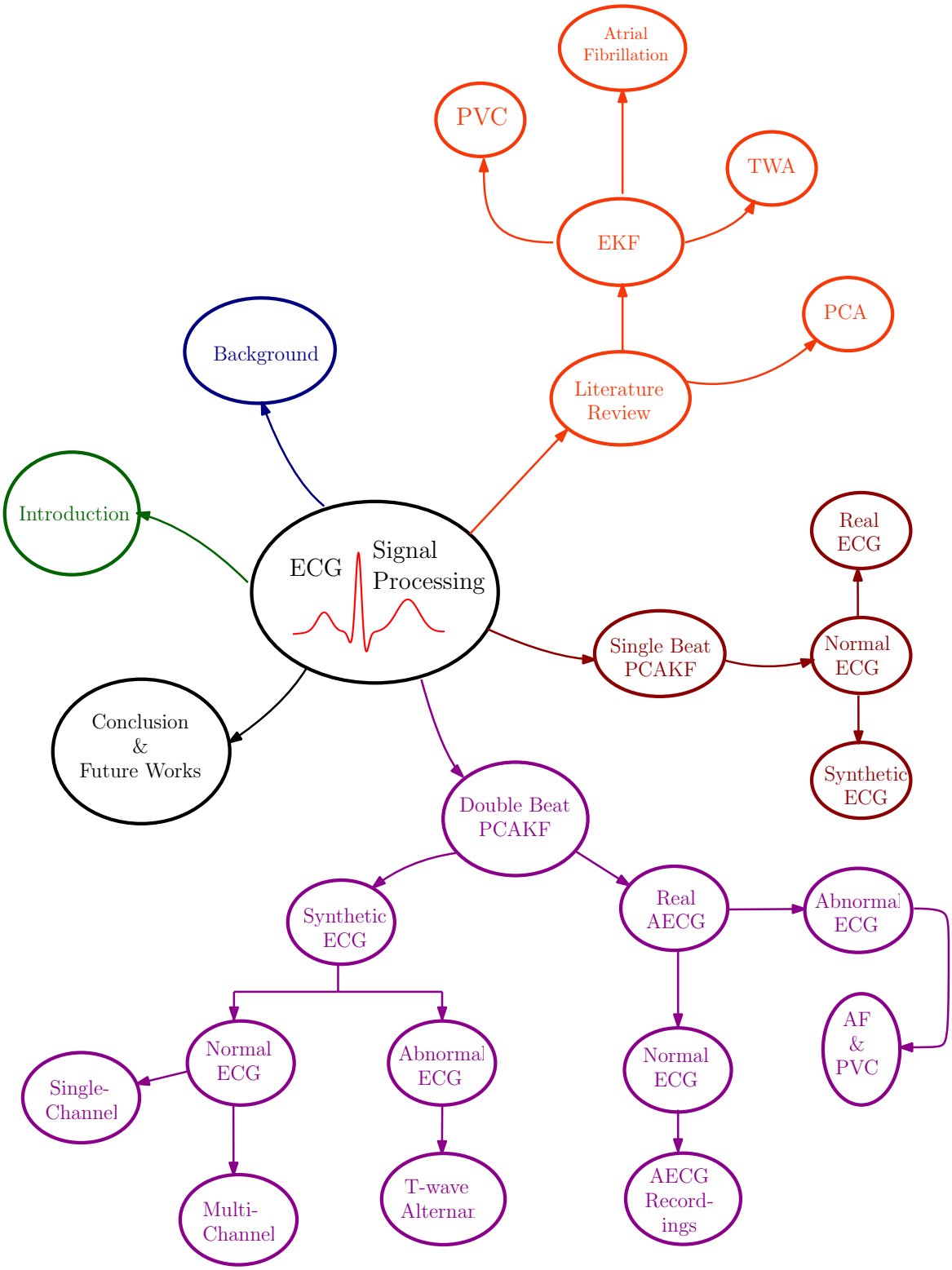


Fig. 1.1 Main and the subtopics covered in this research.

1.3.1 Chapter 2: Background

Chapter 2 introduces the anatomy and electrical function of the heart, the morphology of ECG signals is presented along with a review of ECG noise characteristics and a brief introduction to ECG signal processing.

1.3.2 Chapter 3: Literature Review

Chapter 3 introduces models of ECG signals using sum of Gaussians. This signal parameterisation is fundamental to the gold standard Extended Kalman Filter (EKF) methods, which are reviewed. These methods are later used as comparators for the performance of methods developed as part of this project. Finally, ECG applications of Principle Component Analysis (PCA) are considered as an alternative to sum of the Gaussian signal parametrisations.

1.3.3 Chapter 4: PCAKF Framework

This chapter develops Kalman Filtering techniques applied to heart signals parameterised as a sum of PCA basis signals. Sum of Gaussian ECG models are non-unique, highly over parameterised and yet yield quite a poor fit to many measured heart cycles. Sum of PCA basis signals provide a parametrisation that is unique, optimally parameterised and a good fit. The sum of PCA model may be used in the extrapolation phase of Kalman filtering in a similar way as sum of Gaussian, but requires extrapolation of the whole heart cycle. The proposed PCAKF framework developed in this Chapter and evaluated in later chapters.

1.3.4 Chapter 5: DB-PCAKF Framework

Several variants of the PCAKF have been developed and tested. The PCAKF method is extended from single channel single beat PCAKF framework to the single and multichannel double beat PCAKF algorithm. The double beat (DB) variants use PCA parametrisations

of two consecutive heart cycles and are shown to be more robust than the Single Beat (SB) methods. The methods are tested using synthetic and real, single and multichannel, ECG signals in subsequent chapters.

1.3.5 Chapter 6: Conclusions and Future Works

The final chapter presents a summary of the findings from the current work and critically evaluates the PCAKF methods against the gold standard EKF methods. Finally, some possible directions of future research are identified.

1.3.6 Appendix A

Multichannel AECG filtering results are provided in this section.

1.3.7 Appendix B

A copy of opinions regarding the suggested filtering method from cardiologists and anaesthetists is provided in this section.

1.3.8 Appendix C

Detailed flowcharts of the provided filtering method have been presented in this section.

Chapter 2

ECG Signals and Processing

2.1 Introduction

This chapter introduces a range of topics on the physiology of the heart related to the generation of ECG signals, and ECG signal processing.

2.2 Physiology of the Heart

The heart is an efficient muscular organ whose ultimate purpose is to constantly pump blood to all tissues in the body. The heart consists of four chambers, also called compartments. The top two chambers on each side of the heart are similar and called the left and right atria. They receive blood coming from the body or lungs. Blood passes from the atria to the powerful lower chambers called ventricles. These pump blood out of the heart to the rest of the body or to the lungs. The heart is a dual pump. The left side of the heart receives oxygenated blood from the lungs and then delivers it to the body. The right side of the heart receives deoxygenated blood from the body and delivers it to the lungs to become oxygenated. This process is accomplished by mechanical contraction and relaxation of the cardiac muscle tissue in the ventricles. The myocardium contraction processes are linked to electric fields

around the heart which occur in two phases: the depolarization or *systole* cycle and the repolarization or *diastole* cycle respectively. Increasing pressure due to the contraction in ventricles, causes the blood to flow from the chambers into the arteries. Blood leaves the heart and the decreased pressure, due to the relaxation of ventricles, makes room to accept the blood from atria (Hall, 2015). The anatomy of a heart is depicted in Figure 2.1.

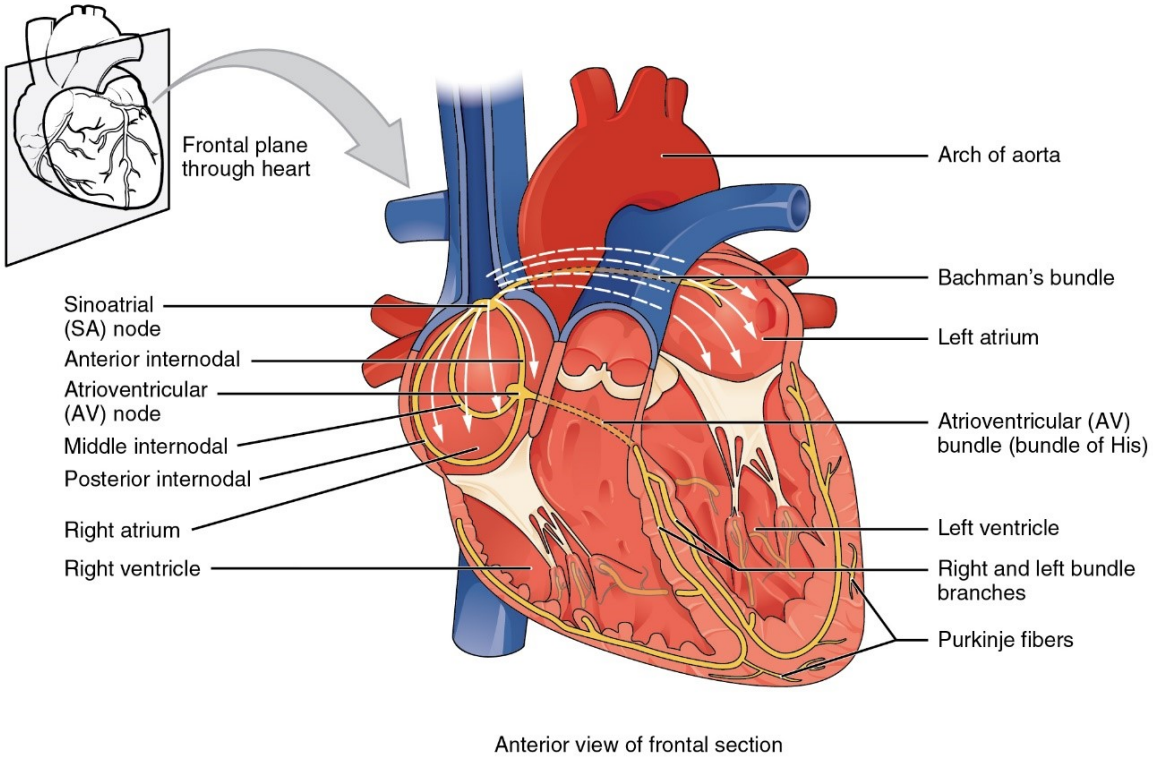


Fig. 2.1 The cross sectional view of the heart, taken from OpenStax (2013)

2.2.1 Electrical Conduction System

The action of the heart is controlled by electrical impulses that travel along nerves. The pumping mechanism of the heart is controlled by a network of nerve fibres which are distributed around the myocardial system, to coordinate the routine rhythmic contraction and relaxation of the cardiac muscle tissue. The main responsibility of the conduction system is to generate an electrical impulse and transfer it in an organized manner through

the myocardium. The strong electrical signal associated with the heart cycle begins in the sinoatrial node (SA-node), which is located in the upper wall of the right atrium. It is the initial source of electrical impulses in the heart and is known as “the pacemaker of the heart”. The SA-node sends a depolarization wave to nearby muscle tissues through the heart. The internodal tracks include the anterior, middle and posterior intermodal pathways. The intermodal pathways transmit and spread the cardiac electric impulses, generated in the SA-node, through both atria. This electric impulse causes both atria to be depolarized. After that, the electric impulse reaches the atrioventricular node (AV-node) located on wall of the right atrium between the upper and lower chambers. Electrical impulses must pass through the AV-node to reach the lower chambers. The AV node is also responsible for controlling the heart rate by slowing the signals, sent by the SA-node, to the ventricles. After the AV-node, the depolarization front arrives at the bundle of His which is the pathway to connect atria and ventricles. The signal then splits into several directions along the Right and Left Branches with Anterior and Posterior deviations. These supply the electrical impulses to the right and left ventricles. The journey ends in the Purkinje fibres or sub-editorial branches, which are located in both the left and right ventricular walls and initiate the ventricular depolarization cycle and ventricular contraction (Hall, 2015, Hampton, 2013).

2.2.2 ECG Morphology

Biological signals are electrical, mechanical or chemical. The activity of the heart muscles leads to measurable electric potential changes on the skin of the chest. A recording of this signal is called an electrocardiogram (ECG). The ECG signal contains very important information about the heart’s condition and function. It is one of the most important sources of information used by physicians to diagnose heart disease and the health of individuals. The ECG signal may be measured using electrodes on the skin. The electrical signal is transmitted via leads to an ECG system which can filter, display and archive data. The typical

ECG signal may be decomposed into P, Q, R, S, T and U waves. This was first suggested by the Dutch physiologist William Einthoven in 1895 (Hurst, 1998). Figure 2.2 shows a schematic of a typical ECG heart cycle with the clinically important features labelled. Each of the six waves may be associated with a phase of the heart cycle. Firstly, the SA node produces a wave of depolarization and spreads the depolarization forward through both upper chambers of the heart via intermodal pathways. This atrial depolarisation causes the P wave. Next, the atria need to repolarize to regain the resting charge. Atria depolarization happens at the same time as ventricular depolarization which generates the QRS complex . After depolarization, the atria begin to repolarize and this is often not observable as it coincides with the high amplitude QRS complex. Next is the T wave, generated by repolarization of the ventricles. In some measurements, a small U wave follows the T wave and is thought to relate to the repolarization of the His Purkinje system, (Hurst, 1998).

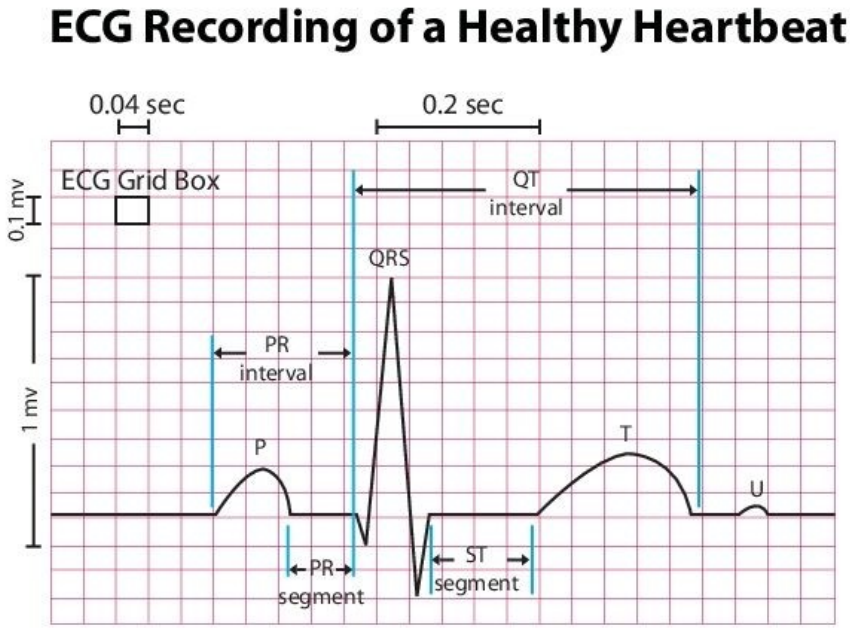


Fig. 2.2 ECG waveform with characteristics waves (Hurst, 1998).

2.2.3 Clinical Analysis of the ECG Signal

P Wave

The P wave duration is about 0.04 to 0.11 seconds. If this wave is wide and dentate, it would indicate enlargement of the left atrium due to mitral stenosis or mitral deficiencies (Garcia and Miller, 2009, Hall, 2015).

QRS Complex

The QRS Complex duration is 0.05 to 0.1 seconds. If this complex is longer than 0.12 seconds, it would be classified as an abnormal wave. The QRS wave amplitude exhibits large natural variation. Very low amplitude QRS complex is considered abnormal and it may indicate diffuse and advanced coronary diseases, ventricular failure, and presence of fluid in the pericardium, myxedema, hypothyroid, as well as obesity, or those with pulmonary emphysema. The amplitude of the QRS complex depends on several factors, including the thickness of the chest wall. A thick chest wall, usually due to obesity, leads to a lower amplitude (Braunwald and Zipes, 2001). A consistently larger QRS amplitude may indicate ventricular hypertrophy. The physical position of electrodes relative to the heart can also effect amplitude (Garcia and Miller, 2009, Hall, 2015).

The QRS complex and ST segment are important for the diagnosis of myocardial infarction (Garcia, 2014).

PR Interval

The PR interval indicates the time required for the signal to be transmitted from sinusoidal node to the ventricular myocardial fibres. The PR interval is the time between the start of the P wave to the start of the QRS complex, and is normally between 0.12 to 0.20 seconds. This PR interval is shorter in children. In adults, a long PR interval may indicate ventricular atrial blockages. A short interval may indicate Wolf-Parkinson-White syndrome, in which there is

an additional conduction path between the atrium and the ventricle (Garcia and Miller, 2009, Hampton, 2013).

ST Segment

This is the segment of the ECG cycle, located between the QRS complex and the beginning of the T wave. This segment begins immediately after the QRS complex, at a point known as the "j point". Amplitude displacement of the ST segment is important for the diagnosis of ischemia and infarction. An increased amplitude of this segment is sometimes seen in normal people, especially black people. A shift to lower amplitudes is known as "ST falling" and may indicate myocardial infarction (Garcia, 2014, Hampton, 2013).

T Wave

The T wave occurs after the ST segment and is asymmetrical: starting with a low gradient but ending in a high gradient. Its peak is curved and has the same polarity as the QRS complex, with both depending upon the placement of the measurement electrode. Sharp or dentate T waves are usually abnormal. An extended T wave is sometimes seen in Hypokalemic infarction (Garcia, 2014, Hall, 2015).

QT Interval

The interval from the beginning of Q to the end of the T wave is the full duration of ventricular systole and diastole, (this includes: the QRS complex, the ST segment and the T wave). The QT duration varies with heart rate, age and gender. Changes within an individual can indicate disease. For example, increased concentrations of potassium or calcium in the blood will lengthen this interval. It also increases in congestive heart failure (CHF), myocardial infarction (MI), hypoxemia, and the use of drugs such as Quinidine and Procainamide (Garcia, 2014, Hall, 2015).

U Wave

The polarity of the U wave is usually the same as that of the T wave. In Hypokalaemia, the U wave is more pronounced. In the ischemic myocardium, it is inverted. In myocardial ischemia, drugs such as digitalis, quinidine, epinephrine and diseases such as thyrotoxicosis may increase the amplitude of U wave (Garcia, 2014, Hall, 2015).

2.3 Biomedical signal processing

The processing of biomedical signals may be required to identify clinically important features. Accurate measurement and feature extraction may be critical steps in diagnosis (Cohen, 1983). Filtering is a basic tool in biomedical signal processing and may be required to separate wanted and unwanted signals, and reducing contamination with extraneous signals and artefacts. The following subsections introduce the processing of ECG signals.

2.3.1 Noise on ECG

Some conditions which can affect ECG signal recording and make the ECG interpretation task difficult are:

- Poor electrical connection between electrodes and the skin due to dirt, hair or variable skin hydration;
- Defective leads connecting electrodes to processing and recording equipment;
- Electrical interference from within the ECG system e.g. mains hum; or from outside the system e.g. power transmission lines, a pacemaker or invasive equipment such as roller pumps (pulmonary and cardiac pumps, etc.);
- Incorrect ECG electrode positioning or mismatch between electrode and lead. Inappropriate placement of ECG electrodes may cause inversion of waves or misinterpretation

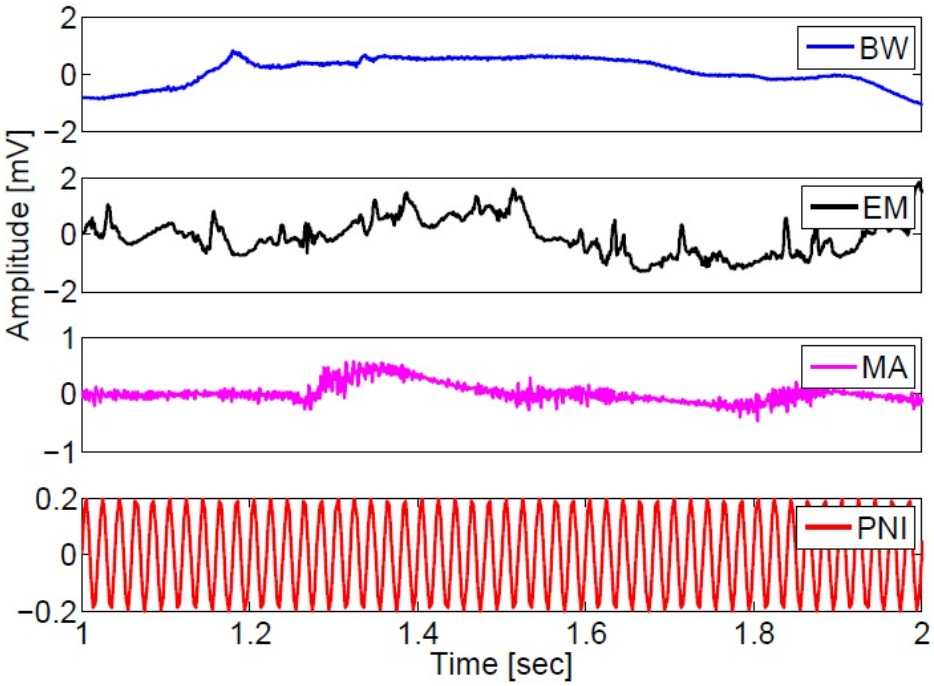
of a measured wave as abnormal. For example, if the precordial lead is much higher or much lower than its proper position relative to the left or right ventricles, left ventricular hypertrophy cannot be seen with electrocardiogram;

- Movement and talking during ECG measurement generates unwanted electrical signals which combine with the desired ECG signal;
- High-intensity exercise before ECG measurement can distort the ECG signal;
- Excitement or deep breathing during ECG test can also distort the signal.

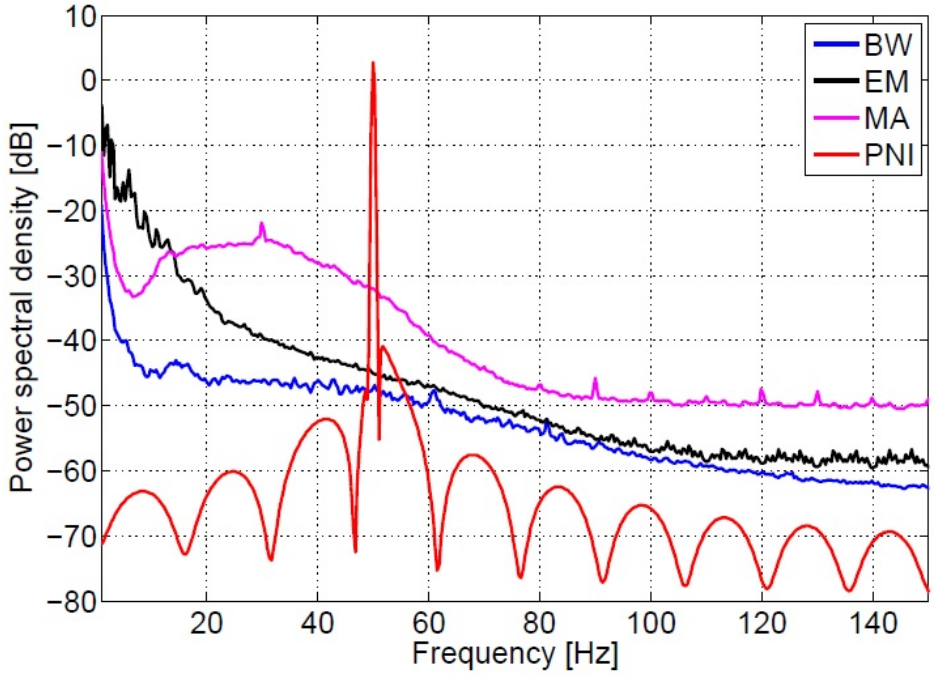
Some common classes of noise in measured ECG signals include: Baseline Wander (BW) and electromyography artifact (MA or EMN), Electrode Motion (EM) and Power Noise Interference (PNI). AECG signals are commonly corrupted with intervals of high amplitude MA, EM and BW noise. Reduction or removal of these noise classes is difficult as the interfering signal overlaps the cardiac signal in both time and frequency, while often having a higher amplitude. Figure 2.3 shows the frequency and temporal overlap present in data from the Physionet Noise Stress Test Database in both time (a) and frequency domain (b), (Behar et al., 2013). For example, the spectral range of MA is from 0.1 to 100 Hz, which overlaps with that of cardiac signals and so these cannot be separated by frequency filters. Therefore, there is an urgent need to use innovative signal processing techniques to increase the visibility of cardiac signals in the presence of MA. These methods should reduce artifact power while preserving the ECG morphological information important for medical diagnosis.

2.3.2 ECG Signal Processing

Many signal processing methods have been proposed to remove contaminates from ECG signals. Ensemble Averaging (EA) is one of the most common filtering methods used for extraction of cardiac characteristics from a noisy ECG signal. However, this method requires



(a) Time Domain



(b) Frequency Domain

Fig. 2.3 BW is baseline wander, EM is electrode motion, MA is muscle artifact and PNI is the power noise interference; reproduced with permission from (Behar et al., 2013)

a great number of heartbeats for averaging. Intermittent cardiac anomalies and important changes in heart rhythm cannot be identified using this method (Lander and Berbari, 1997).

Wiener Filtering (WF) optimizes the Minimum Mean Squared Error (MMSE) when a known cardiac signal of unknown amplitude is contaminated with stationary noise. However, apriori the cardiac signal is not known. Furthermore, cardiac signals are non-stationary. Researchers have proposed the use of WFs in two-dimensions i.e. time-frequency domain or time-scale domain (Kestler et al., 1998, Lander and Berbari, 1997). Adaptive filters (AF) effectively convolve with a time varying kernel and have been shown to reduce EM and MN noise, baseline drift and electrical interference (Laguna et al., 1992).

Wavelet denoising (WD) is a common noise removal technique as it can be applied to non-stationary signals and noises, particularly signals that scale, as cardiac signals do, with heart rate. In these methods, the base wavelet is similar to the cardiac signal in the frequency or temporal domains (Agante and De Sá, 1999, Kestler et al., 1998). Then, the denoising process is usually performed based on a soft thresholding rule on the output coefficients of wavelet transform in different scales or classes.

Artificial neural networks have also been used to extract the statistical content of cardiac signals immersed in high frequency noise. However, these networks require a global supervised learning algorithm and large amounts of training data similar to the data to be filtered (Clifford et al., 2001, He et al., 2006).

All filtering methods rely upon prior knowledge of some characteristic of the cardiac signal. Some of these are universal, such as the periodicity with heart rate, but most are determined by the individual and specific ECG measurement e.g. electrode placement. Many filtering techniques rely upon a flexible ECG Dynamical Model (EDM) that can be adapted to a particular ECG dataset. This requires the dataset to be split into a training section where the EDM parameters are determined, and a test section where the EDM based filter is applied.

In some cases the test section can overlap the training section. The post-filtering EDM may be used to identify clinically important features in the data.

Bayesian model-based statistical methods have also been proposed to separate cardiac signal and noise (Akhbari et al., 2013b, Hesar and Mohebbi, 2017a, Roonizi and Sassi, 2016, Sameni, 2008a, Sameni et al., 2007b, Sayadi and Shamsollahi, 2008). These methods are based on the most widely used dynamic ECG models: the McSharry Cartesian nonlinear dynamic model (McSharry et al., 2003) and its derivatives in polar coordinates.

These methods use Extended Kalman Filters (EKF), which are a generalized version of the standard Kalman filter. Nonlinearities in the system response are addressed by local linearization around the current system state. EKF filters for ECG signals have dominated for most of the last decade and many studies have shown their effectiveness in the presence of non-stationary additive Gaussian noise. In many cases they yielded better Signal-to-Noise ratio (SNR) improvements than WD and AG methods applied to the same data. EKF methods can yield artifacts when applied to signals with non-stationary non-Gaussian noises, such as MN. In this case, it is important that the quality metrics applied measure the visibility of clinically important features rather than focus solely on improvement in the signal to noise ratio. Various metrics evaluate how well filtering maintains the important clinical characteristics in the underlying cardiac signal e.g. the WDD, WEDD and MSEWPRD criteria (Manikandan and Dandapat, 2007, 2008, Zigei et al., 2000).

ECG segmentation refers to the identification of the clinically important cardiac signal characteristics introduced in Section 2.2.3. This is an important step in the interpretation of an ECG signal, in forming a diagnosis and choosing a treatment. The efficiency of the ECG segmentation method depends on the fidelity of the filtering system and the underlying EDM.

Kohler et al. (2002) reviews some ECG segmentation techniques, including AF filters, wavelet transforms, neural networks, Hidden Markov Models (HMM), fuzzy logic and Support Vector Machines (SVM). EKF methods based on McSharry EDMs yield signal

parameters directly linked to cardiac signal morphology and clinically important features such as wave amplitudes, centre phases and phase widths (Akhbari et al., 2018, 2013a, Sayadi and Shamsollahi, 2009).

ECG signal segmentation can be improved by processing the EKF filtered ECG signal using nonlinear Bayesian structures, such as a Particle Filter (PF) (Lin et al., 2010). However, the computational cost of such methods is very high compared to linear-Bayesian methods.

Application of statistical denoising techniques such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are also used to compress, classify and remove artefacts from the ECG signal (Barros et al., 1998, He et al., 2006, Kotas, 2006, Palaniappan and Khoon, 2004, Romero, 2010, Sharma et al., 2010). Combined statistical and other algorithms are also proposed based on Neural Networks (NN) and PCA, wavelet-PCA and ICA, PCA-ICA methods have been implemented to remove motion artefacts from ECG signals (Clifford et al., 2001, Martis et al., 2013, Sarfraz et al., 2015). Each of these methods has its own merits and limitations. ICA and PCA aim to identify maximally statistically independent signals and orthogonal signals explaining maximum variation respectively. On their own they are unlikely to yield maximal separation of cardiac signal and noise. However, the combination of statistical and AF methods has the potential to be effective in a range of scenarios.

2.4 Problem Statement

Any changes in the rhythm and principal morphology of the cardiac signal are known as arrhythmia or cardiac dysrhythmia (Hall, 2015). A large proportion of cardiovascular disease mortality is linked to arrhythmia. Preventing any death requires the quick and accurate diagnosis of arrhythmia, often from long duration AECG recordings. Some of the heart abnormalities that can be identified in the ECG signals include:

- amount and place of the heart infarct;
- level of myocardial infarction development;
- the diagnosis of heart rhythm disorders;
- the diagnosis of atrial and ventricular hypertrophy;
- pericardial diagnosis;
- acute pulmonary embolism diagnosis.

Anomaly identification systems need to be developed that can cope with long duration and very noisy AECG recordings from a large number of patients. They should also be able to quickly and automatically identify the type of arrhythmia and its occurrence time, in long ECG recordings with nonstationary signals and noises. The availability of accurate ECG denoising and segmentation methods can reduce the incidence of death due to arrhythmia by the identification of heart abnormalities. The dynamics and morphology of the underlying cardiac signal should not be distorted by such system as this could lead to false positive diagnoses.

This thesis presents an ECG denoising method based on the integration of PCA and Kalman Filter (KF). The PCA yields an optimal parameterization of a specific ECG dataset, and is calculated using a training segment of the data. Standard KF is then used to track the weights used in a linear weighted sum of PCA signals model of the underlying cardiac signal. This method is both simpler than the nonlinear EKF methods and computationally less intensive. Later, it is shown to be better in preserving clinically important features of the cardiac signal.

The McSharry-based dynamic polar models have shown their capability in a variety of ECG processing applications, such as denoising, segmentation, etc. However they do not perform well in ECG records corrupted with high levels of noise, as commonly occurs in

AECG data. The proposed integrated PCA-KF methods is shown to perform better than the most recent EKF methods when applied to simulated and real AECG data.

2.5 Summary and Conclusions

This chapter provided a brief introduction to heart physiology, the electrocardiogram and clinical features of ECG signal. A range of ECG signal processing methods are reviewed in terms of denoising with their limitation and challenging issues. In the following chapter, we will discuss previous work on Bayesian methods for ECG denoising, segmentation and diagnosis of abnormalities.

Chapter 3

Literature Review

3.1 Introduction

This chapter reviews previous ECG studies based on Bayesian frameworks, e.g. Kalman filters; and statistical methods such as Principal Component filters. First, a McSharry nonlinear dynamic model for ECG signals is described. Then, ECG filters based on this model are introduced. These include the Extended Kalman filter framework which has dominated the literature for the last decade. Finally, some conventional methods based on statistical methods, particularly PCA filters, are reviewed.

3.2 ECG Dynamic Model

An ECG dynamic model (EDM) was introduced by McSharry et al. (2003). It is based on a set of nonlinear state space equations in Cartesian coordinates. Each of the P, Q, R, S and T waves are modelled as a Gaussian pulse with three parameters: centre, width and amplitude. The EDM is defined by three ordinary differential equations:

$$\begin{cases} x' = \alpha x - \omega y \\ y' = \alpha y + \omega x \\ z' = -\sum\limits_{i \in \{P,Q,R,S,T\}} a_i \Delta\theta_i exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right) - (z - z_0) \end{cases} \tag{3.1}$$

where x, y and z are the state variables and $x' = \frac{\partial x}{\partial y}$ (similarly for y' and z'), $\alpha = 1 - \sqrt{x^2 + y^2}$, $\Delta\theta_i = (\theta - \theta_i) mod(2\pi)$ with $-\pi < \theta = tan^-1(\frac{y}{x}) < \pi$, angular velocity of the trajectory is $\omega = \frac{2\pi}{T}$, and T is the heart cycle period taken to be the time between R-peaks. The parameter, z_0 , sets the virtual ground.

Rotation in the $x-y$ plane defines a notional phase which varies by 2π over each heart cycle, and is equivalent to time. The phase θ can either grow monotonically with time or be used modulo 2π. It can be visualised as a saw-tooth pattern with zeros at R-peaks. Figure 3.1 illustrates an ECG signal and the phase. It is important to note that a valid phase observation depends on quality of R peak detection.

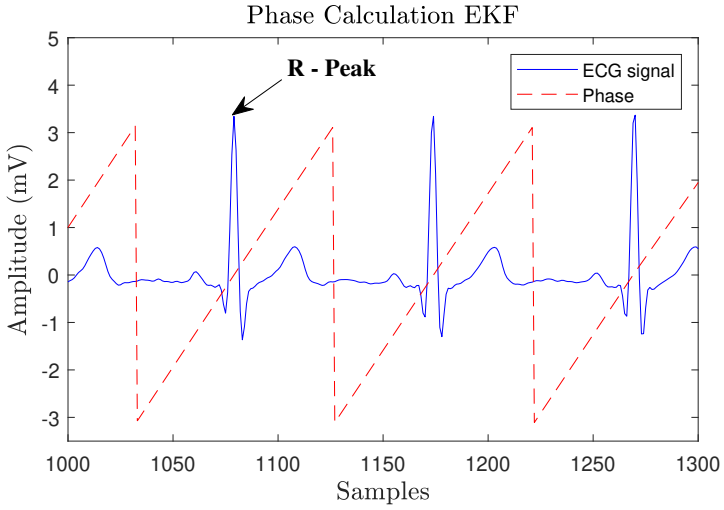


Fig. 3.1 Phase modulo 2π for an ECG signal, repeated using algorithm on Sameni et al. (2007b)

Each Gaussian wave is defined by the parameters a_j, b_j, θ_j for ($j \in P$, P, Q, R, S, T), which correspond to the amplitude, width, and centre phase. A total of 15 wave parameters plus the angular velocity are required. Table 3.1 provides some typical values of the these parameters adapted from McSharry et al. (2003).

Index	P	Q	R	S	T
$\theta_j(rad)$	$-\pi/3$	$-\pi/12$	0	$\pi/12$	$\pi/2$
$a_j(mV)$	1.2	-5.0	30	-7.5	0.75
$b_j(mV)$	0.25	0.1	0.1	0.1	0.4

Table 3.1 Parameters of the synthetic ECG model in McSharry et al. (2003)

The resulting dynamic moves in a circle in the x-y plane, once per heart cycle, while the z coordinate represents the ECG signal in mV. A range of heart abnormalities can be simulated using the EDM, while Figure 3.2 illustrates a typical cycle. Dotted circles indicate the centres of P, Q, R, S and T waves. The dashed lines reflects the limit cycle of unit radius in x and y coordinates. Methods have been developed to fit EDM parameters to measured ECGs.

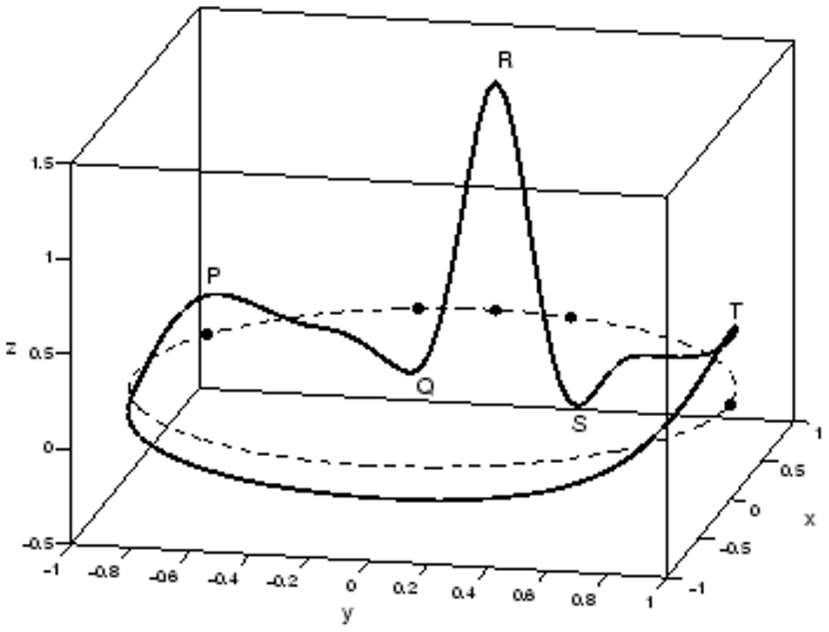


Fig. 3.2 Typical artificial ECG signal (McSharry et al., 2003)

3.3 ECG Signal Processing Using Extended Kalman Filters

This section reviews the EDM-based Extended Kalman filter techniques that have dominated ECG signal processing for the last decade. The McSharry EDM is not linear with respect to the wave centre and width parameters. The standard Kalman Filter (KF) assumes linear variation of the system with respect to the state parameters. To apply KF using an EDM model, requires an extension to the algorithm. Extended Kalman filtering (EKF) uses a linearization of the EDM model around the current state vector. A sequence of EKF methods have been developed tracking ever larger subsets of the EDM parameters. These include methods known as EKF2 and EKF17 (Sameni et al., 2007b, Sayadi and Shamsollahi, 2008). Furthermore, when filtering ECG data off-line, the filters can be applied in both the forward and reverse time directions. When applied in the negative time direction the methods are known as Extended Kalman Smoothing (EKS): e.g. EKS2 and EKS17. Often, EKF is followed by EKS.

3.4 EKF2

The continuous ECG dynamic equations (3.1) need to be discretised to be applied to sampled signals. Sameni et al. (2007b) transformed the EDM model to polar coordinates before discretising:

$$\begin{cases} \theta_{k+1} = (\theta_k + \omega\delta) mod(2\pi) \\ z_{k+1} = -\sum_i \delta \frac{\alpha_i \omega_k}{b_{i,k}^2} \Delta\theta_{i,k} exp(-\frac{\Delta\theta_i^2}{2b_i^2}) + z_k + \eta \end{cases} \tag{3.2}$$

where z_k is the k^{th} ECG signal sample. The angular velocity ω_k may change between samples. The sample period is δ and η is additive white Gaussian noise, which accounts for baseline wander and another additive noises. Sameni et al. (2007a) developed an EKF method with two state variables, θ and z, known as EKF2, with:

$$\mathbf{x}_k = [\theta_k, z_k]^T \tag{3.3}$$

Other parameters are fixed or describe noise and are:

$$\mathbf{w}_k = \left[\alpha_p, ..., \alpha_T, b_P, ..., b_T, \theta_P, ..., \theta_T, \omega, \eta\right]^T \tag{3.4}$$

where η is the measurement noise.

Noise is assumed to be additive and Gaussian. During periods of high noise, the EKF output follows the EDM model more closely while, during low noise periods, the measurements are tracked more closely. The EKF methods rely strongly on an accurate R-peak detection, as well as band-pass pre-filtering to remove baseline wander and high frequency artefacts. The fifteen Gaussian peak parameters are initially fitted using a period of training data. The success of EKF2 filtering relies upon the accuracy of this fit and the stationary of the shape of the cardiac signal. R-peak detection removes the sensitivity to heart rate.

In Sameni et al. (2007b) the EKF frame work is described in detail. A dynamical model of the discrete-time, nonlinear system is represented as follows:

$$\begin{cases} \mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{w}_k, k) \\ \mathbf{y}_k = h(\mathbf{x}_k, \mathbf{v}_k, k) \end{cases} \tag{3.5}$$

where $f(\bullet)$ and $h(\bullet)$ are nonlinear functions describing the evolution of the state vector $\mathbf{x}_k$ and the observation vector $\mathbf{y}_k$, respectively. In equation (3.5) the random variables, $\mathbf{w_k}$ represents the process noise and $\mathbf{v}_k$ is the measurement noise with associated covariance matrices $\mathbf{Q}_k = E\{\mathbf{w}_k\mathbf{w}_k^T\}$ and $\mathbf{R}_k = E\{\mathbf{v}_k\mathbf{v}_k^T\}$. The initial estimate of the state vector is given by $\overline{\mathbf{x}_0} = E\{\mathbf{x}_0\}$ which is assumed to be known, with $\mathbf{p}_0 = E\{(\mathbf{x}_0 - \overline{\mathbf{x}_0})(\mathbf{x}_0 - \overline{\mathbf{x}_0})\}^T$. The linearized approximation of the process and observation models in equation (3.1) are derived around the desired reference point $\left(\hat{\mathbf{x}}_k, \hat{\mathbf{w}}_k, \hat{\mathbf{v}}_k\right)$ (Sameni, 2008a):

$$\begin{cases} \mathbf{x}_{k+1} \approx f\left(\hat{\mathbf{x}}_k, \hat{\mathbf{w}}_k, k\right) + \mathbf{A}_k\left(\mathbf{x}_k - \hat{\mathbf{x}}_k\right) + \mathbf{F}_k\left(\mathbf{w}_k - \hat{\mathbf{w}}_k\right) \\ \mathbf{y}_{k+1} \approx h\left(\hat{\mathbf{x}}_k, \hat{\mathbf{v}}_k, k\right) + \mathbf{C}_k\left(\mathbf{x}_k - \hat{\mathbf{x}}_k\right) + \mathbf{G}_k\left(\mathbf{v}_k - \hat{\mathbf{v}}_k\right) \end{cases} \tag{3.6}$$

where $\mathbf{A}_k$ and $\mathbf{F}_k$ are the Jacobian matrices of partial derivatives of f with respect to $\mathbf{x}$ and $\mathbf{w}$ respectively. $\mathbf{C}_k$ and $\mathbf{G}_k$ represent the Jacobian matrix of partial derivatives of f with respect to $\mathbf{v}$:

$$\begin{aligned} \mathbf{A}_k &= \left.\frac{\partial f\left(\mathbf{x}, \hat{\mathbf{w}}_k, k\right)}{\partial \mathbf{x}}\right|_{\mathbf{x}=\bar{\mathbf{x}}_k} \quad \mathbf{F}_k = \left.\frac{\partial f\left(\hat{\mathbf{x}}_k, \mathbf{w}, k\right)}{\partial \mathbf{w}}\right|_{\mathbf{w}=\bar{\mathbf{w}}_k} \\ \mathbf{C}_k &= \left.\frac{\partial f\left(\mathbf{x}, \hat{\mathbf{v}}_k, k\right)}{\partial \mathbf{x}}\right|_{\mathbf{x}=\bar{\mathbf{x}}_k} \quad \mathbf{G}_k = \left.\frac{\partial g\left(\hat{\mathbf{x}}, \mathbf{v}, k\right)}{\partial \mathbf{v}}\right|_{\mathbf{v}=\bar{\mathbf{v}}_k} \end{aligned} \tag{3.7}$$

To simplify the matrix notations, both $\mathbf{F}_k$ and $\mathbf{G}_k$ are usually absorbed into the noise covariance matrices as follow:

$$\mathbf{F}_k\mathbf{Q}_k\mathbf{F}_k^T \rightarrow \mathbf{Q}_k \quad \mathbf{G}_k\mathbf{R}_k\mathbf{G}_k^T \rightarrow \mathbf{R}_k \tag{3.8}$$

In order to implement EKF, a linearization of equation (3.2) is needed. The process and observation models use the first order of Taylor series. The nonlinear model is linearized using equations (3.6) and (3.7) by defining:

$$\begin{cases} \theta_k = F_1(\theta_{k-1}, \omega) \\ z_k = F_2(z_{k-1}, \theta_{k-1}, \omega, a_i, b_i, \theta_i, \eta)_{i\in\{P,Q,R,S,T\}} \end{cases} \tag{3.9}$$

The Jacobian elements with respect to the hidden variables are:

$$\begin{cases} \dfrac{\partial F_1}{\partial \theta} = 1 \quad \dfrac{\partial F_1}{\partial \omega} = \delta \quad \dfrac{\partial F_1}{\partial a_i}, \dfrac{\partial F_1}{\partial a_i} = \dfrac{\partial F_1}{\partial b_i} = \dfrac{\partial F_1}{\partial \theta_i} = 0 \quad i \in \{P,Q,R,S,T\} \\ \dfrac{\partial F_2}{\partial a_i} = -\delta \dfrac{\omega \Delta\theta_i}{B_i^2} exp\left(-\dfrac{\Delta\theta_i^2}{2b_i^2} \right) \\ \dfrac{\partial F_2}{\partial b_i} = 2\delta \dfrac{a_i \omega \Delta\theta_i}{b_i} \left[1 - \dfrac{\Delta\theta_i^2}{2b_i^2} \right] exp\left(-\dfrac{\Delta\theta_i^2}{2b_i^2} \right) \\ \dfrac{\partial F_2}{\partial \theta_i} = \dfrac{a_i \omega}{b_i^2} \left[1 - \dfrac{\Delta\theta_i^2}{2b_i^2} \right] exp\left(-\dfrac{\Delta\theta_i^2}{2b_i^2} \right) \\ \dfrac{\partial F_2}{\partial \omega} = \sum_{i\in\{P,Q,R,S,T\}} \partial \dfrac{a_i \Delta\theta_i}{b_i^2} exp\left(-\dfrac{\Delta\theta_i^2}{2b_i^2} \right) \end{cases} \tag{3.10}$$

Finally, the time propagation and the measurement propagation equations are summarized by Sameni et al. (2007b):

$$
\begin{cases}
\hat{\mathbf{x}}_{k+1}^{-} = f\left(\hat{\mathbf{X}}_k^{+}, \mathbf{w}, k\right)\Big|_{\mathbf{w}=\bar{\mathbf{w}}_\mathbf{k}} \\
\mathbf{R}_k = \mathbf{Y}_k - h\left(\hat{\mathbf{X}}_k^{-}, \mathbf{v}, k\right)\Big|_{\mathbf{v}=\bar{\mathbf{v}}_\mathbf{k}} \\
\mathbf{P}_{k+1}^{-} = \mathbf{A}_k \mathbf{P}_k^{+} \mathbf{A}_k^{T} + \mathbf{F}_k \mathbf{Q}_k \mathbf{F}_K^{T} \\
\mathbf{S} = \mathbf{C}_k \mathbf{P}_k^{-} \mathbf{C}_k^{T} + \mathbf{G}_k \mathbf{R}_k \mathbf{G}_k^{T} \\
\mathbf{K}_k = \mathbf{P}_k^{-} \mathbf{C}_k^{T} S^{-1} \\
\hat{\mathbf{X}}_k^{+} = \hat{\mathbf{X}}_k^{-} + \mathbf{K}_k \mathbf{R}_k \\
\mathbf{P}_k^{+} = \mathbf{P}_k^{-} - \mathbf{K}_k \mathbf{S}_k \mathbf{P}_k^{-}
\end{cases}
\tag{3.11}
$$

Where $\hat{\mathbf{X}}_k^{-}$ and $\mathbf{P}_k^{-}$ are the prior estimates of the state vector and the covariance matrix respectively before using the k^{th} observation, $\hat{\mathbf{X}}_{k\backslash k}$ and $\mathbf{P}_{k\backslash k}$ are posteriori estimates of the state vector and covariance matrix respectively after using the k^{th} observation. In the above, $\bar{\mathbf{w}}_k = E\{\mathbf{w}_k\}$ and $\bar{\mathbf{v}}_k = E\{\mathbf{v}_k\}$.

Figure 3.3 illustrates the application of EKF2 and EKS2 to an ECG signal with additive White noise. Sameni et al used 190 short (30 s) ECG segments from the MIT-BIH normal sinus rhythm data base, Data 16265 (Goldberger et al., 2000, Moody, 2000b) in the experiment.

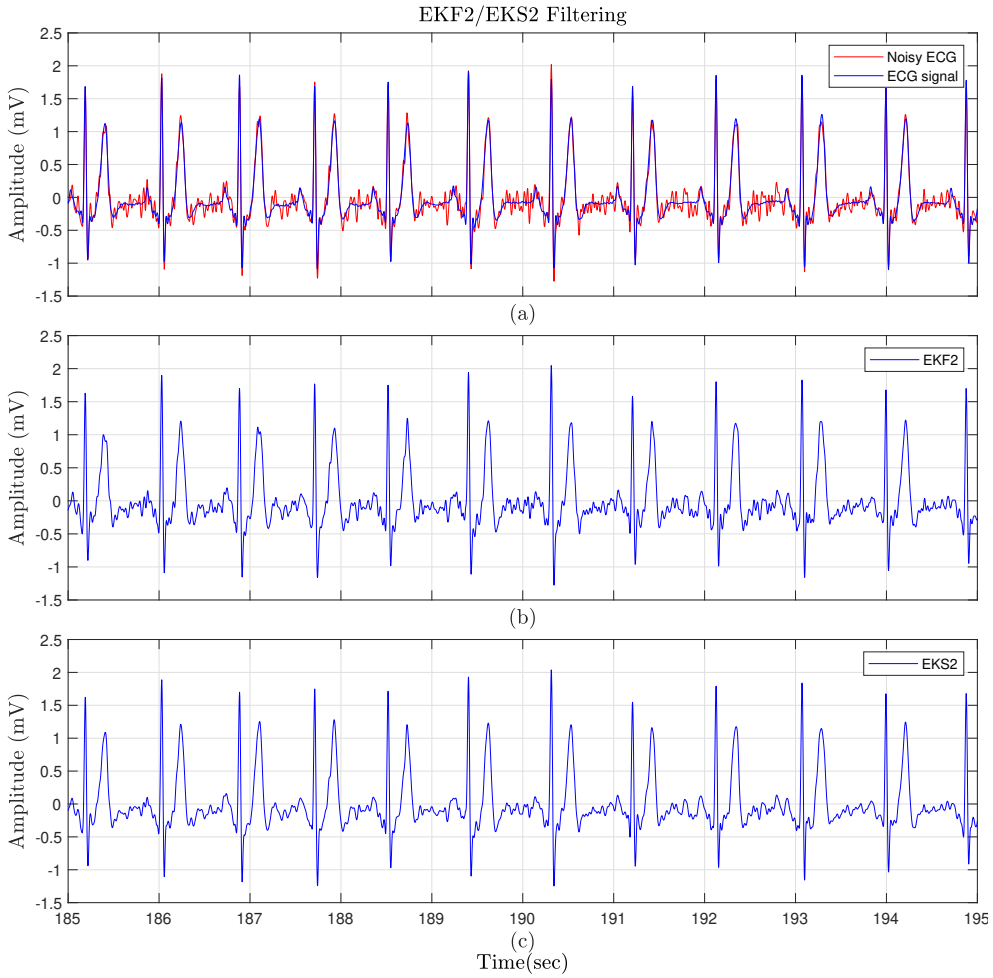


Fig. 3.3 Typical filtering result for EKF2/EKS2 with an input signal with 4 dB AWGN; a) shows the original signal with and without noise, b) is the result after EKF2 and c) is after EFF2 and EKS2.

In, Sameni et al. (2008a) the EKF/EKS methods were used to separate brain and cardiac signals from EEG measurements where the cardiac signal was considered to be noise. EKF2 has also been used to separate fetal from maternal ECG signals from single channel fetal ECG measurements (Sameni et al., 2008b).

3.5 EKF17 and EKS17

Sayadi et al. extended the EKF2 method of Sameni et al. (2007b) by adding the 15 Gaussian parameters as state vector variables, (Sayadi and Shamsollahi, 2008), to produce an algorithm known as EKF17. The linearized state vector update can be written:

$$
\begin{cases}
\theta_k = (\theta_{k-1} + \omega\delta) mod(2\pi) = F_1(\theta_{k-1}, \omega) \\
z_k = -\sum\limits_{i \in P,Q,R,S,T} \omega\delta \dfrac{a_{i,k}\Delta\theta_{i,k}}{b_{i,k}^2} exp\left(-\dfrac{\Delta\theta_{i,k}^2}{2b_{i,k}^2}\right) + z_{k-1} + \eta = F_2(z_{k-1}, \theta_{k-1}, \omega, a_{i,k}, b_{i,k}, \eta) \\
a_{p,k} = a_{p,k-1} + u_{1,k} = F_3(a_p, u_1, k) \\
. \\
. \\
. \\
b_{p,k} = b_{p,k-1} + u_{6,k} = F_8(b_p, u_6, k) \\
. \\
. \\
. \\
\theta_{p,k} = \theta_{p,k-1} + u_{11,k} = F_{13}(\theta_p, u_{11}, k) \\
. \\
. \\
. \\
\theta_{T,k} = \theta_{T,k-1} + u_{15,k} = F_{17}(\theta_T, u_{15}, k)
\end{cases}
\tag{3.12}
$$

where $\mathbf{x}_k = [\theta, z_k, a_{P,k}, .., a_T, b_{P,k}, ..., b_T, \theta_{P,k}, ..., \theta_{T,k}]^T$ represented the expanded state vector. The process noise vector is $\mathbf{w}_k = [\omega, \eta, u_{1,k}, ..., U_{15,k}]^T$ with the similarly expanded

covariance matrix of $\mathbf{Q}_k = E[w_k w_k^T]$.The number of EKF state equation variables has increased to 17:15 Gaussian parameters, the angular velocity and cardiac signal. Sayadi named the new algorithm EKF2+15 or EKF17, and the equivalent smoother EKS17. Figure 3.4 illustrates a typical result applying EKF17 and EKS17 to a measured ECG signal with added White Gaussian noise. Sayadi's experiments showed that EKF17/EKS17 performed denoising better than EKF2/EKS2, particularly for low SNR signals. Figure 3.5 shows the mean and standard deviation of the SNR improvements of EKF17 compared to EKF2 and the Multiadaptive Bionic Wavelet Transform (MABWT) (Sayadi and Shamsollahi, 2007) denoising methods.

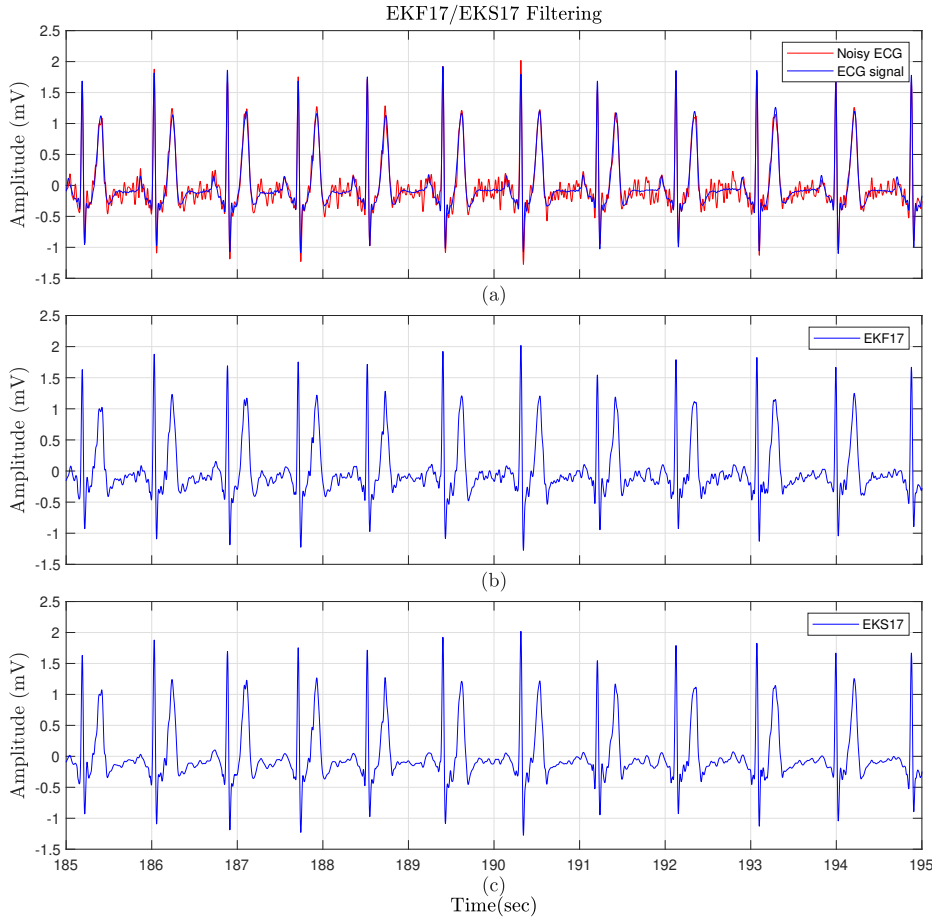
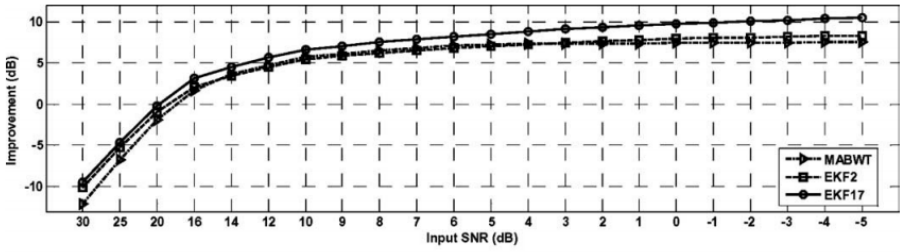
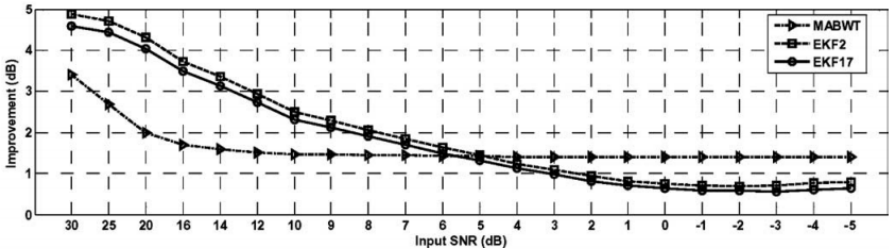


Fig. 3.4 Typical filtering result for EKF17\EKS17 with an input signal with 4 dB AWGN; a) shows the original signal with and without noise, b) is the result after EKF17 and c) is after EFF17 and EKS17.



(a) Original ECG and Noisy Signal



(b) Output of EKF17 on the Noisy Signal

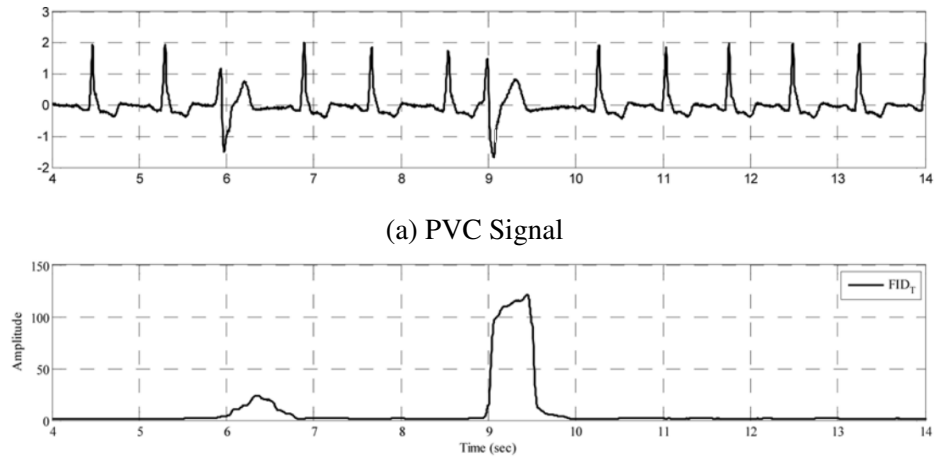
Fig. 3.5 Results of EKF17 and EKS17 using MIT-BHI SNR with 100 time repetition. (a) SNR improvements (b) standard deviation between all 18 subjects.(Sayadi and Shamsollahi, 2008)

3.5.1 Diagnosing Abnormalities Using EKF

ECG denoising using EKF relies upon an apriori model of signal morphology, often obtained by fitting the McSharry EDM to a training subset of measured data. Where the measured signal is considered noisy, the EKF output tracks the model more closely than the measurements. There is a danger that the filter could remove cardiac signal variation that is clinically important. Many common heart conditions lead to irregular signal shapes and arrhythmias e.g. atrial fibrillation (AF), premature ventricular contractions (PVCs) and supraventricular tachycardia (SVT) (Harvard-Medical School, 2014). Variation can be periodic or non-periodic i.e. affecting sequences of heart cycles or occasional single heart cycles. It is important that ECG denoising does not change the real morphology of cardiac signals for the correct diagnosis to result. To identify and classify heart diseases it is important that the duration, amplitude and shape of each ECG wave is preserved. The following sections review the performance of EKF algorithms on abnormal cardiac signals.

3.5.2 ECG Denoising and PVC Detection

The frequent occurrence of Premature Ventricular Contractions (PVC) in an ECG records may indicate serious heart diseases. Sayadi et al. (2010) proposed a variation of EKF with four state variables known as EKF4/EKS4. The cardiac signal was assumed to be the sum of three signals corresponding to the P-wave, QRS complex and T-wave. The state vector contained the phase and the three values of these signals. The grouping of dependent waves reduced the number of variables from EKF17 while allowing the three signals to become time-shifted relative to each other. Further more, Sayadi developed a Fidelity Factor in this research as a measure of the presence of PVC. Normal ECG morphology yields a Fidelity Factor close to zero while abnormal heart cycles yield a value much larger. The Fidelity Factor may be calculated after each heart cycle to yield a near instantaneous alert to PVC. Figure 3.6a illustrates an example of ECG signal with PVC at times $t = 6$ and $t = 9$. Figure 3.6b shows the Fidelity Factor after EKF4 processing with peaks corresponding to the PVC events.



(a) PVC Signal

(b) Fidelity Coefficient

Fig. 3.6 PVC monitoring using the EKF4 (a) ECG signal with PVC (b)fidelity coefficient; adopted from Sayadi et al. (2010).

The method has been tested on 40 ECG segments from the MIT-BIH Arrythmia database (Goldberger et al., 2000, Moody, 2000a). The method detected PVC with an accuracy of 98.43% (Sayadi et al., 2010).

3.6 ECG Denoising and T-wave Alternans Detection

Akhbari et al. (2014) have used the EKF method to detect T-wave Alternans (TWA) (Akhbari et al., 2014). In TWA arrhythmia, the T wave flips between opposing polarity, as can been seen in Figure 3.7.

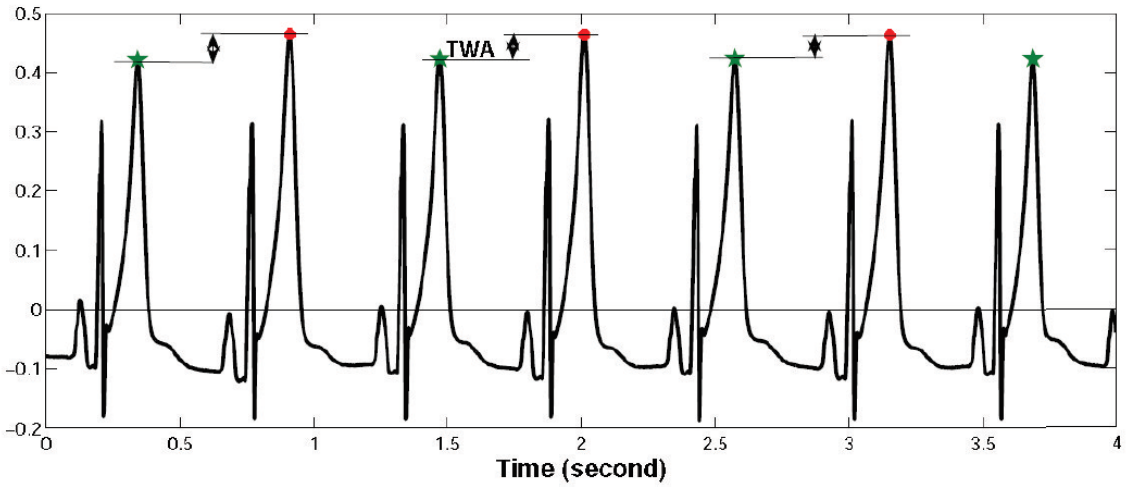


Fig. 3.7 T Wave Alternans, associated with repolarization abnormalities and an increase in serious ventricular arrhythmias (Akhbari et al., 2014).

Akhbari et al. (2014) proposed an EKF variant with six state variables known as EKF6. The cardiac signal is assumed to be the sum of two signals: one formed by the PQRS waves and the second by the T-wave. The state variables correspond to phase and two signal values, and the three Gaussian parameters defining the T-wave. EKF6 was tested on the TWA challenge data base (Goldberger et al., 2000, Moody, 2008) and it was demonstrated that TWA could be identified and T-wave peak times could be estimated to within 8 milliseconds(Paradis et al., 2007).

3.7 EKF and Apnoea Bradycardia

Following further research based on EKF framework, Ghahjaverestan et al. (2015) devised a further variant known as Switching Kalman Filter (SKF) to detect Apnoea Bradycardia (AB), which is associated with decreased heart rate in foetuses.

3.8 EKF and Atrial Fibrillation

Atrial fibrillation (AF) is one of the most common heart arrhythmias that is typified by irregular heart rhythm (Le et al., 2008). Figure 3.8 presents an example where irregular R-R intervals can be observed. Furthermore, the P wave is absent and replaced by fluctuation of the baseline known as fibrillation or f-waves.

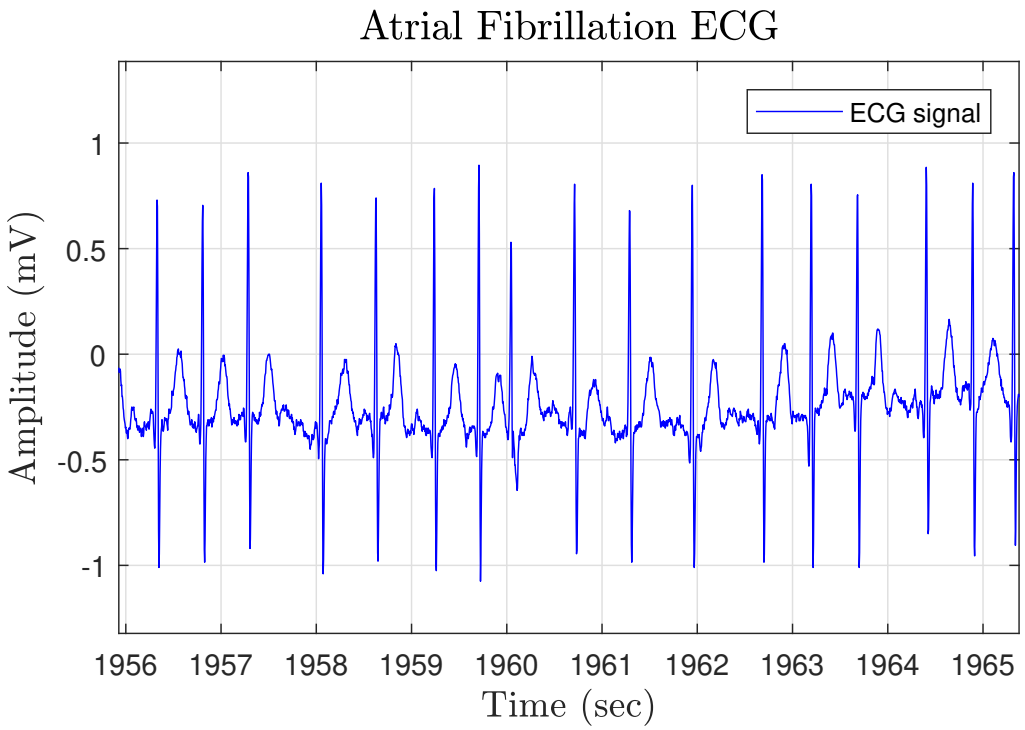


Fig. 3.8 Example of atrial fibrillation. Data 4746 collected from PhysioNet (Goldberger et al., 2000)

Recently, Roonizi and Sassi (2017) adjusted the standard EKF\EKS method for the analysis of AF. The standard sum of Gaussians model was used but with no P-wave component. The model ECG signal was augmented with a trigonometric wave between T and Q waves, with parameters of amplitude, frequency and phase, to represent the f-wave. This yields 17 parameters, i.e. 12 for the four QRST waves, 3 for the sinusoid and 2 for phase and cardiac signal.

The method was tested with normal data from the PhysioNet PTB Diagnostic ECG database (Goldberger et al., 1995). The real data was transformed to mimic AF by the addition of synthetic f waves generated using algorithms devised by Petrenas et al. (2012), Stridh and Sornmo (2001). As the R-R interval is normalised by R-peak detection followed by scaling of the phase, it was considered unnecessary to simulate irregular heartbeats. Figure 3.9 illustrates results of simulated AF using synthetic f wave and real ECG signals.

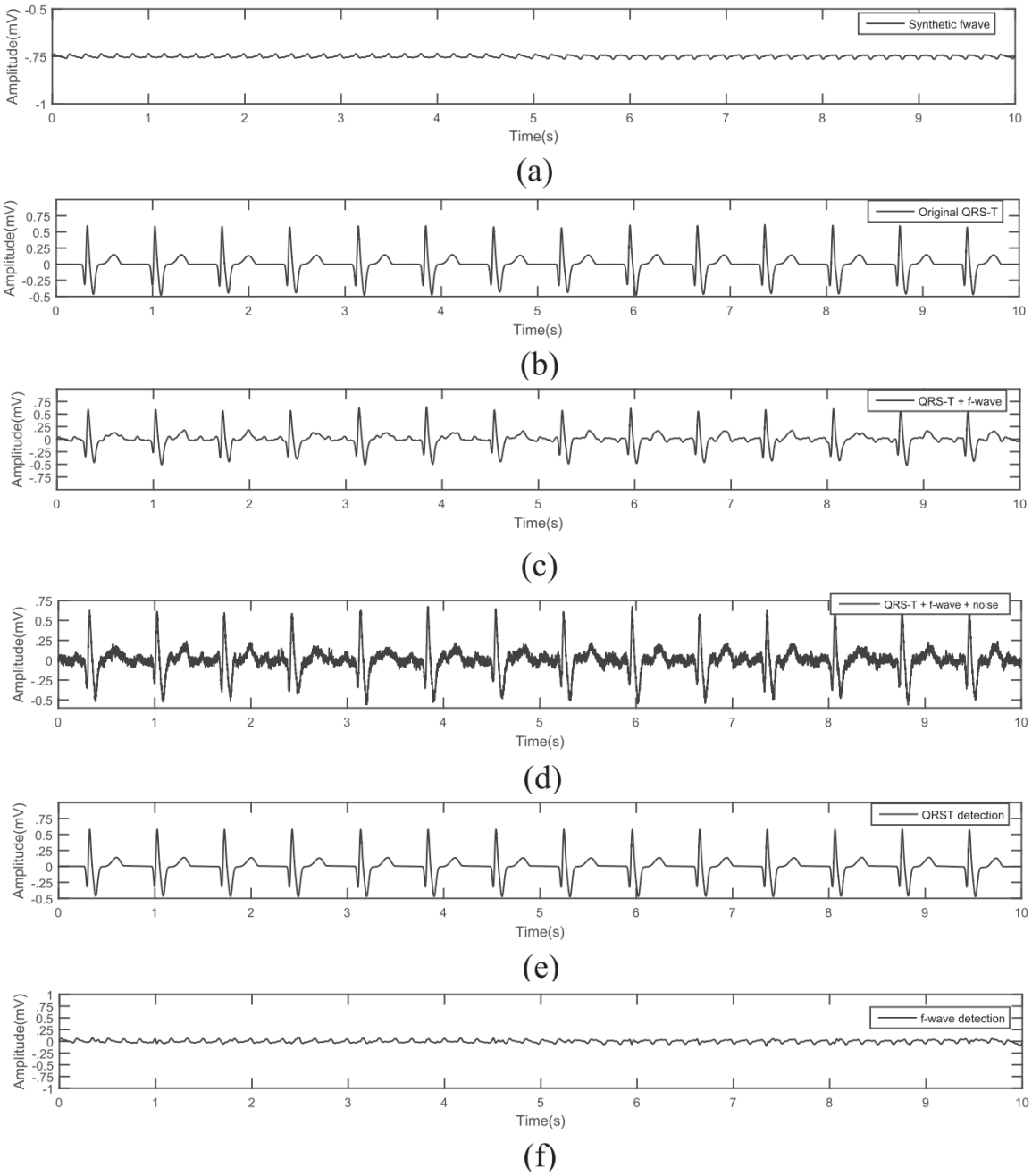


Fig. 3.9 Example of atrial fibrillation. (a) Synthetic f wave; (b) QRST; (c) sum of f waves and QRST signal; (d) Noisy ECG signal; (e) QRS-T extraction using EKS framework; (f) f wave extraction using EKS framework (Roonizi and Sassi, 2017).

The method was also tested on 10 segments of measured ECG containing AF from the AF termination challenge database each with duration of 2 minutes,(Roonizi and Sassi, 2017).

3.9 Use of PCA in ECG signal Processing

Principal Component Analysis (PCA) is a statistical technique able to reduce the dimensionality of data. A set of correlated random variables is expressed in terms of new, uncorrelated variables. The new variables, known as the principal components, are linear combinations of the correlated variables. The principal components are ordered in terms of the proportion of the total variance they explain. The first principal component is the unit norm, linear combination of variables that explains the most variance. The second is the unit norm, linear combination of variables that explains most of the remaining variance. The principal components may be collected into an orthonormal matrix which can be calculated by performing a symmetric singular value decomposition of the covariance matrix of the original variables.

The PCA basis has several desirable properties. The singular values indicate the rank of the set of random variables and so allows redundancy to be identified and eliminated. In this case a vector of random variables can be projected onto the PCA basis and represented by a smaller number of variables. Furthermore, the random variables can be expressed as a linear combination of PCA basis vectors. This known as a Karhunen–Loève transform. This basis can be truncated to approximate the random vector (Maglaveras et al., 1998). The truncated PCA basis is the minimum sized parameterisation that explains a particular proportion of variance. This is often used in data compression. Additive noise is often spread across the PCA basis and so truncated PCA representation can also increase the SNR.

The literature related to ECG denoising using PCA is much smaller than that of EKF methods. However, PCA has been widely used to compress ECG data, and has also been used in denoising and segmentation. PCA was used in Chawla et al. (2006) to identify QRS complex morphology in signals contaminated with muscle noise. Olmos et al. (1999)

employed PCA to compress ECG signals and to filtering additive noise. They used PCA in the reconstruction and denoising of the state space trajectory points. Kotas (2006) presented a denoising algorithm using a robust extension of classical PCA. PCAs of short signal segments were used to filter ECGs while preserving the shapes of the individual heart cycles. In 2010, Yan and Li (2010) applied the Karhunen Loève Transform (KLT) to denoising, compressing and feature extraction of ECG signals. They observe that the KLT concentrates the ECG information into a minimum number of parameters while truncated KLT reduces noise.

PCA techniques also combined with different methods to enhance the ECG signal. PCA is combined with a wavelet transform in Kher et al. (2014) to reduce motion noise in ambulatory ECG. In Rodríguez et al. (2015), PCA is used to identify the QRS complex then a Hilbert transform, combined with adaptive threshold techniques, are used to detect R peaks in noisy ECG signals.

3.10 Summary and conclusions

The EKF method, based on the McSharry sum of Gaussians model of a heart cycle, has become the gold standard ECG denoising algorithm over the last decade. Typically it is applied off-line in the forward time direction (EKF) and then in the reverse time direction (EKS). Many variations exist where the McSharry model is adjusted to better match some medical condition, so that clinically important parameters can be identified.

The McSharry model has 15 shape parameters, each related to a specific feature within the cardiac signal. However, the parameterisation is not unique and this leads to instabilities when the algorithms are applied to long-term data. PCA yields a linear description of heart cycle variation that explains the maximum variation for the least number of parameters. As a PCA based heart cycle model is linear, standard Kalman Filter may be used. Further, the reduction in model parameters simplifies both the algorithm and computation.

In later chapters, a combined PCA and KF method will be developed, known as PCAKF. It will be applied to ECG data from standard databases and the denoising performance will be compared to that of EKF2\EKS2 and EKF17\EKS17. The EKF2 framework was adopted with permission from Dr Sameni and extended to EKF 17. The algorithms are used as bench mark methods during this study.

Chapter 4

Principal Component Analysis Kalman Filter for Single Channel ECG

4.1 Introduction

Previous chapters have introduced ECG signals and their importance in diagnosing heart conditions. The McSharry ECG Dynamic Model (EDM) provided a method that allowed the simulation of ECG signals for a wide range of individuals, electrode positions and heart conditions. The EDM was the basis for a wide variety of Extended Kalman Filters applied to ECG denoising. EDM parameters can be grouped and tracked to yield different EKF methods applicable to different conditions. All rely upon a training period of ECG data to identify the starting EDM parameters. However, the McSharry EDM is over-parameterised and this leads to instabilities. A minimum parameterisation can be derived from a PCA applied to a training set of data. Using a PCA description of the ECG, heart cycle by heart cycle, allows standard Kalman filtering to be used to both track and filter the ECG signal. This chapter develops the PCA based KF for ECG denoising and compares its performance to standard EKF methods presented in previous chapters.

4.2 Methodology

This section develops the PCAKF method. The algorithm is based on a combination of statistical and adaptive methods. Standard Kalman Filtering (KF) is used to track a single channel ECG signal. Instead of the McSharry sum-of-Gaussians EDM heart-beat model, each beat is approximated by a weighted sum of a small number of principal component (PCA) basis signals. The PCA basis signals are calculated from a short period of training data and may be augmented with signals specific to particular cardiac conditions. The KF tracks the basis weights from heart-beat to heart-beat. The method includes three different phases. First, the ECG signal is passed through the standard pre-processing filters to remove common noises from the recorded signal, such as baseline wander. Second, a small part of the pre-processed ECG signal, known as training data is used to calculate the PCA basis signals. Finally, the pre-processed ECG signal and PCA basis signals are used in the Kalman filter to track and denoise the ECG signal. This new method has been called PCAKF. The flowchart in Figure 4.1 illustrates the overall algorithm used to test the PCAKF filtering method on signals with additive synthetic and measured noise. The output of PCAKF, along with EKF methods, is compared with the original ECG signal from both a signal power and clinical viewpoint. The detailed version of the flowcharts for the PCAKF algorithm are provided in appendix C.

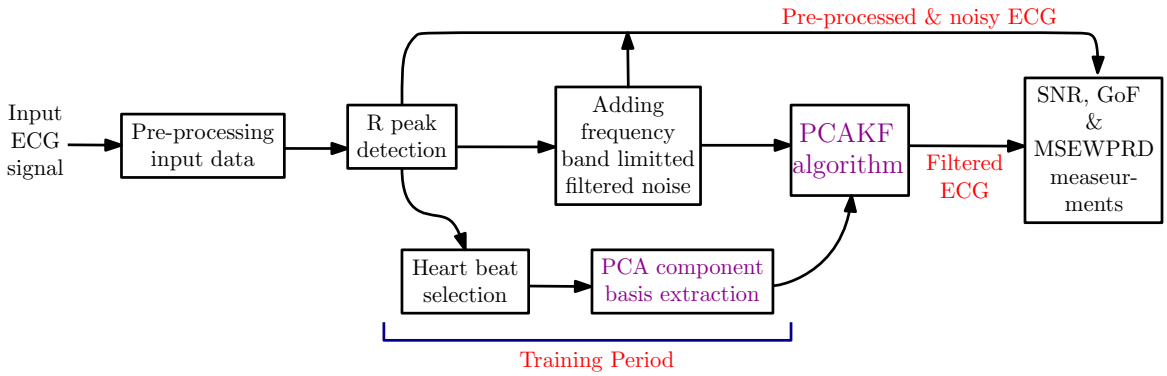


Fig. 4.1 The process flowchart used to implement and test the PCAKF algorithm.

4.2.1 Baseline Wander Removal and Pre-processing

This section specifies the standard ECG signal pre-processing used in this project. The aim of pre-processing is to remove noise outside the frequencies spanned by cardiac signals, and to remove clear artefacts such as spikes and step changes in the baseline. The algorithms are those described in Di Marco et al. (2012). They are designed to detect and reduce the number and amplitude of QRS band artefacts, baseline drift, pacemaker (PM) spikes, spurious step changes and high frequency noise; often due to interference from other electronics. A linear filter is used to reduce noise outside the ECG signal range from 0.5 to 30 Hz.

4.2.2 Estimation of the PCA Model

The proposed method is based on an approximation of the ECG signal over each heart cycle, based on a linear combination of basis signals derived from Principal Component Analysis (PCA) of a sample of the signal. These basis signals are calculated from a training period of clean data acquired under ideal conditions. This has several advantages over the sum-of-Gaussians model. Firstly, a minimum number of basis signals is required to fit a known proportion of signal variation. Secondly, there is a linear relationship between basis weights and the signal, and so a set of parameters yields a unique signal.

After pre-processing, a standard R-peak detection algorithm is applied to identify heart cycle boundaries. Each heart cycle is interpolated onto a fixed number of samples. In this project, samples were used for each heart cycle. The PCA basis was chosen to be the Eigen-vectors of the training period heart cycle covariance matrix, with the largest Eigen-values. Each element of the 1000×1000 covariance matrix is the expected second moment of two heart cycle samples. The Eigen vectors are the discretized patterns of heart cycle variation that explain the largest proportion of the observed variation over the training period. The signals corresponding to the Eigen vectors, known as the PCA basis functions, generally correspond to changes in P, Q, R, S and T wave, amplitudes and temporal displacements.

In the following paragraphs, notation is introduced which will be used to define the PCAKF algorithm. Consider a training interval where N_T heart cycles are measured. The k^{th} beat time-series is denoted: $ECG^k(\theta)$. The vector $\theta \equiv (\theta_j)$ for $j = 1$ to N_S contains the heart-cycle phases of the measurement time-series, assumed to vary linearly from $-\pi$ to $+\pi$ over each heartbeat. Each filtered ECG signal, over each heart cycle, is approximated by the weighted sum of N_B basis signals, with weights W_i^k, for $i = 1$ to N_B.

$$ECG^k(\theta) \cong \sum_{i=1}^{N_B} w_i^k PCA_i(\theta) \tag{4.1}$$

When the training data from a clean ECG signal is identified and divided into N_T heart-beats, an uncentred covariance matrix for heart-cycle data is calculated. From the covariance matrix, the PCA basis signals $PCA_i(\theta)$ are calculated. The calculation of the PCA basis relies upon the quality of pre-processing, baseline removal and R-peak identification. In practice, some selection is required to remove heart cycles strongly contaminated by noise and artefacts, such as those in Figure 4.4. This selection is automated by rejection of heart cycles that differ too much from a mean or target cycle. The more and cleaner the available training data, the lower the noise in the PCA basis signals. Pre-processing of the data before PCA decreases this noise. Irregularity in the heart rate is largely normalized by R-peak detection and heart cycle identification.

4.2.3 Kalman Tracking of Single Channel PCA Weights

In the proposed method, Kalman Filter (KF) is applied to track the ECG basis weights w_i^k as in equation (4.1), from one heartbeat to the next. This is achieved by a combination of extrapolation from the previous beat weights and correction based on the current observed beat. Standard KF notation is used, as in Simon (2006). For ECG filtering, the state vector is the vector of weights for the single-lead ECG signal:

$$\mathbf{X}_k \equiv (w_i^k) \tag{4.2}$$

This is assumed to be multidimensional Normally distributed with covariance matrix $\mathbf{P}_k$. The extrapolation process assumes that the heart signal does not change from one cycle to the next, i.e. there is no forcing:

$$\mathbf{X}_{k/k-1} = \mathbf{X}_{k-1} + \mathbf{W}_{k-1} \tag{4.3}$$

The noise vector $\mathbf{W_k}$ represents the extrapolation uncertainty due to variation between consecutive heart beats and is assumed to be zero mean and with covariance $\mathbf{Q}_k$. The uncertainty in the extrapolation estimate of the current state vector is a combination of the uncertainty in the last heart cycle and the extrapolation uncertainty:

$$\mathbf{p}_{k/k-1} = \mathbf{P}_{k-1} + \mathbf{Q}_k \tag{4.4}$$

In our case, the state vector is observed by the measurement of a heart cycle contaminated with noise:

$$\mathbf{Z}_k = \mathbf{H}_k\mathbf{X}_k + \mathbf{V}_k \tag{4.5}$$

where $\mathbf{V}_k$ represents the observation noise, which is assumed to be zero mean and with covariance $\mathbf{R}_k$. The matrix of PCA basis vectors has been given an index k to indicate that the PCA basis can evolve in time. In the results presented in this document, the PCA basis is calculated once at the beginning. However, there is the possibility for the PCA basis to be refined iteratively. Due to the simplification of the KF process, the ECG PCA parameters and their uncertainties can be tracked using the following process in equations (4.6) to (4.8).

The Kalman factor $\mathbf{K}$ is calculated by solving:

$$\mathbf{K}_k(\mathbf{H}_k\mathbf{P}_{k/k-1}\mathbf{H}_k^T + \mathbf{R}_k) = \mathbf{P}_{k/k-1}\mathbf{H}_k^T \tag{4.6}$$

then the estimate of the state vector for this heart cycle, and its uncertainty, using the Joseph Form, is calculated using:

$$\mathbf{X}_k = \mathbf{X}_{k/k-1} + \mathbf{K}_k(\mathbf{Z}_k - \mathbf{H}_k\mathbf{X}_{k/k-1}) \tag{4.7}$$

$$\mathbf{P}_{k/k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k/k-1}(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \tag{4.8}$$

Training data are used to estimate the initial PCA parameter uncertainty $\mathbf{P}_1$ and extrapolation uncertainty $\mathbf{Q}_1$. $\mathbf{P}_1$ is the covariance of the training heart cycle PCA weight vectors while is the covariance of the difference between consecutive weight vectors. The difference between the reconstructed heart cycle time series (4.1) and the measured ECG signal is assumed to be noise with mean squared value N^2 and uniformly and independently spread across all samples which is:

$$N^2 = E\left[\left(ECG^k(\theta) - \sum_{i=1}^{N_B} W_i^k H_i(\theta)\right)^2\right] \tag{4.9}$$

This suggests a diagonal uncertainty matrix:

$$\mathbf{R}_k = N^2\mathbf{N}_{N_S} \tag{4.10}$$

4.3 Test Data

To evaluate ECG filtering algorithms, it is necessary to have ECG signals without noise to be the perfect, uncontaminated output. All measured ECG signals are contaminated with some noise, even if the ECG is taken in perfect recording conditions. Noise could be internal to the equipment, internal to the patient, or from the electrode interface. To overcome this, the use of synthetic ECG and noise signals allows the evaluation and comparison of different filtering methods. Both synthetic ECG and real AECG are used to test the proposed method against gold standard methods.

4.3.1 Synthetic ECG

A highly realistic synthetic ECG signal is generated with sample rate of 512 Hz in order to study the quantitative performance of filtering techniques. Generation of synthetic ECGs, based on the three-dimensional canonical model of a single dipole vector of the heart, was proposed by Sameni et al. (2007a). A sum-of-Gaussian model is assumed where the Gaussians represent P, Q, R, S and T peaks. The Gaussian parameters, three for each peak, are each assumed to be samples from Gaussian distributions. A 10% multiplicative variation is generated in each parameter using a multiplier of the form: $(1+0.01\times\varepsilon)$, where ε is an independent sample from a Standard Normal distribution. The ECG signal is the concatenation of heart cycles with similar random variation in the R-R intervals. Figure 4.2 illustrates a multi-channel synthetic ECG signal (Sameni, 2008b).

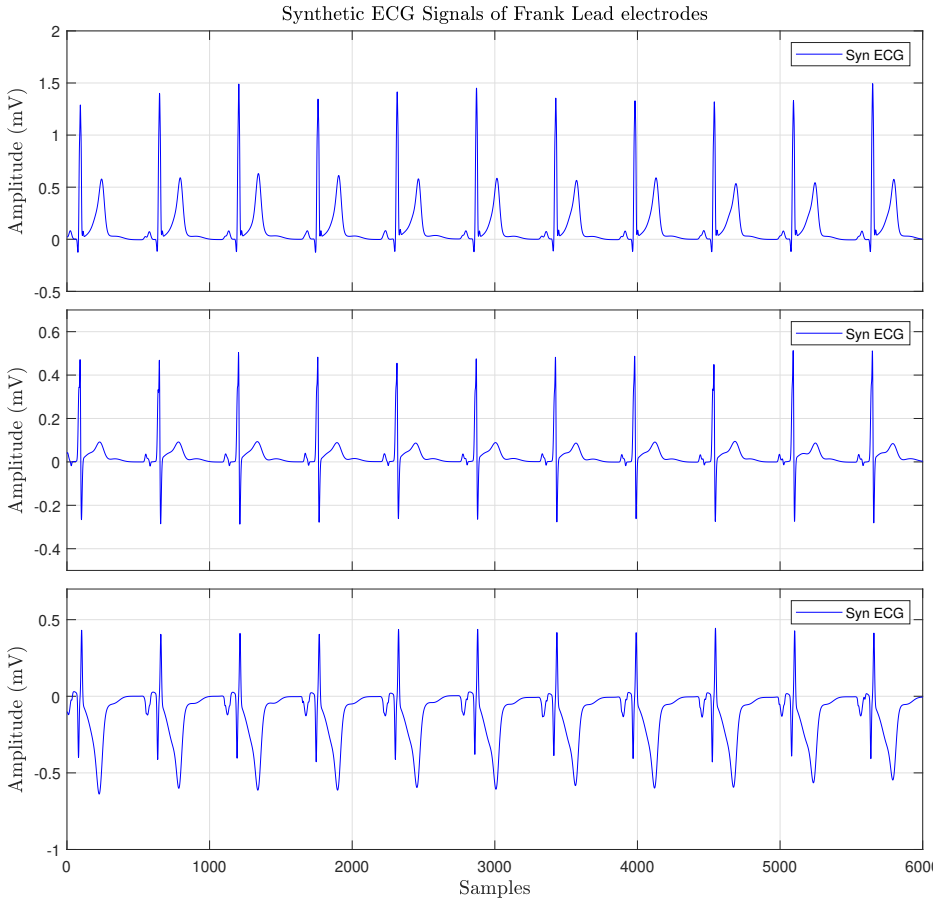


Fig. 4.2 An interval of realistic synthetic multi-channel ECG, (Sameni et al., 2007a).

4.3.2 Data Collection

In the second part of the reported analysis, measured AECG signals were used to study the performance of the proposed method. Short segments of ECG signal were selected from twenty five records measured using an AECG system. Ethical approval was received from The University of Hull, to collect AECG data from healthy male volunteers. A Mortara 12 lead Holter system, with sample rate of 1000 Hz, is used to collect data. Healthy male participants were recruited, with no history of heart disease, aged between 20 and 55 and with a resting heart rate between 70 and 85 beat per minute. Figure 4.3 shows a participant with the electrode array attached to his chest. A data collection protocol was designed to simulate a range of normal daily activities, which included six different physical activities.

First, 10 minutes of data was recorded while the participant was sitting in an upright position without movement. These data are assumed to have as few noise artefacts as is possible. Data was then recorded in five minute intervals for five further activities. we asked the participants to walk normally, while swinging the arms which can add movement and muscle noise to the ECG signal. To collect extremely noisy data, the participants were asked to jog then run. Between all of these activities, we asked the participants to stop and rapidly move their arms up and down, while shaking their body(phase 5). After running and phase 5, the participants were asked to take three rapid deep breath. This breathing exercise was repeated four more times.

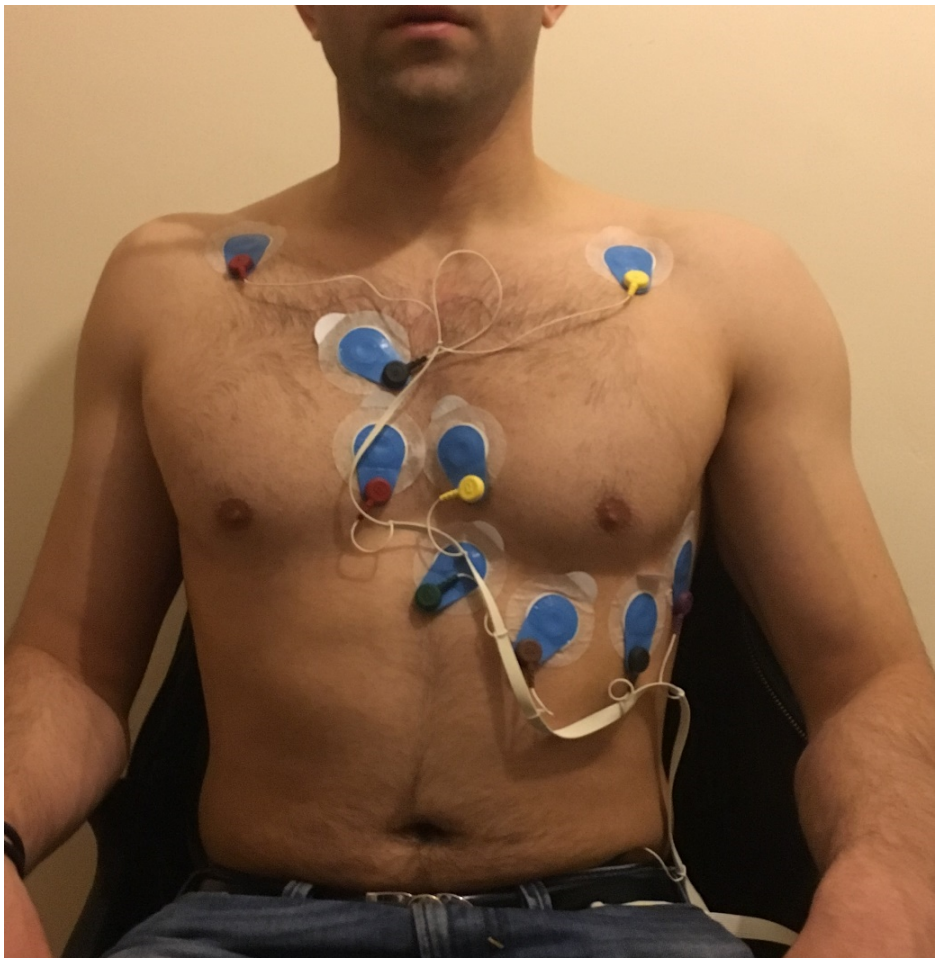


Fig. 4.3 AECG data collection while sitting upright.

4.4 Noise Signals

The performance of different ECG filters can be quantitatively measured by filtering clean ECG signals with added noise. Four types of noise were used in the tests: Additive White Gaussian Noise (AWGN), Additive Pink Gaussian Noise (APGN), measured muscle noise and measured electrode movement noise. Coloured noise has a monotonically decreasing power law spectral density given below in equation (4.11):

$$S(f) \propto \frac{1}{f^{\beta}} \tag{4.11}$$

where $S(f)$ is noise spectral density, f is the frequency in Hz and β is a measure of noise colour. White noise has colour parameter of $\beta = 0$ while pink noise has $\beta = 1$. Pink noise is both more realistic and more challenging for filters. It has more power at low frequencies leading to correlation between noise time samples and stronger overlap with the significant cardiac signal frequencies. To ensure the consistency of the results, the addition of white and pink Gaussian noise was repeated for hundred times in all tests, and summarizing statistics calculated.

Two different types of real measured noise were used in the tests. These were selected from standard recordings by selecting intervals of real muscle artefact (MA) and electrode motion artefact (EM). Noise signals are from the Massachusetts Institute of Technology-Beth Israel Deaconess Medical Centre (BIDMd) and were collected with sample rate of 360 Hz (Goldberger et al., 2000, Moody, 2000c). For this study, the noise signals were resampled to 512 Hz and 1000 Hz. Motion artefacts are generally considered the most troublesome because this type of noise can mimic the appearance of ectopic beats and it cannot be removed easily by filtering methods, as can noise of other types.

4.5 Quality Measurements

Three common metrics of filter performance are used in this chapter. When synthetic ECG signals are used, the signal to noise ratio (SNR) before and after filtering may be calculated. Similarly, the before and after normalised root mean square error (GoF) may be calculated. Both SNR and GoF are measures familiar to engineers. However, signals with the same SNR or GoF can look very different due to the phases and spectral density of the noise. Clinically important parameters may or may not be identifiable in signals with the same SNR or GoF. Therefore, a third metric is used which measures the visibility of clinically important features: the MSEWPRD is specified in Section 4.5.3.

4.5.1 SNR Analysis

The signal to noise ratio (SNR) is an engineering term for the ratio of signal to noise powers, often expressed in decibels (dB). The power in a time-series is often taken to be the sum of squared amplitudes. The SNR improvement produced by an ECG filter can be defined as(Sayadi and Shamsollahi, 2008):

$$\begin{aligned} SNR_{improve}[dB] &= SNR_{output} - SNR_{input} \\ &= 10log\left(\frac{\sum_{i=1}^{N_s}|X_n(i)-X(i)|^2}{\sum_{i=1}^{N_s}|X_f(i)-X(i)|^2}\right) \end{aligned} \tag{4.12}$$

where X represents the pre-processed clean ECG signal, X_f is the filtered signal and X_n denotes the pre-processed noisy segment of ECG signal. A higher SNR improvement indicates better filtering results.

4.5.2 Goodness of Fit for ECG Signals

Further to SNR calculation, the normalised root mean square error (GoF) is calculated as a measure of goodness of fit. GoF is the normalized standard deviation of the difference between original clean ECG and output of the filter under consideration. The Goodness of Fit (GoF) is defined by:

$$GoF = 1 - \frac{\|X_f(i) - X\|_2}{\|X(i) - X\|_2} \tag{4.13}$$

where X is the reference ECG signal, $X(i)$ is the reference signal plus the ith noise signal, while $X_f(i)$ is the filtered ECG signal. The root mean square error is the mean of this value over a large number of tests with different noise signals. Assuming filtering reduces the noise power then GoF varies between 0 and 1 where 0 represents very poor fit and 1 indicates a perfect fit with the original ECG.

4.5.3 MSEWPRD

SNR and GoF are valuable methods to quantify quality of synthetic ECG signals, from a signal processing point of view. However, it is important to demonstrate that ECG filtering retains clinically important features of the signal. These includes intervals on the ECG signal such as the R-R interval, ST segment, QT interval, amplitude and direction of P wave. This allows diagnosticians to read and interpret the ECG signal correctly so that patients can be treated effectively, with the correct procedures and medication. A misinterpretation of the ECG could result in unnecessary testing (which wastes medical funding and may cause substantial discomfort to the patient) or misdiagnosis.

The Multi Entropy-Based Weighted Distortion Measure (MSEWPRD) (Manikandan and Dandapat, 2008), is a measure of how well a filter preserves the clinically important features

in an ECG signal. It compares the features of the original ECG with those of the filtered ECG and a measure of the difference is calculated. MSEWPRD measures the Weighted Percentage Root-mean-square Difference (WPRD) between the sub-band wavelet coefficients of the original ECG signal and the filtered ECG signal, with weights equal to the multi-scale entropies of the corresponding sub-bands. To calculate this metric, signals to be compared must be decomposed using wavelet filters up to W_L levels. The wavelet levels separate the sharp QRS complexes from the smoother P and T waves. The figures reported in this study use Daubechies 9/7 bi-orthogonal wavelet filters with $W_L = 4$ (Hesar and Mohebbi, 2017b). A small MSEWPRD value indicates that clinically important features have been preserved by the filter.

4.6 Results

Two different sets of ECG signals are used in this chapter. First, the performance of the proposed method is studied using highly realistic synthetic ECG to yield quantitative results. Then the analysis is repeated using real ECG data to qualitatively test the proposed method. In both synthetic and real signals, PCA results are studied to find the ideal model to use in the Kalman filter. In the second step, the quantitative results are calculated using short term synthetic ECG signals with additive white and coloured Gaussian noise with different powers. Additive real measured muscle and electrode movement noises were also tested. In the third step, the effectiveness of the proposed method is examined using a typical real ECG signal with additive real noises.

4.6.1 Pre-processing of ECG Signals

Standard pre-processing algorithms are used on both the ECG and additive noise signals. The algorithms are those described in Di Marco et al. (2012) which won the PhyisioNet computing in cardiology challenge in 2011. They are designed to detect and reduce the number and amplitude of QRS band artefacts, baseline drift, pacemaker (PM) spikes, spurious step changes and high frequency noise often due to interference from other electronics. Figure 4.4 shows the real ECG signal mixed with real muscle noise to achieve a SNR of -1 dB. It can be observed that this leads to large baseline wander in the ECG signal. Pre-processing removes this baseline wander and spikes from noisy time series. Figure 4.5 illustrates the signal after pre-processed (blue) with R peaks (red). R-peak detection is a critical step in both EKF and PCAKF filtering methods, and pre-processing is important to correctly identify R-peaks. This example also highlights that it is important that additive noise is pre-processed before mixing, or frequencies outside the cardiac signal are re-introduced. The test protocol performs R-peak detection on the clean signal before adding noise. This means that observed differences between the outcomes after filtering are not due to sensitivity to R-peak detection.

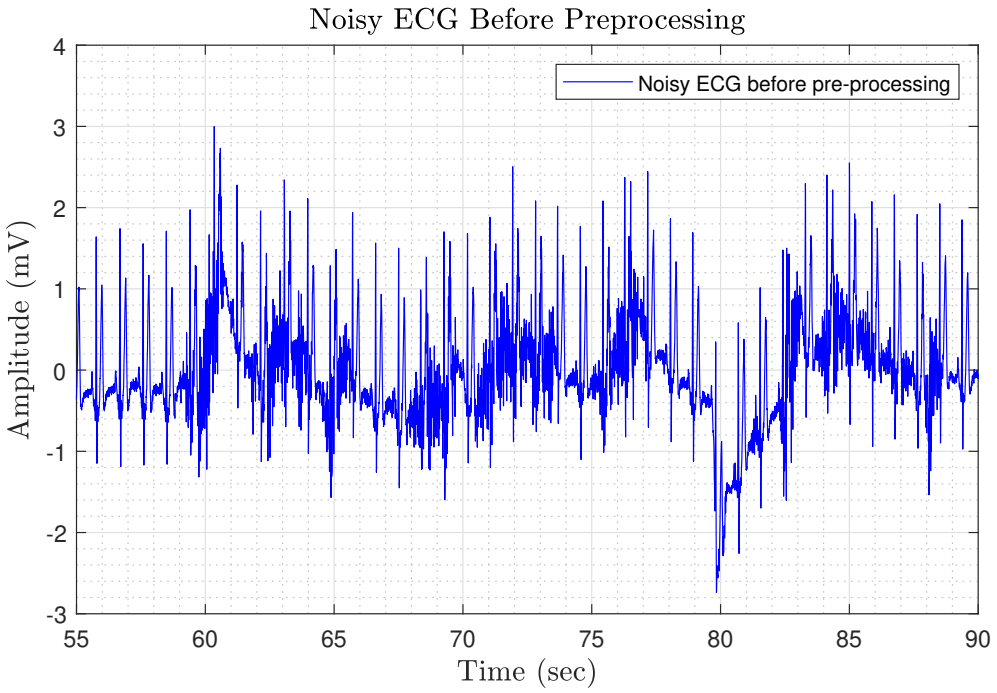


Fig. 4.4 AECG with additive real muscle noise to achieve a SNR of -1 dB, before pre-processing.

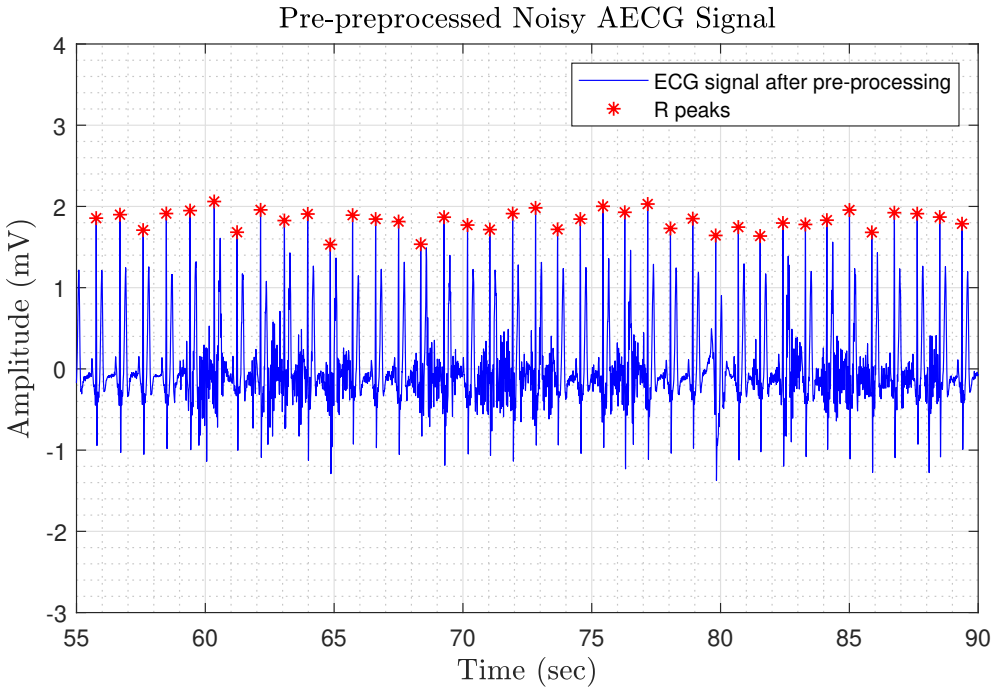


Fig. 4.5 AECG after pre-processing with R peak detection.

4.6.2 Training Data for PCA

In this section, the estimation of the Principal Component basis signals is described. It is important to extract PCA components from clean ECG signals. Pre-processing is used to remove most of the high and low frequency noise and other artefacts. Generally, in any long interval of training data, heart cycles exist that are strongly affected by noise. Including these cycles leads to distortion in the estimated PCA bases. An automated system has been developed to reject contaminated heart cycles. The method maintains an average heart cycle time series and each subsequent heart cycle is rejected from PCA analysis if it deviates too far from the average cycle. Accepted heart cycles are included in the running average. Figure 4.6 illustrates the ECG mean heart cycle calculated from training period using real AECG, while the results of ECG automated selection is illustrated in Figure 4.7. The blue cycles have been accepted as valid for PCA calculation while the red signals have been rejected. In the provided typical example, 10 minutes of measured ECG signal is used.

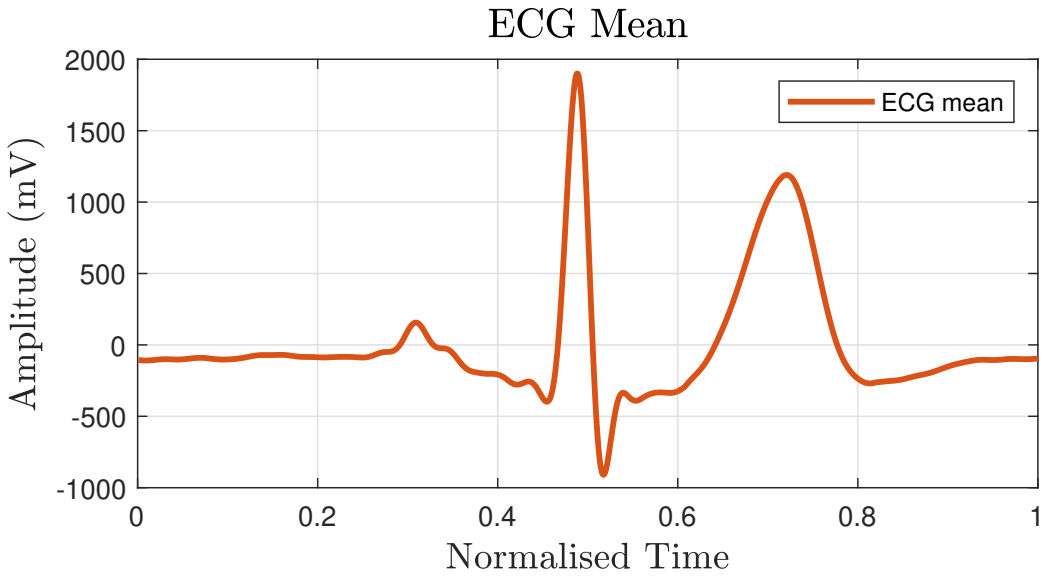


Fig. 4.6 Mean ECG using real data.

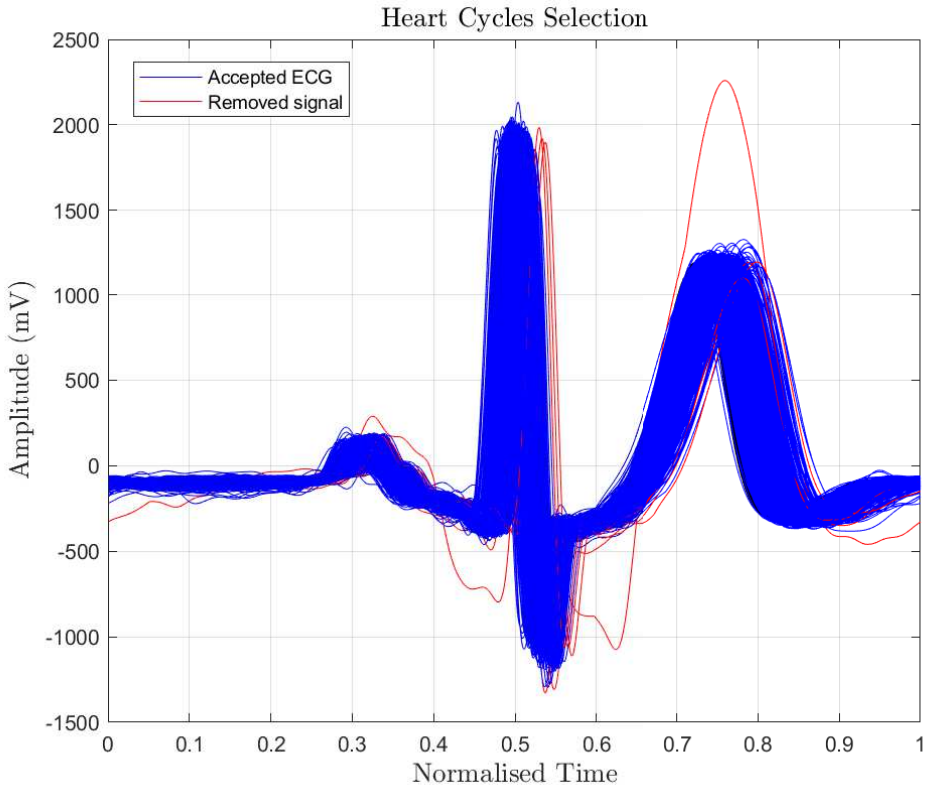


Fig. 4.7 Heart cycle selection for PCA using Real ECG data.

4.6.3 PCA Component Extraction

In this section, the PCA components are calculated from a ten minute training period of synthetic or measured ECG signals. The consequences of choosing different numbers of PCA basis signals are determined. Figure 4.8 illustrates the first 10 PCA basis signals from synthetic ECG data. Additionally, the first two basis signals are also presented. The choice of the number of PCA basis signals is tested by applying PCAKF to 100 samples of two minutes of synthetic ECG signal contaminated with APGN to yield 21 input SNRs. The average SNR improvements are provided on the Figure 4.9. It shows that using one or two PCA basis signals yielded the best performance and then performance monotonically decreased. The use of 2 PCA components yields the best results in terms of SNR, GoF and MSEWPRD measurements. The results become significantly poorer when using more than 5 PCA basis

signals. There is a trade-off in this choice. If too few basis signal are used then less of the variation in the heart signal can be captured. However, the PCA basis signals have decreasing SNR with basis number and so including more basis signals than necessary can introduce more noise into the filtered signal. Table 4.1 illustrates the typical results while the signal filtered using PCAKF while contaminated with added white Gaussian noise at 0 dB.

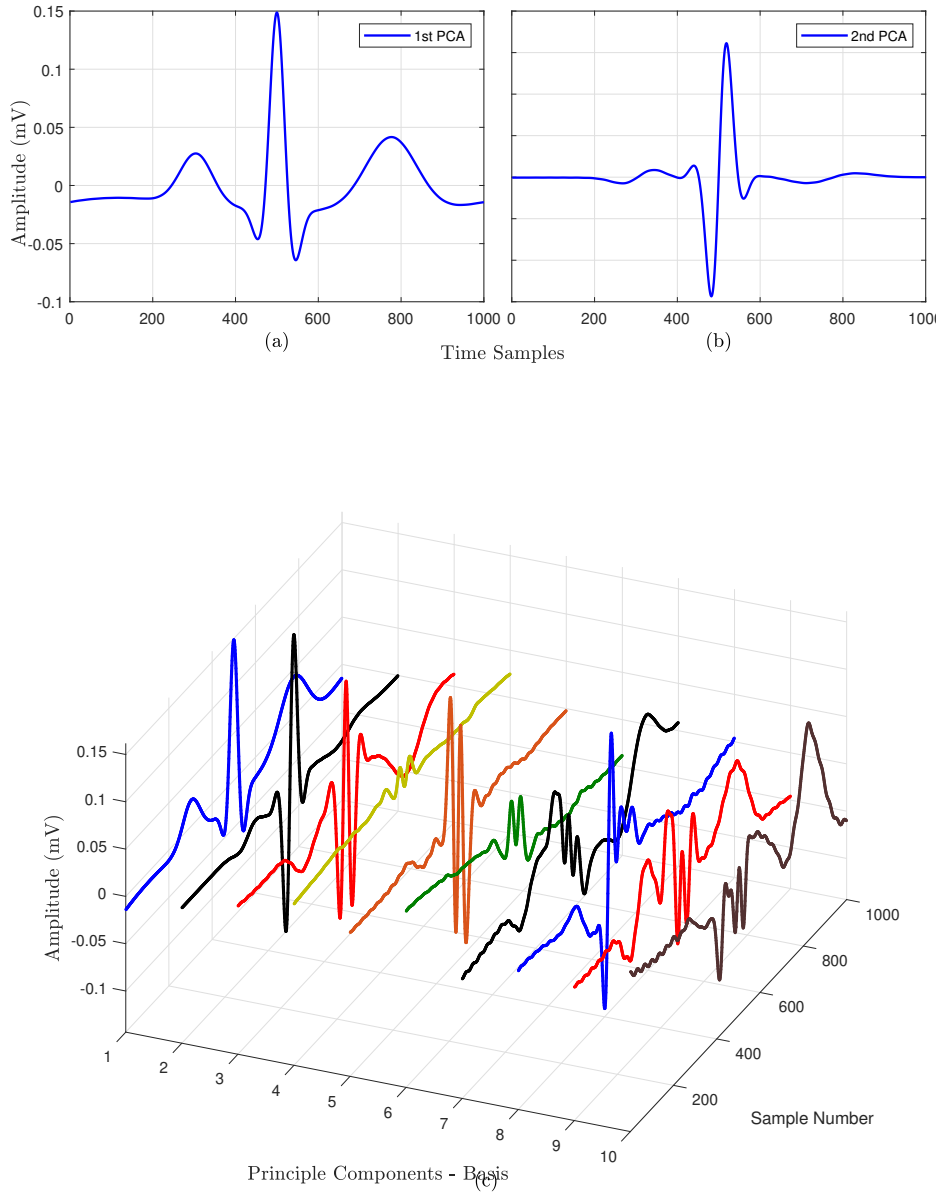


Fig. 4.8 The PCA basis signals for synthetic ECG data (a) 1st (b) 2nd (c) first 10 signals.

Number of PCA Basis Signals	SNR(dB)	GoF	MSEWPRD
1	11.9	0.94	0.0342
2	13.6	0.96	0.0247
3	12.8	0.95	0.0274
4	12.2	0.94	0.0283
5	11.5	0.93	0.0394
6	10.2	0.90	0.0392
7	8.70	0.87	0.0432
8	7.99	0.84	0.0502
9	7.43	0.82	0.0570
10	6.50	0.78	0.0617

Table 4.1 Typical filtering results for synthetic ECG data with AWGN at initial SNR of 0 dB.

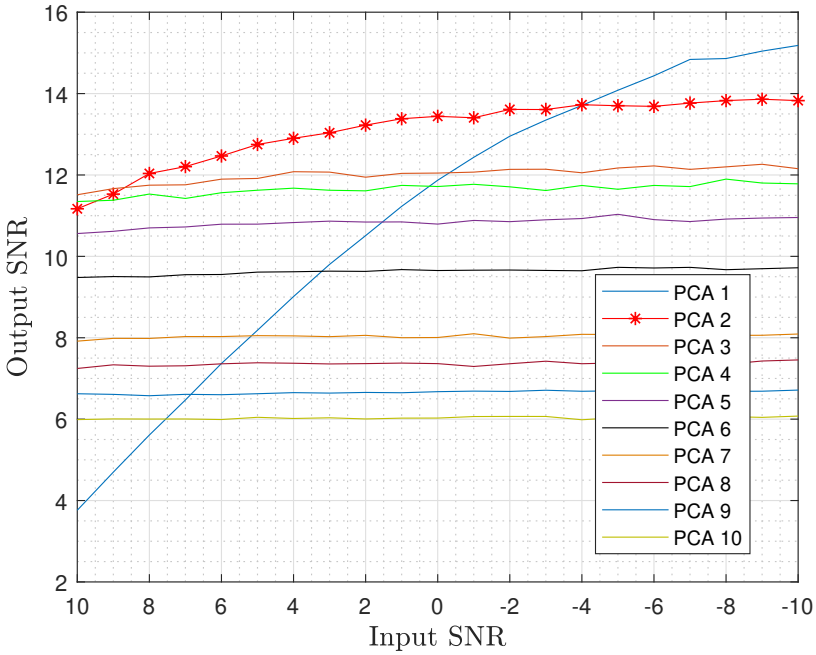


Fig. 4.9 The PCA basis signals for synthetic ECG data (a) 1st (b) 2nd (c) first 10 signals.

Testing with synthetic ECG signals shows that most of variation in synthetic ECG signal is spanned by the first two PCA components. A similar test has been performed using measured ECG data in section 4.7 and the results suggest that first 5 PCA span most of the variation in a measured normal ECG signal. Therefore, we use 5 PCA components in both synthetic and real signals for this thesis. In the next section, PCAKF is applied to synthetic ECG signals.

4.6.4 PCAKF Applied to Synthetic ECG

This section quantifies the performance of the PCAKF method when applied to synthetic ECG signals with four different types of noises including AWGN, APGN, Additive Muscle Noise (AEMN) and Additive Electrode Movement Noise (AEMN). Adding noise to clean ECG signals allows the de-noising performance of the methods to be tested and quantified against a known target signal. Figure 4.10 shows the test reported in this Chapter using synthetic ECGs.

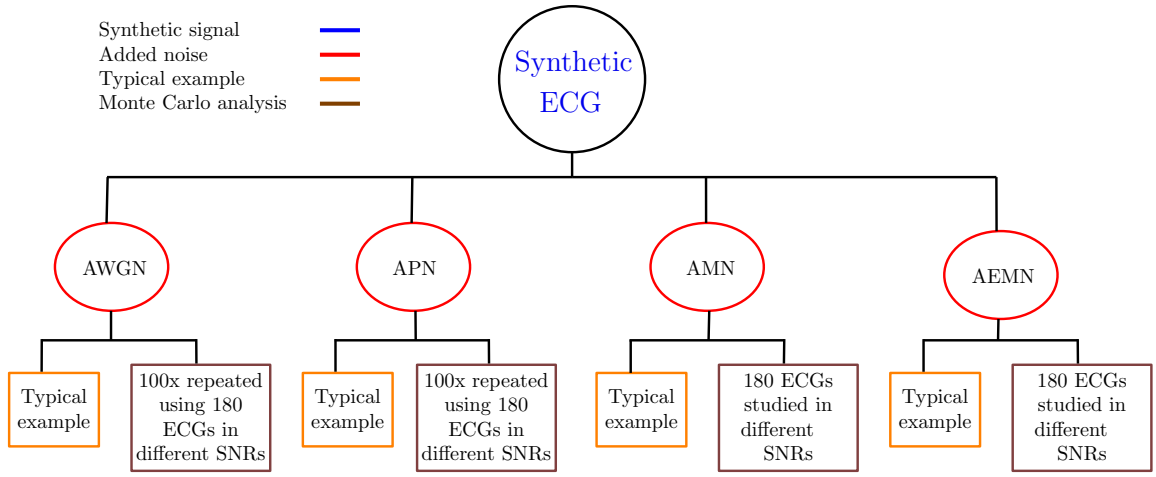


Fig. 4.10 The tests performed in this chapter using synthetic ECG data.

The following sections present the results of filtering synthetic ECG signals with additive noise. Five different filtering methods have been tested: EKF2, EKS2 followed by EKS2, EKF17, EKF17 followed by EKS17, and PCAKF. Four different noises are considered: AWGN, APGN, AMN and AEMN. For each noise type, a typical result of each filter method is illustrated. Monte Carlo results are calculated using 180 synthetic ECG segments each of 2-minutes duration.

Synthetic ECG with Additive White Gaussian Noise

In this section, synthetic ECG data contaminated with AWGN. The Monte Carlo statistical tests are used for applying 21 different input SNRs with 100 random noise signals for each.

Figure 4.11 represents the statistical analysis of the proposed method against benchmark methods where the proposed method filter better than benchmark methods. Also, typical results of the same procedure are illustrated in figure 4.12 which shows efficiency of the proposed method compare to benchmark methods. Table 4.2 represents the comparison between PCAKF algorithm and benchmark methods over two minutes of analysis where PCAKF algorithm improved SNR, GoF and MSEWPRD comparing EKF2 , EKS2 and their smoother filters.

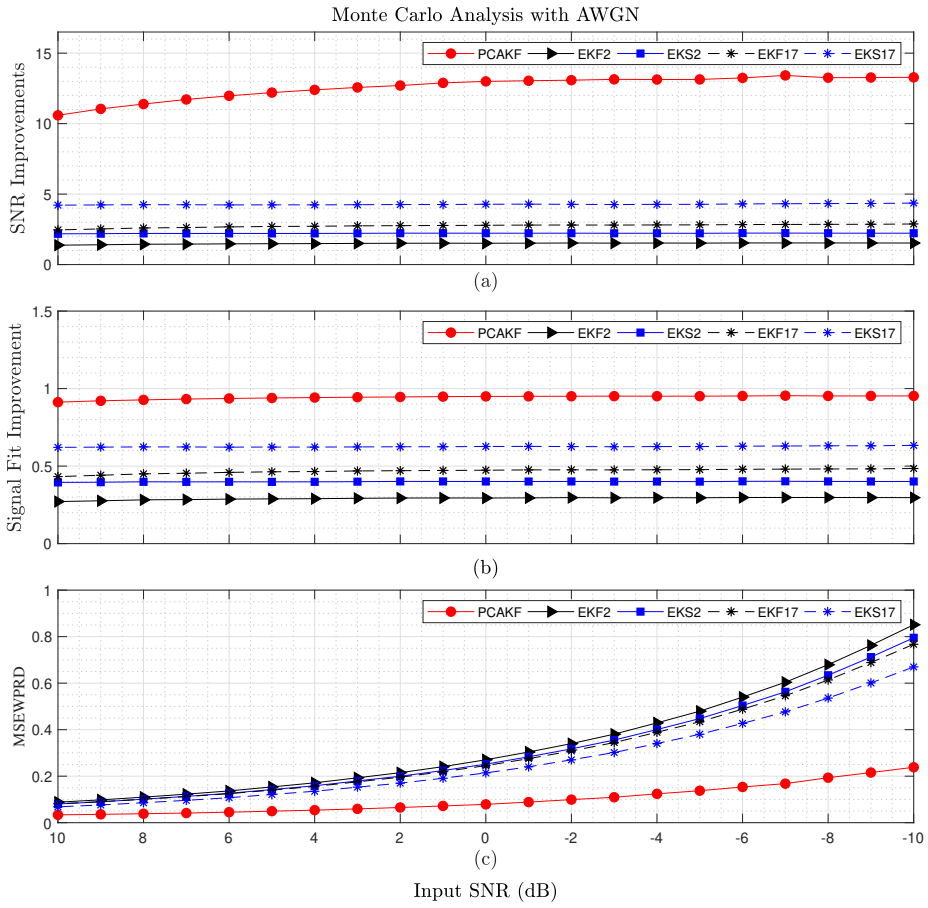


Fig. 4.11 Mean quality measures using 100 two minute ECGs with AWGN with a range of initial SNRs (a) SNR improvement, (b) GoF, (c) MSEWPRD.

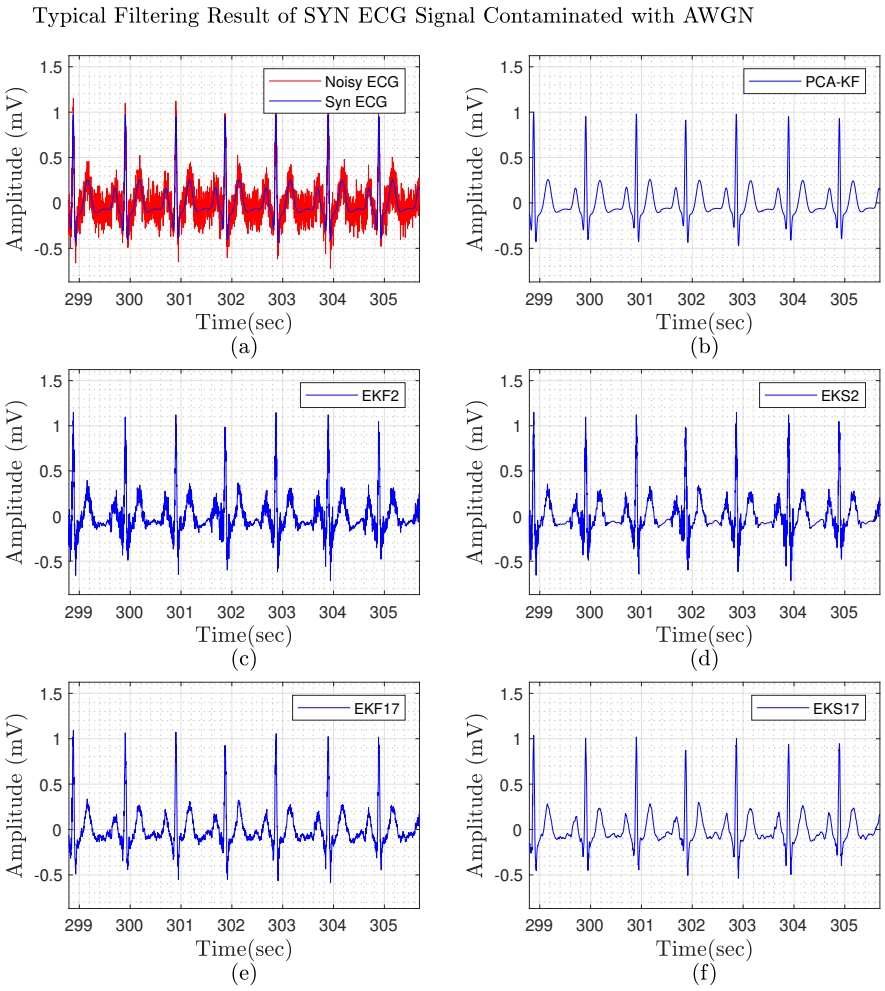


Fig. 4.12 Typical filtering results on two minutes of synthetic ECG data with AWGN at 5dB: (a) Noisy signal and synthetic ECG, (b) PCAKF, (c) EKF2, (d) EKS2, (e) EKF17, (f) EKS17.

Methods	SNR(dB)	GoF	MSEWPRD
EKF 2	1.52	0.30	0.15
EKS 2	2.24	0.40	0.13
EKF 17	2.76	0.47	0.13
EKS 17	4.29	0.63	0.11
PCAKF	11.6	0.93	0.05

Table 4.2 Typical filtering results for two minutes of synthetic ECG data with AWGN at an initial SNR of 5 dB.

Synthetic ECG with Additive Pink Gaussian Noise

This section provides statistic and typical results for PCAKF and benchmark methods using synthetic ECG data contaminated with APN. Again, 21 different SNRs are tested with 100 random noise signals each. Figure 4.13 represents the statistical analysis of the proposed method against benchmark methods. Figure 4.14 and Table 4.3 illustrate typical results while synthetic data contaminated with APN at -1 dB. As it seen in figures 4.13, 4.14 and table 4.3 , the proposed method has better filtering results compare to benchmark methods.

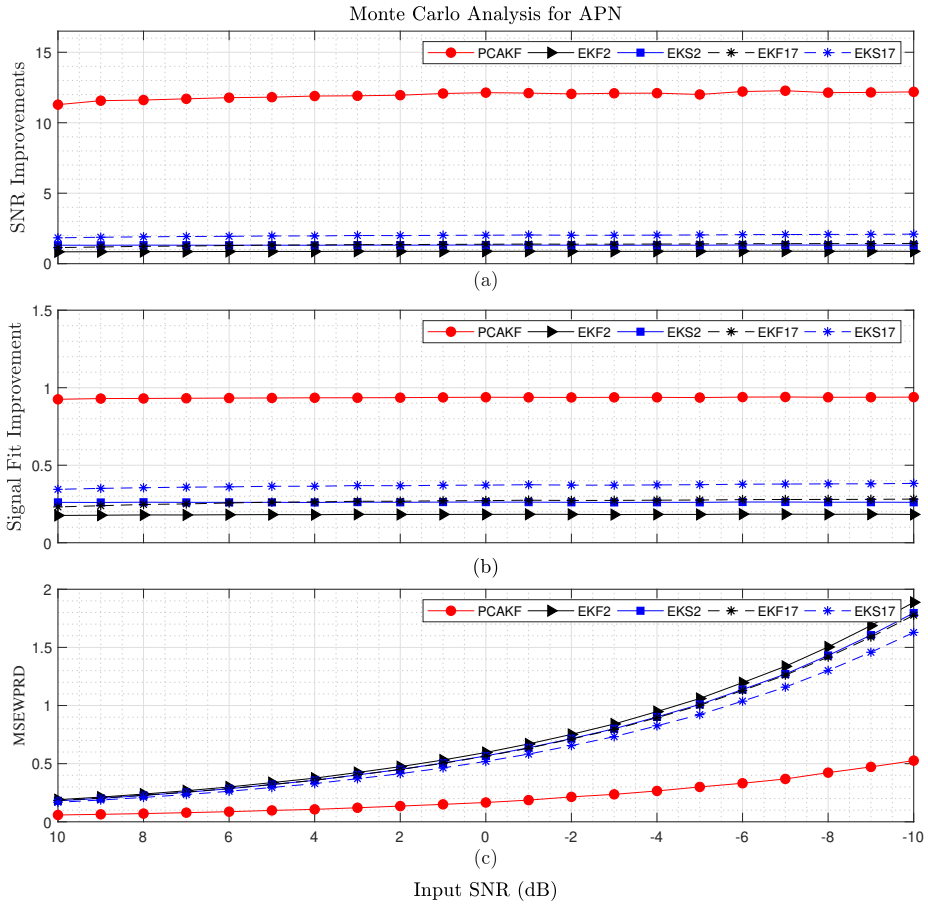


Fig. 4.13 Mean quality measures using 100 two minute ECGs with APN with a range of initial SNRs (a) SNR improvement, (b) GoF, (c) MSEWPRD.

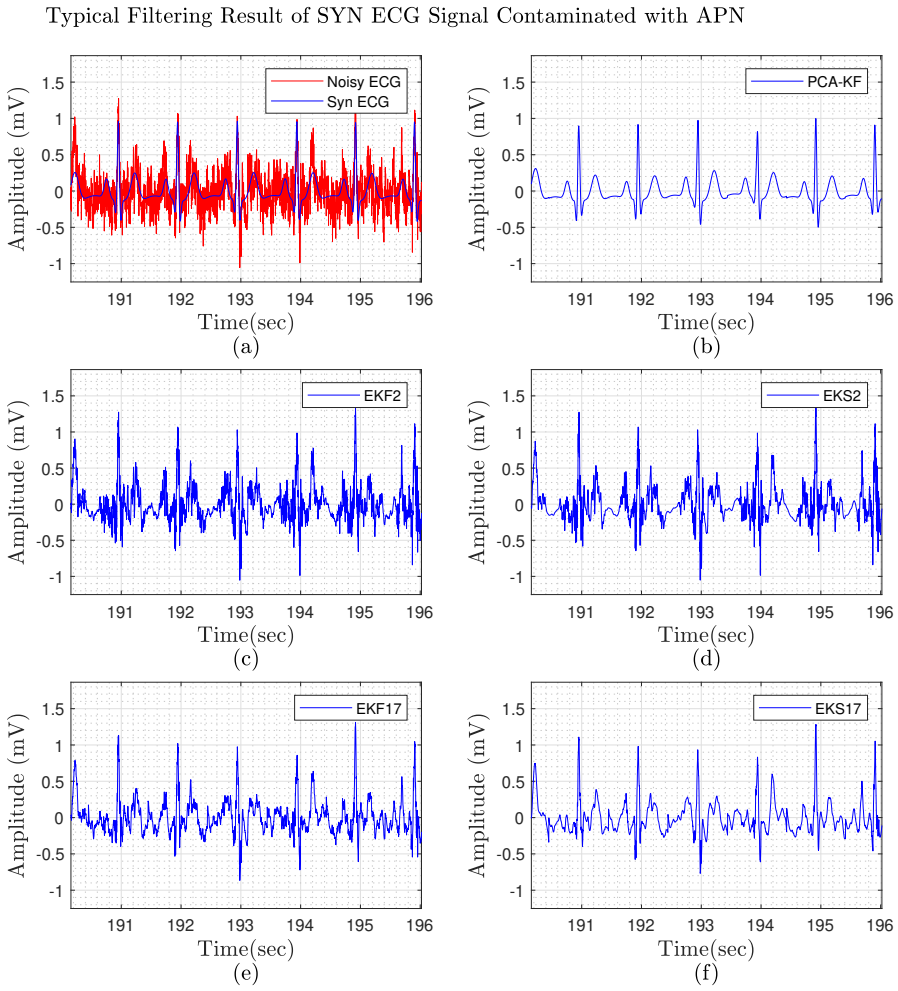


Fig. 4.14 Typical filtering results on two minutes of synthetic ECG data with APGN at -3 dB: (a) Noisy signal and synthetic ECG, (b) PCAKF, (c) EKF2, (d) EKS2, (e) EKF17, (f) EKS17.

Methods	SNR	GoF	MSEWPRD
EKF 2	0.815	0.17	0.82
EKS 2	1.17	0.24	0.79
EKF 17	1.32	0.26	0.77
EKS 17	1.92	0.36	0.71
PCAKF	12.0	0.94	0.22

Table 4.3 Typical filtering results for two minutes of synthetic ECG data with APN at an initial SNR of -3 dB.

Synthetic ECG with Additive Real Muscle Noise

In this section synthetic ECG data is contaminated with real muscle noise to study the performance of the PCAKF framework in a more realistic simulation. Noise was resampled from 360 to 512 Hz and 100 intervals, with duration of two minutes, are selected and added to 100 ECG signals. Noisy data were filtered using PCAKF algorithm and benchmark methods on different SNRs. Results are provided in Figure 4.15. Typical results of test is given at figure 4.16 and Table 4.4. It experimentally proves that the PCAKF algorithm able to remove muscle noise from ECG signal better than benchmark methods in both typical and statistic analysis for different input SNRs.

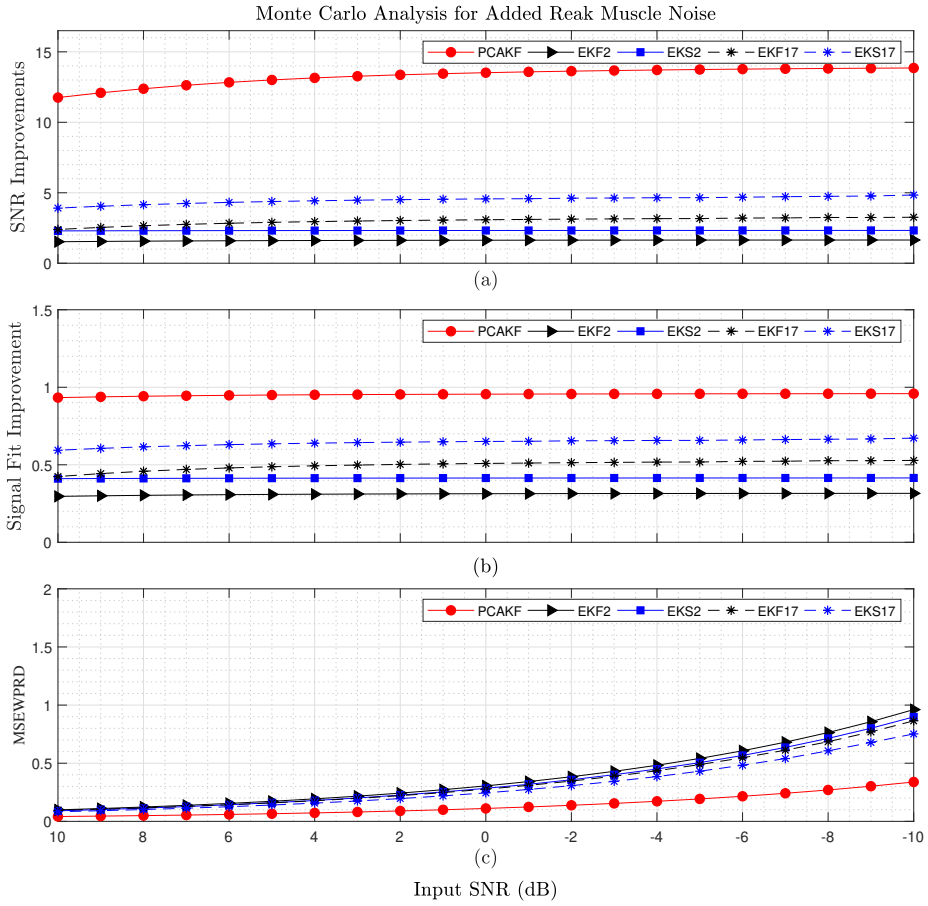


Fig. 4.15 Mean quality measures using 100 two minute ECGs with AMN with a range of initial SNRs (a) SNR improvement, (b) GoF, (c) MSEWPRD.

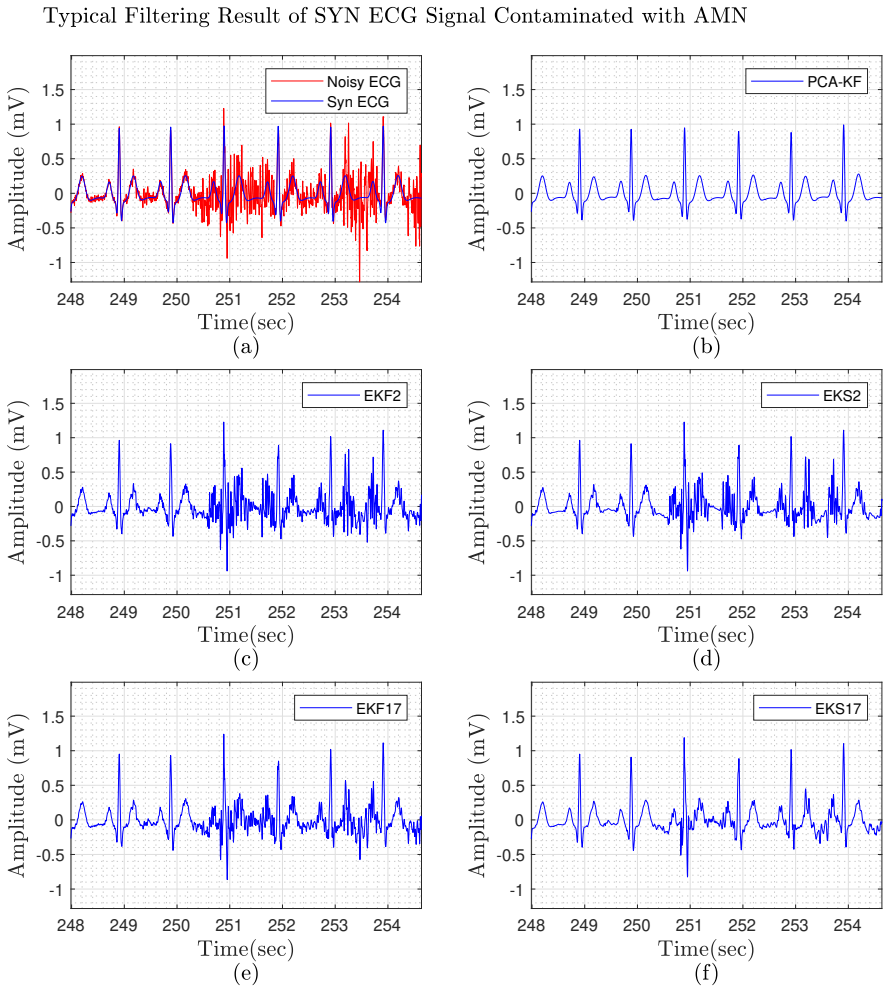


Fig. 4.16 Typical filtering results on two minutes of synthetic ECG data with AMN at 0 dB: (a) Noisy signal and synthetic ECG, (b) PCAKF, (c) EKF2, (d) EKS2, (e) EKF17, (f) EKS17.

Methods	SNR	GoF	MSEWPRD
EKF 2	1.58	0.31	0.31
EKS 2	2.34	0.42	0.29
EKF 17	3.06	0.51	0.28
EKS 17	4.60	0.65	0.25
PCAKF	14.5	0.96	0.083

Table 4.4 Typical filtering results for two minutes of synthetic ECG data with AMN at an initial SNR of 0 dB.

Synthetic ECG with Additive Electrode Movement Noise

In this section, the performance of the PCAKF is studied using synthetic ECG data and measured AEMN. The same procedure as in Section 4.6.4 is used, but with measured AEMN. The PCAKF algorithm examined Figure 4.17 shows results for statistical analysis and typical results are provided in Table 4.5 with the signal contaminated with -2 dB AEMN.

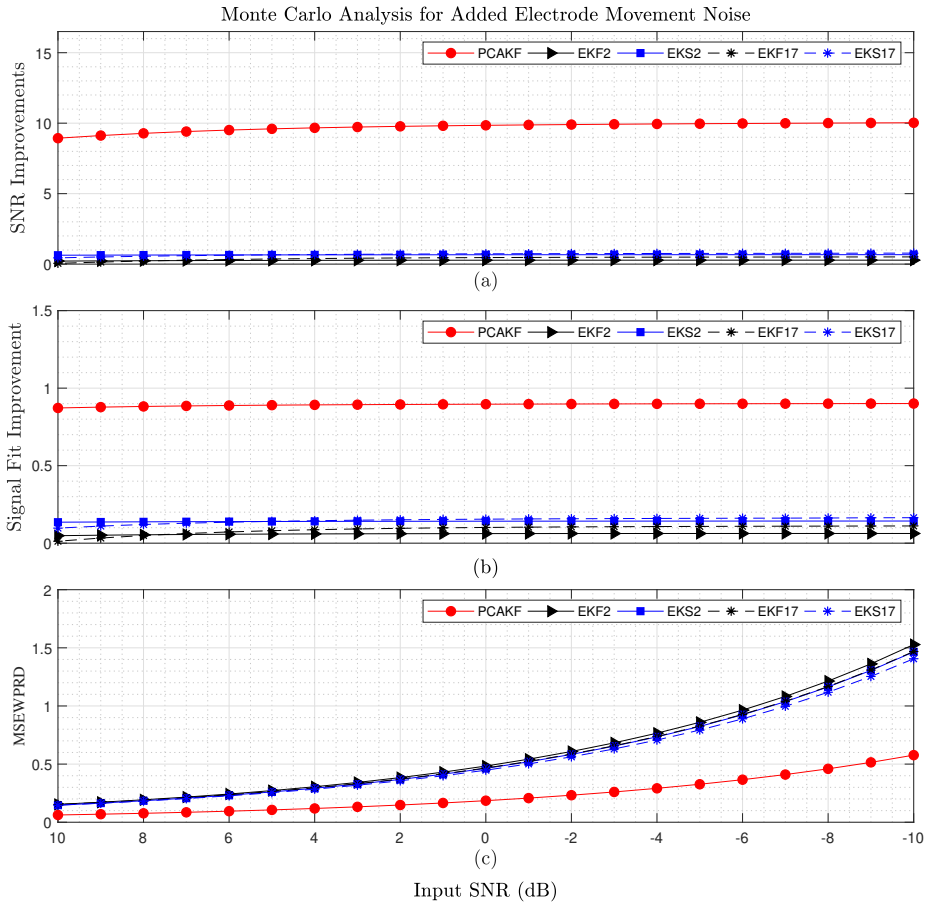


Fig. 4.17 Mean quality measures using 100 two minute ECGs with AEMN with a range of initial SNRs (a) SNR improvement, (b) GoF, (c) MSEWPRD.

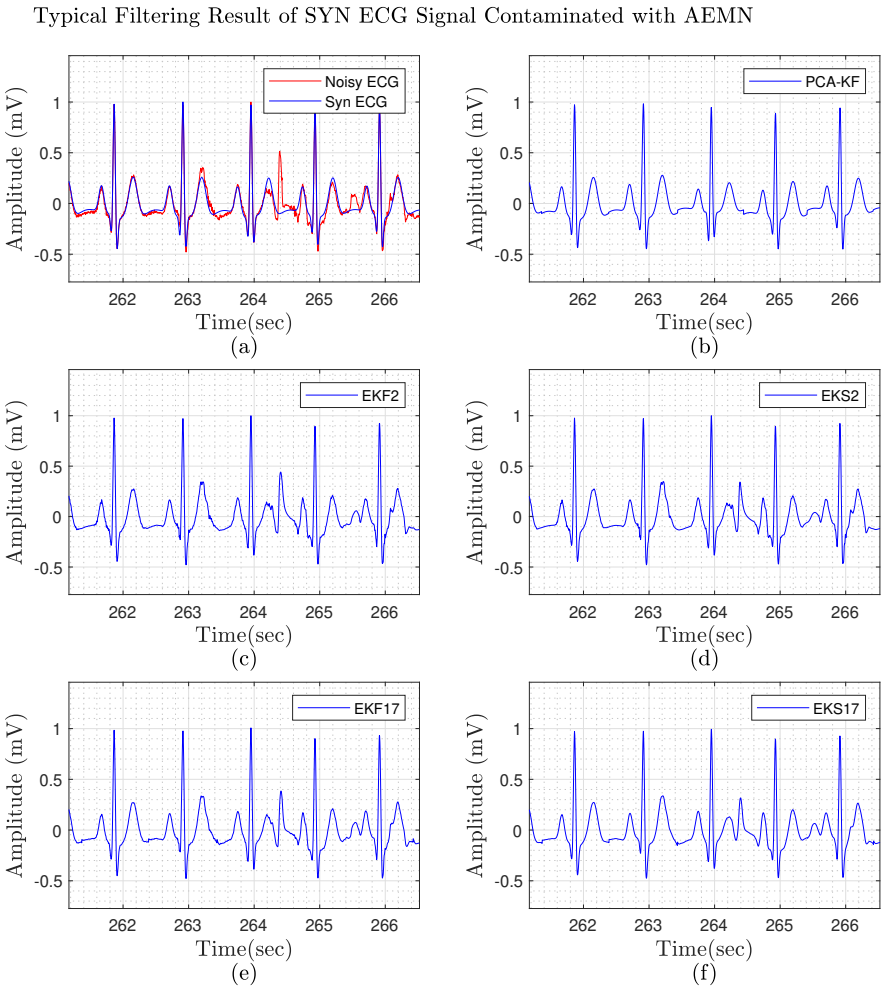


Fig. 4.18 Typical filtering results on two minutes of synthetic ECG with AEMN at -2 dB: (a) Noisy signal and synthetic ECG, (b) PCAKF, (c) EKF2, (d) EKS2, (e) EKF17, (f) EKS17.

Methods	SNR	GoF	MSEWPRD
EKF 2	0.28	0.063	0.62
EKS 2	0.57	0.12	0.60
EKF 17	0.47	0.10	0.60
EKS 17	0.72	0.15	0.58
PCAKF	9.83	0.90	0.21

Table 4.5 Typical filtering results for two minutes of synthetic ECG data with AEMN at an initial SNR of -2 dB.

4.7 Measured Data

In this section, the filtering methods are applied to 30 minutes of measured AECG data. Measured muscle noise has been scaled and added to the measured AECG signals to yield a range of initial SNRs (calculated assuming no noise in the measured AECG). Although SNR improvement and GoF are unreliable, due to the known noise in the measured signals, MSEWPRD can be calculated. Figure 4.19 shows the 5 PCA basis signals calculated from 10 minutes of measured AECG signal. Figure 4.20 shows the proportion of ECG signal variance explained by each PCA component. It is a feature of PCA that each successive basis signal explains a smaller proportion of the variance. Figure 4.21 illustrates the output MSEWPRD comparison. Based on this test, 5 PCA basis signals were selected as a compromise between performance at low and high initial SNRs.

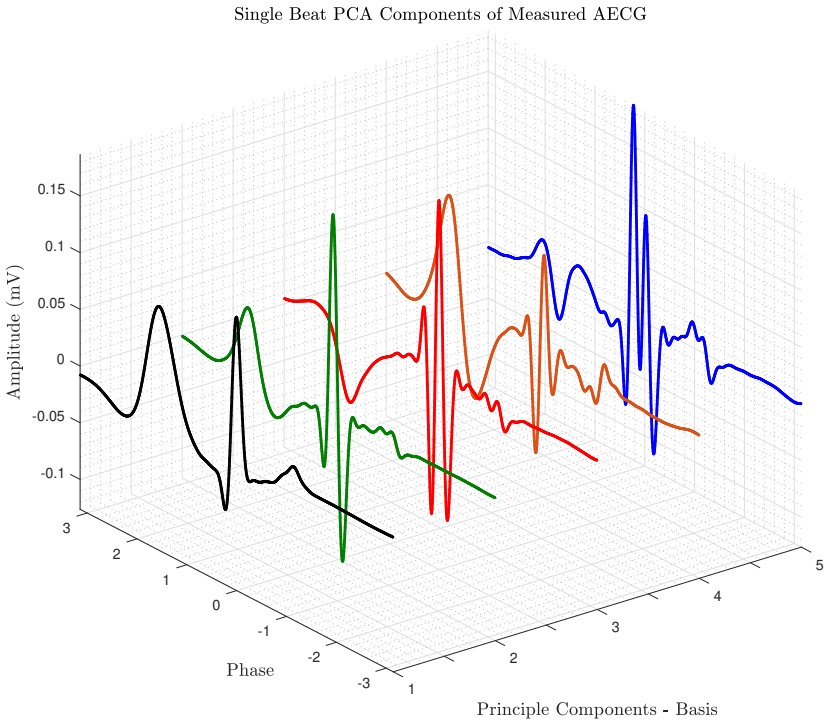


Fig. 4.19 PCA components using real AECG signal.

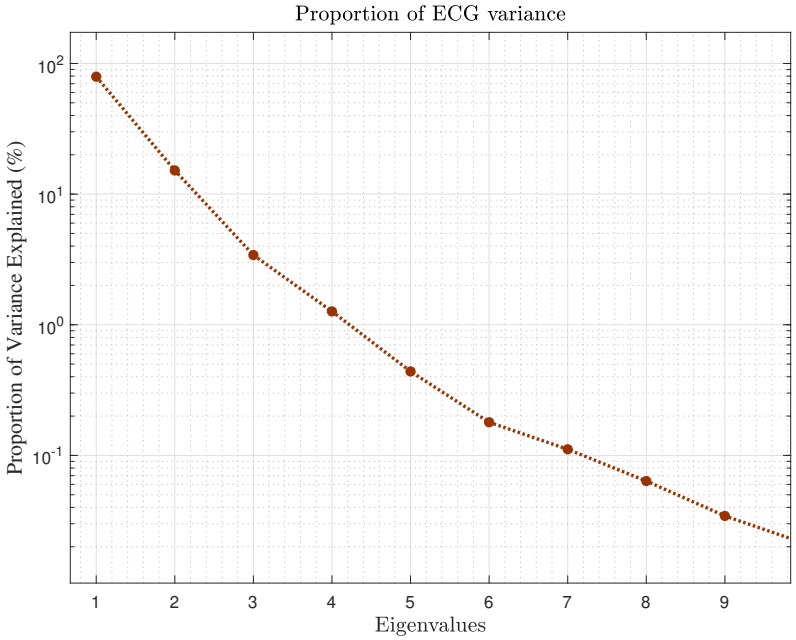


Fig. 4.20 Proportion of ECG variance explained by PCA basis signals.

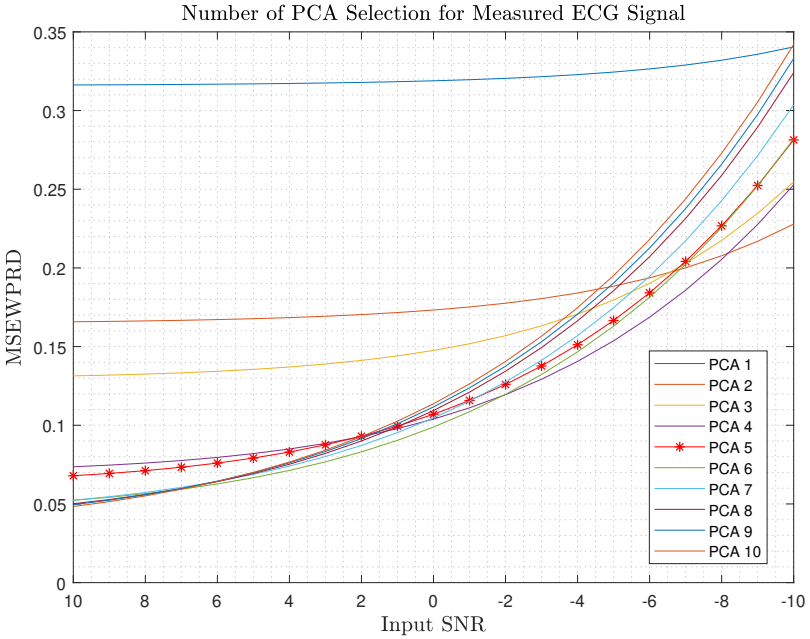


Fig. 4.21 Effect of the number of PCA basis signals on PCAKF denoising quality.

4.7.1 Typical Example of Measured Data

Figure 4.22 illustrates the filtering portion of 4-seconds AECG signal contaminated with AMN at -5 dB. Results of denoising for two minutes are presented on 4.6.

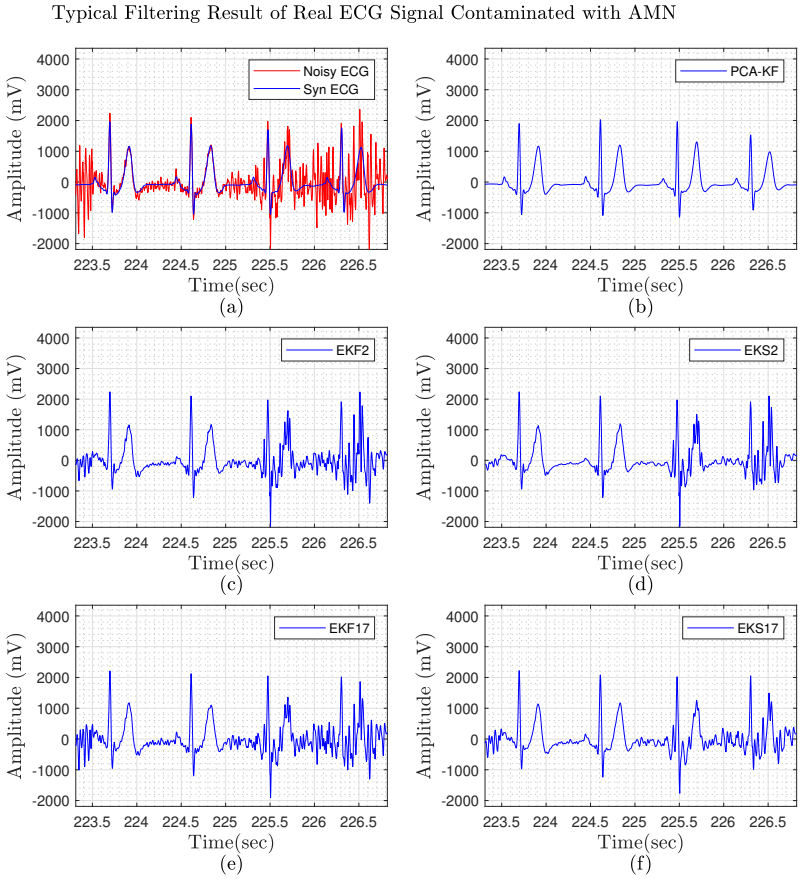


Fig. 4.22 Typical filtering period from two minutes of measured AECG with AMN at -5 dB: (a) Noisy signal and synthetic ECG, (b) PCAKF, (c) EKF2, (d) EKS2, (e) EKF17, (f) EKS17.

Methods	EKF 2	EKS 2	EKF 17	EKS 17	PCAKF
MSEWPRD	0.471	0.424	0.492	0.416	0.180

Table 4.6 Typical MSEWPRD results for 2-minute real AECG signal with AMN at -5 dB.

4.8 Discussion

In this chapter, numerical experiments have been performed on the ECG denoising properties of EKF\EKS and the PCAKF filter, for both synthetic and measured AECG data and noise. Standard pre-processing algorithm has been applied to remove baseline drift, high frequency noise and common artefacts. Figure 4.5 showed an example of AECG signal with added -1 dB real muscle noise. The baseline drift in the noise has changed the shape of the signal. As this would normally be removed by pre-processing of the measured signal plus noise, it has been important to apply pre-processing to all additive noise signals before initial SNR calculations. Standard R peak detection is also shown in Figure 4.5, and the clean signal before addition of noise is used for this. Although, in practise, R peak detection would use pre-processed data including noise, it was important that the filtering methods were compared without the complication of R peak detection faults.

In order to implement PCAKF, a portion of the signal is used as training data to calculate the PCA basis signals. EKF methods also use a training interval to select the EDM parameters. A method has been developed to automate the selection of low-noise heart cycles to be used when calculating the signal covariance.

To calculate the best number of PCA basis signals to use, various tests have been performed on synthetic and real ECG data. Using two PCAs had the best results on synthetic signals however for the measured data, the first five PCA components contain over 99% of the variance. This choice is a compromise and so a choice has been made that will be applicable to a wide range of ECG signals and measurement scenarios. Longer periods of training data yield PCA basis signals with lower noise contamination, and so more basis signals could potentially be used. The length of training data is also a compromise and so a realistic duration of ten minutes was used in these tests. Around 3960 analyses have been performed using synthetic ECG signal with AWGN, APGN, AMN and AEMN. In all cases, PCAKF performed better than the benchmark methods. Table 4.7 shows the mean quality

measures overall the tests. It is clear that PCAKF yields significantly better signal and noise separation while maintaining the clinically important properties of the signal.

Methods	SNR	GoF(%)	MSEWPRD
EKF 2	0.94	18	0.53
EKS 2	1.46	27	0.50
EKF 17	1.43	31	0.51
EKS 17	2.90	45	0.46
PCAKF	12.2	93	0.16

Table 4.7 Mean quality measures over all numerical experiments.

Measured ECG signals are contaminated with some level of noise, even when recorded in perfect condition. This makes it more difficult to quantify filter performance with real data. However realistic synthetic ECG test signals do not exhibit the complexity of real data. In the final section, measured ambulatory ECG signals are contaminated with AMN. The results are broadly similar to those produced with synthetic ECG signals: PCAKF results are smoother and more plausible as cardiac signals than those of the benchmark methods. In some cases, the PCAKF output is more plausible than the initial AECG data, suggesting that initial noise, albeit small, has been reduced.

The PCAKF algorithm is an order of magnitude computationally faster than EKF2, EKF17, and requiring about 1 ms to process a heart cycle on a standard PC, compared to 7 ms for EKF2-EKF17, for 128 Hz data. For more modern 1 kHz sampled data, the difference is larger at 1.4 ms compared to 115 ms. The PCAKF algorithm may easily run on a smartphone in real time and may be part of a hierarchical monitoring and decision making system. Matlab code for single beat single channel PCAKF algorithm are provided on D.

4.9 Conclusions

Despite extensive literature in ECG denoising, there are few methods effective on non-stationary Ambulatory ECG signals, experiencing intermittent high amplitude noise from

a range of sources. This chapter demonstrated the filtering of synthetic and measured ambulatory AECG time-series after the addition of synthetic or measured noise. A new method known as PCAKF has been presented, based on Kalman tracking of single heart cycles expressed as a weighted sum of PCA basis signals. In these tests, the PCAKF method performed well on normal heart signals. The next chapter will test the method on ECG with arrhythmias.

Chapter 5

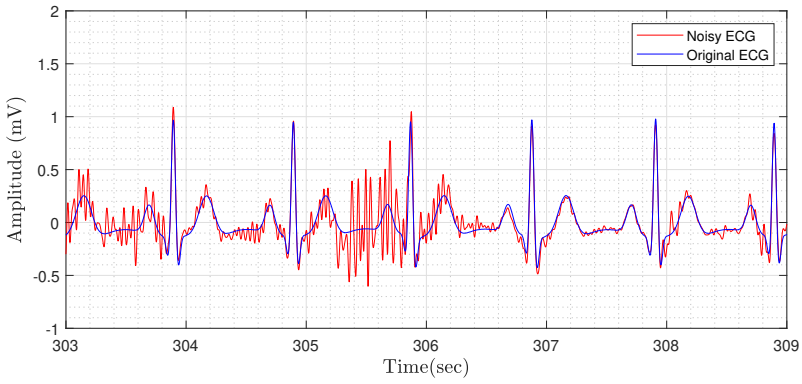
Double Beat - Principal Component Analysis Kalman Filter

5.1 Intorduction

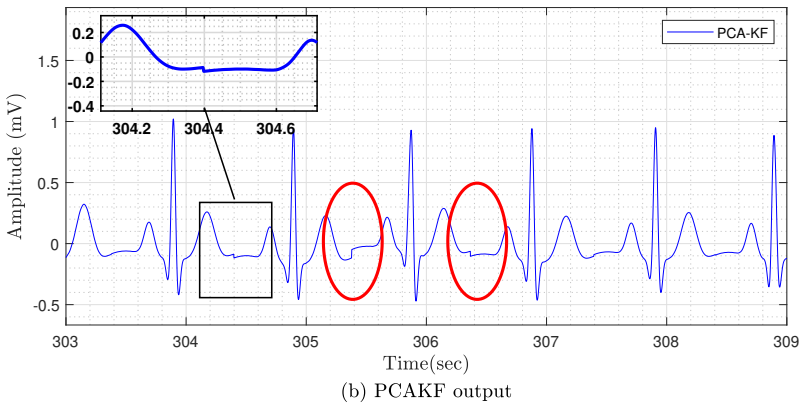
This chapter develops the method presented in the previous chapter with the innovation that two consecutive heart cycles as a time are modelled. In the previous chapter, the PCAKF denoising algorithm was introduced, which processes a single heart cycle at a time from a single ECG channel. The single-beat single-channel PCAKF method performed better than gold standard EKF2\ EKF17-EKS2\EKS17 methods, using both synthetic and real ECG signals and noise. A criticism of the single beat PCAKF method is that it yields time series with discontinuities at heart cycle boundaries. The double beat variant was developed to address this but has shown other advantages as well. The double beat method is presented in this chapter and it is tested using normal, abnormal, AF signals and then extended to multi-channel ECG recordings. The current chapter has three phases. In the first step, DB-PCAKF framework is introduced. Then the PCA components are calculated for different types of ECGs. Finally,the extracted PCA components are used on the suggested algorithm to denoise noisy signals.

5.2 Discontinuity on Single Beat PCAKF

It was observed in several tests, that discontinuities can occur between heart beats in the output of the SB-PCAKF method. For low noise signals, the discontinuity is very small and often not visible. However, even for low noise signals, if the training data is limited, anomalous beats or noisy epochs may be included; the PCAKF output can have visible discontinuities. These discontinuities effect the perceived morphology of signal and could hide a real cardiac anomaly or be interpreted as a cardiac signal. Figure 5.1 illustrates a discontinuity in the output of an ECG signal contaminated with AMN to yield a SNR of -4 *dB*.



(a) Noisy and Original ECG



(b) PCAKF output

Fig. 5.1 (a) original clean measure data and after adding AMN; (b) SB-PCAKF output showing discontinuities.

When discontinuities are present, they are common. This suggests that they are introduced through the PCA bases. If the training data is noisy, and in particular if the baseline wander has not been fully removed, then the PCA basis signals may not converge to 0 at phases: $\phi = \pm\pi$. These phases are at the midpoints between R peaks and have been chosen as the cardiac signal and should be near to the baseline. If the PCA basis signals include some residual baseline wander then the filter output, which is a weighted sum of these basis signals, will not start and end at the baseline. In this situation the discontinuities will become apparent. This could be addressed by tapering the PCA basis signals to zero at $\phi = \pm\pi$. However, in this project a different approach has been taken. The beat transition phase points are arbitrary and if the heart cycles are irregular then the PCA basis signals may be a poor representation. Numerical experimentation has shown that the single beat algorithm does not perform well in this common clinical scenario. The new approach that has been developed, uses over lapping analysis intervals spanning two consecutive heart cycles from R peak to R peak. The R peaks are a very distinct feature of the cardiac cycle and using these as boundaries removes the arbitrary use of the mid-points between R peaks. The new variant of the algorithm is known as double beat (DB) PCAKF: DB-PCAKF.

5.3 Double Beat PCA Components

The DB-PCAKF method uses an approximation of the ECG signal over two heart cycles as a linear combination of double beat PCA basis signals derived from a training period. The method steps through the ECG signal to be filtered, moving forward one heart cycle at a time, but considering two consecutive heart cycles at each iteration. Each heart cycle is considered twice: first when it is the 2^{nd} cycle of two beats and then in the next iteration when it is the 1^{st} cycle. A weighted mean of these two approximations is taken as the filter output. Each filtered double beat interval is added to the output signal after scaling with a triangular weight function that ramps linearly from a value of 1 at the central R peak, to zero

at the beginning and end of the interval. Each point in the output signal is derived from two analysis intervals with weights that heavily favour one analysis period near an R peak to equal weighting of two intervals half-way between R peaks. Due to the smooth tapering of mixing weights, no discontinuities are possible in the output.

Another innovation is the non-linear variation of phase. The EKF\EKS and SB PCAKF methods, assume that the phase varies linearly over each heart cycle. DB-PCAKF does not make this assumption. A phase of $\phi_n = 2n\pi$ is assigned to the nth R peak at time t_n^R. A quadratic transformation is used to transform times to phases i.e. when considering the interval with the n^{th} R peak in the middle the transformation is the quadratic that passes through the points: $\{(t_{n-1}^R, -2\pi), (t_n^R, 0), (t_{n+1}^R, +2\pi)\}$.This transformation has the advantage that some anomalous cardiac signals are normalised by this process, for example AF where some heart cycles are very long. After quadratic distortion, each heart cycle has the same duration in phase, and so some of the other variations in wave positions and distortions are now spanned by the normal variation in the training period. Other transformations were tested in developing this method. A cubic spline could be fitted to all the $\{(t_n^R, \phi_n), n = 1, 2, ..., N_T\}$ data to yield a transformation that had continuous derivatives. However, in rare cases with extreme heart cycle variation, the transformation became non-monotonic.

The DB-PCAKF requires the same pre-processing to remove baseline wander and high frequency noise as other methods, see Section 4.6.1. The training ECG data is assessed by an automated heart cycle selection to eliminate data with strong non-cardiac signals and to prevent them contaminating the PCA basis calculation. Figure 5.2 shows the ECG mean using two heart cycles and figure 5.3 illustrates the automated heart cycle rejection using 10 minutes of measured AECG data. A set of PCA basis signals spanning two heart cycles are calculated from the clean data.

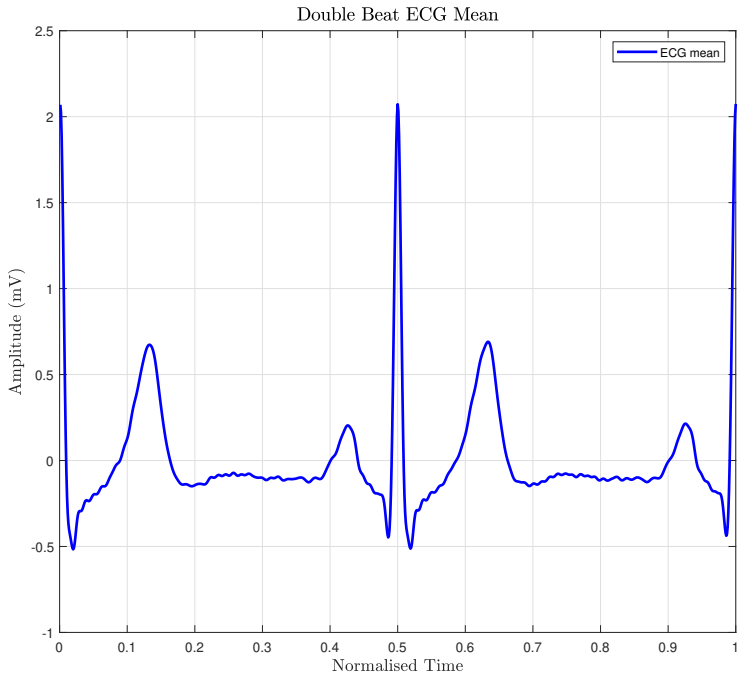


Fig. 5.2 Mean of two consecutive heart cycles calculated from good AECG data.

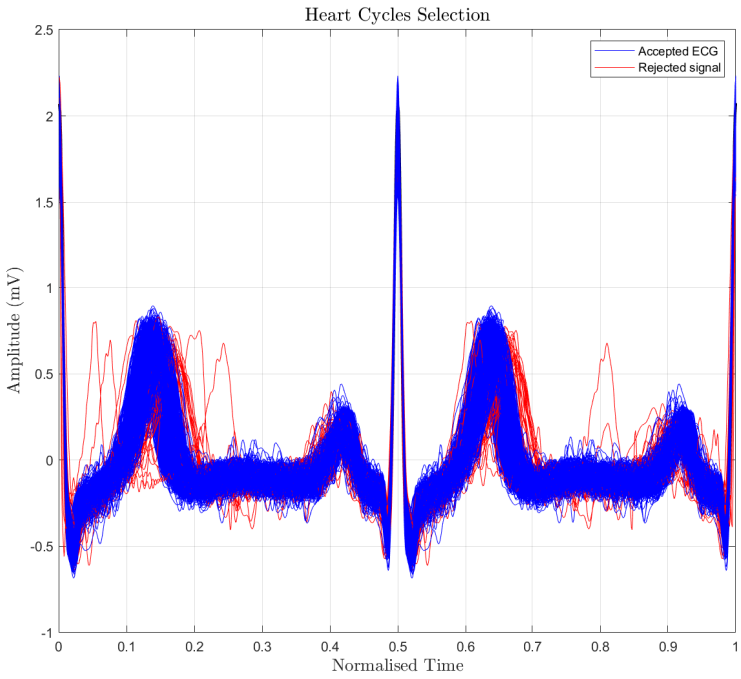


Fig. 5.3 ECG Selection for DB-PCA calculation.

5.4 PCA on Normal ECG

First, normal ECG signal is selected to study using PCAKF. Figure 5.4 illustrates the first five DB-PCA basis signals derived from measured AECG data. The first basis signal is the mean of the clean beats and clearly shows the PQRST waves of the second beat, along with the RST of the first beat and the PQR of the next beat. The middle section of the basis signals is the most important as the ends are multiplied by a small weight when mixing the DB output intervals. After temporal transformation, the largest variations are in the QRS amplitudes and the position and amplitude of the two T waves. Due to transformation normalisation, it is possible that fewer basis functions can be used, but in these tests the same number (5) as is SB-PCAKF has been used.Once the DB-PCA basis signals have been calculated, the KF process continues just as with SB. Each double beat interval is approximated by a weighted sum of basis functions and KF is used to track the weights from one double beat to the next. The only other difference is the tapered weights applied to each double beat as it is accumulated into the output signal.

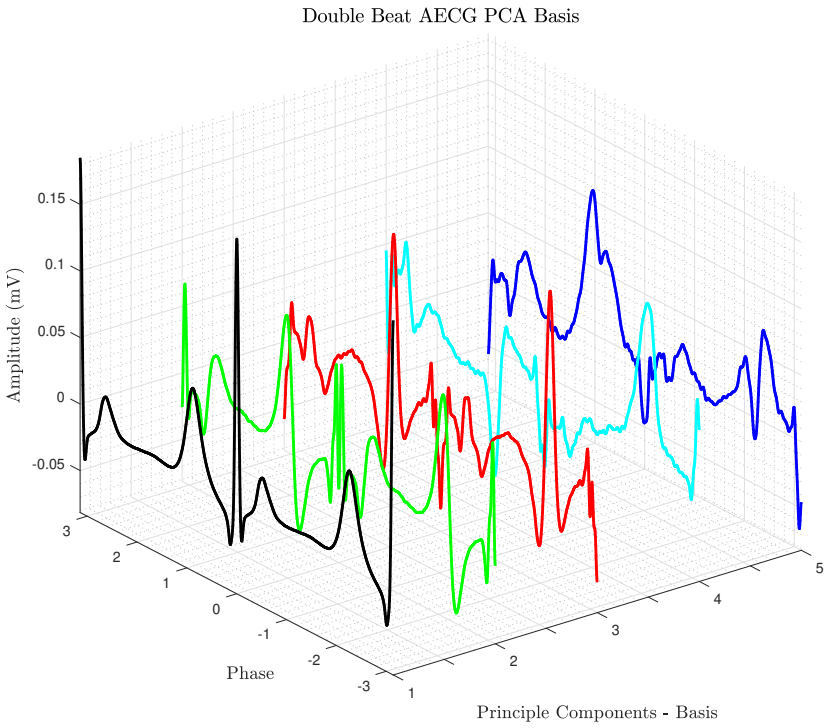


Fig. 5.4 DB PCA basis signals derived from measured Normal AECG data.

5.5 Filtering of Abnormal ECG Signals

In the previous section DB-PCAKF was applied to normal ECG signals. In this section DB-PCAKF is applied to ECG signals with abnormalities. Arrhythmias are due to abnormal contractions of the heart muscles, leading to irregular heartbeats. This irregularity can be a sequence of beats that may be too fast (tachycardia), or too slow (bradycardia), or with an irregular pattern from beat to beat such as flutter or fibrillation. In this thesis, three common abnormal heart signals were selected to study using the suggested method. These include:

- Early Ventricular Contractions;
- Atrial fibrillation;
- T-wave alternans.

5.5.1 Early Ventricular Contractions

For a healthy heart, the normal heart rhythm is controlled by electrical messages produced by sinus node (SA), located on right atrium and known as the heart's natural pacemaker. The normal electric messages are sent out through the heart regularly and each message triggers the heart to contract and pump blood around the body. When this system becomes disturbed, the result may be arrhythmia. Many healthy people experience abnormal heart beats during periods of stress. Over a protracted period this can even lead to heart failure.

A particular abnormality is known as premature ventricular contractions (PVCs). Occasionally PVCs have no physical signs other than their appearance in ECG data. With early and correct diagnosis, it is possible to prevent heart failure. In this chapter, the performance of DB-PCAKF is studied using real abnormal ECG signals. The Physionet abnormal ECG data base has a sample rate of 128 Hz and duration of 5 minutes (Goldberger et al., 2000, Moody, 2000a). For this test only, the training and test data are the same. Figure 3.6a illustrates an abnormal ECG while Figure 5.6 illustrates the PCA basis signals derived from the ECG signals with PVC.

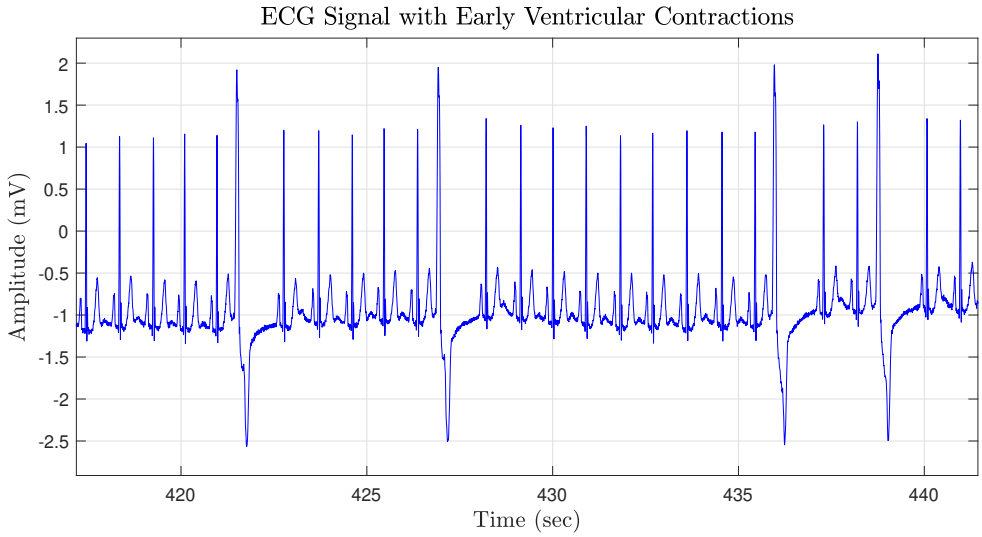


Fig. 5.5 Early Ventricular Contractions ECG (Goldberger et al., 2000, Moody, 2000a).

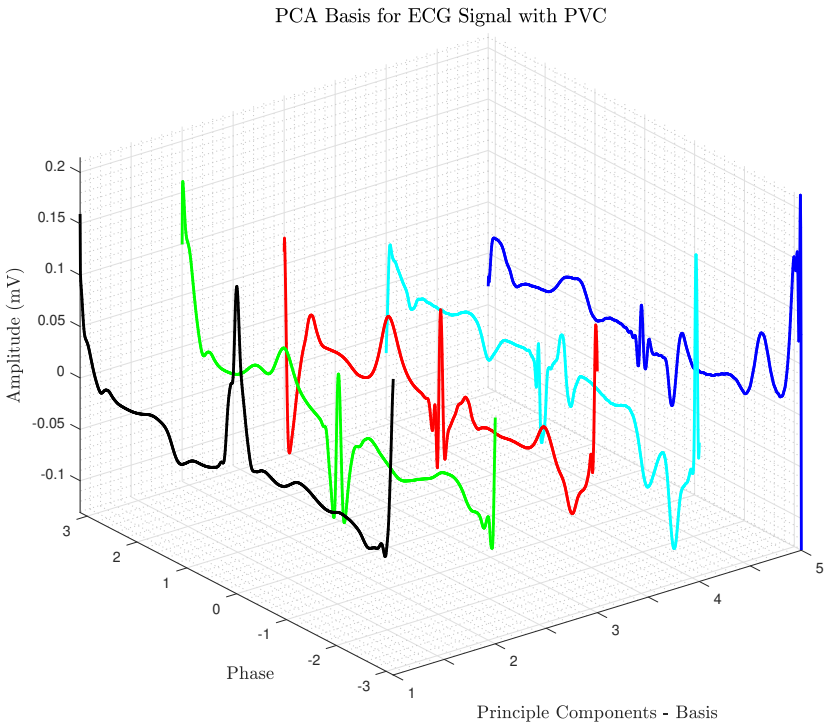


Fig. 5.6 Double Beat PCA basis signals derived from measured AECG data with PVCs.

5.5.2 PCA on Atrial Fibrillation

Many heart problems present as cardiac dysrhythmia visible in ECG data. Some arrhythmias may be dangerous and lead to emergency hospital admission. One of the most common arrhythmias is atrial fibrillation (AF) for which early and correct diagnosis is vital. AF occurs when the electrical impulses that trigger the heart are initiated from different places in the two upper chambers (atria). This leads to the atria contracting faster than normal. In AF, atria start to quiver or fibrillate and produce chaotic electrical signals. These chaotic signals pass to the AV node and on to the lower chambers. The ventricles then also beat irregularly and rapidly, but not as rapidly as the atria. The result is an irregular heart rhythm that is inefficient at pumping blood. AF can increase the risk of stroke dramatically as it increases the risk of a blood clot forming inside the heart (Pollock, 1986). The heart physiology and ECG signal for normal heart rhythm and AF are compared in Figure 5.7. Many factors may increase

the risk of developing atrial fibrillation but the exact cause of AF is unknown. Associated conditions include:

- High blood pressure;
- Heart attack;
- Coronary artery disease;
- Congenital heart disease;
- Above 60 years of age.

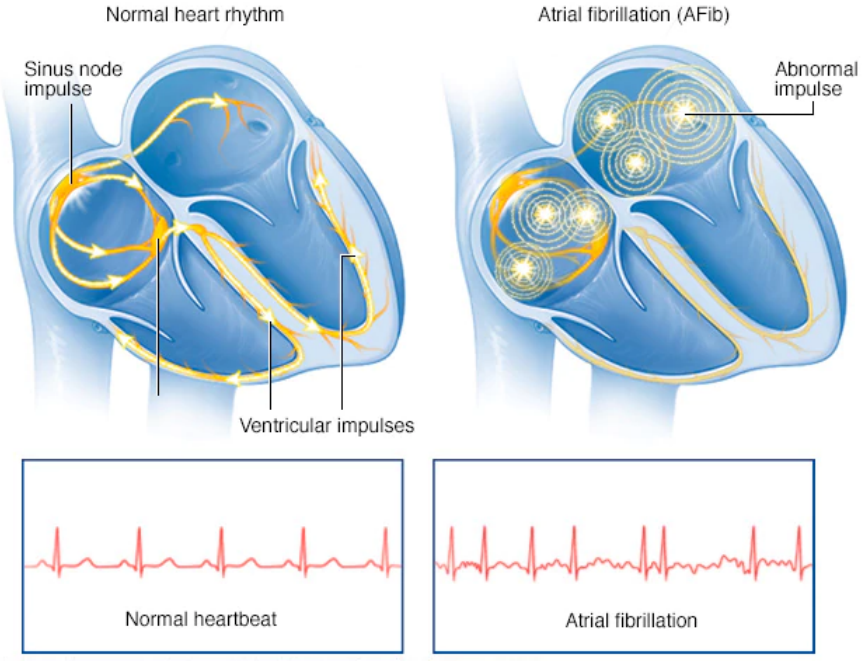


Fig. 5.7 Normal and AF heart rhythm, adopted from MayoClinic (2016).

A person can be completely unaware that their heart rate is irregular due to AF. However, a Holter AECG monitoring test can be used to diagnosis AF by monitoring the heart rhythm and electrical activity over 24 hours or more (Stahrenberg et al., 2010). Each year, millions

of people globally experience AF and this is a major cause of stroke (Bai et al., 2017). According to the British Heart Foundation report, updated in February 2018, around 1.3 million people in the UK have been diagnosed with AF and approximately 500,000 people are living with undiagnosed AF (British Heart Foundation, 2018). It is also the most common abnormal rhythm in elderly people.

Few ECG filtering methods are applicable to AF signals. EKF\EKS methods are predicated on the McSharry model which includes a P wave and no f waves. EKF\EKS methods interpret the missing P wave as noise. Recently, an EKF\EKF variant has been produced, EKF5, based on a model with no P wave and a parameterised f wave, see Section 3.8.

In this section, DB-PCAKF is applied to AF signals. The PCA basis functions are calculated using 200 heart cycles that contain AF, and Figure 5.8 illustrates first 5 basis signals.

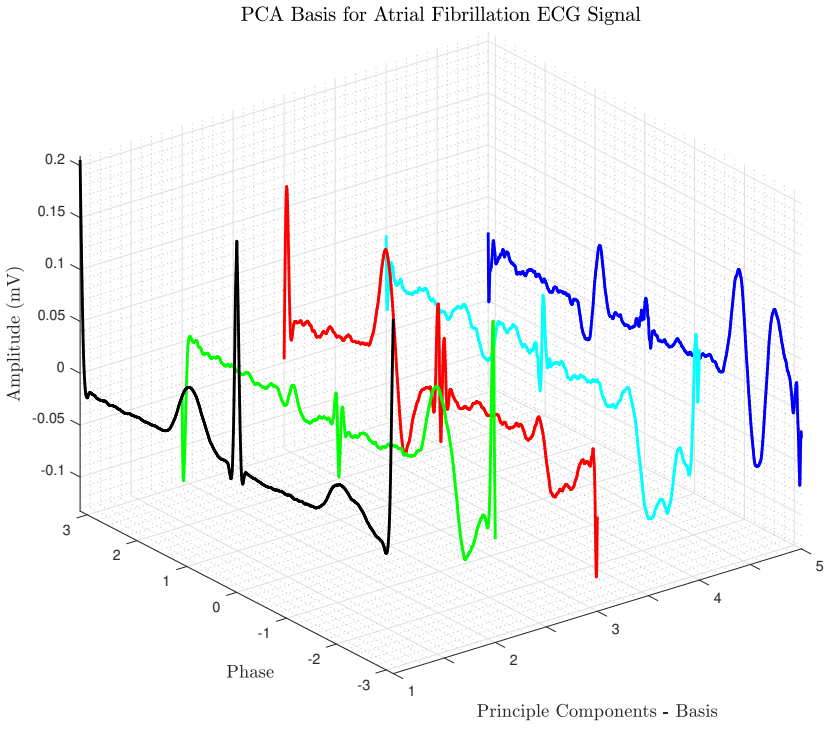


Fig. 5.8 PCA basis signals derived from ECG data with Atrial Fibrillation.

5.5.3 PCA on T-wave alternans

T-wave alternans (TWA) is a marker of repolarization instability in the heart. TWA is associated with significantly increased mortality risk and cardiac problems. TWA is associated ischemic cardiomyopathy (IC). Early detection of TWA is one of the most efficient ways to reduce cardiac mortality (Zipes and Libby, 2018). As it is difficult to see TWA in hospital based ECG recordings, Holter monitoring is commonly used. Figure 3.7 illustrates a typical TWA ECG signal. Figure 5.9 shows the PCA component extractions using synthetic TWA ECG.

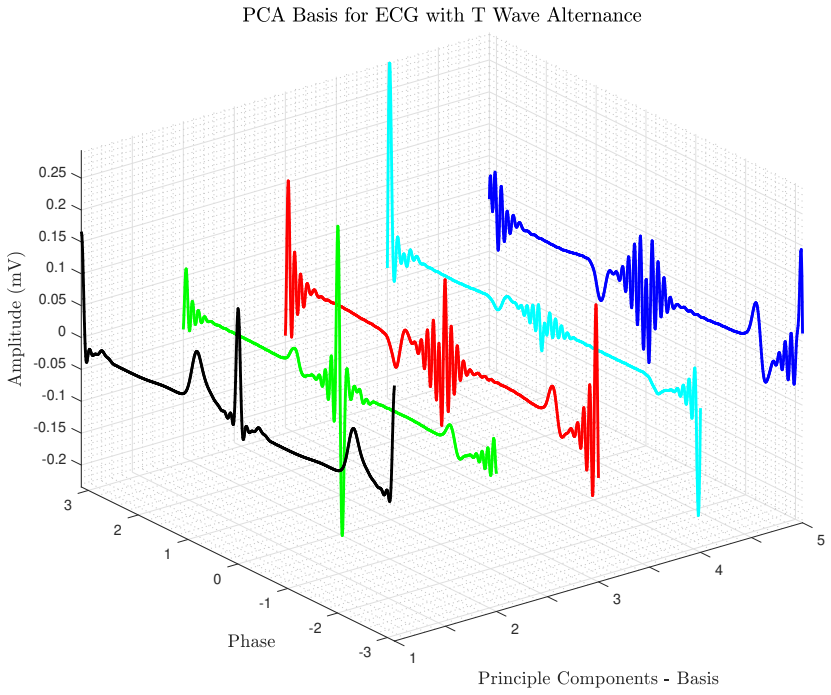


Fig. 5.9 PCA basis signals derived from T-wave alternans ECG.

5.6 PCA on Multi-channel AECG

The changes required to apply PCAKF simultaneously to signals from many electrodes is small but powerful. Signals from different electrodes are highly correlated. This allows

multi-electrode methods to predict the beat signal even in the temporary presence of high amplitude noise. It also allows errored data to be identified when unlikely combinations of signals occur. A PCA basis needs to be calculated for each electrode, using the data from the N_T training heart cycles. This is done independently for each electrode. Then the PCA fitting weights are calculated for each double heart cycle, for each electrode, using that electrode's PCA basis. These fitting weights are formed into a measured state vector of length $N_E \times N_B$:

$$\mathbf{X} \equiv (\alpha_i^l)^T = (\alpha_1^1, ..., \alpha_1^1, \alpha_1^2, ..., \alpha_{N_B}^2, ..., \alpha_1^{N_E}, \alpha_{N_B}^{N_E}) \tag{5.1}$$

The state vector for each of the N_T training heart cycles can be collected into a N_T by $N_E \times N_B$ array $X_{Training}$. The PCA parameter uncertainty $\mathbf{P}_1$ and the extrapolation uncertainty $\mathbf{Q}_1$ can be estimated from $X_{Training}$ in the same way as for a single electrode system. The information in these two arrays contains important information on inter-electrode dependencies that will constrain the Kalman filter.

The other important uncertainty is the measurement uncertainty quantified by the covariance$\mathbf{R}_l$ of the measurement noise. This is assumed to be diagonal with the same diagonal values for all parameters from the same electrode. As PCA basis fitting is carried out independently for each electrode-interval signal, there are no inter-electrode correlations in the measurement noise. The figure 5.10 shows a typical segment from 8 channels of a standard 12 lead AECG system collected using Holter system and figure 5.11 illustrates first 5 PCA's of multichannel data.

Fig. 5.10 Typical 12 lead AECG Signal.

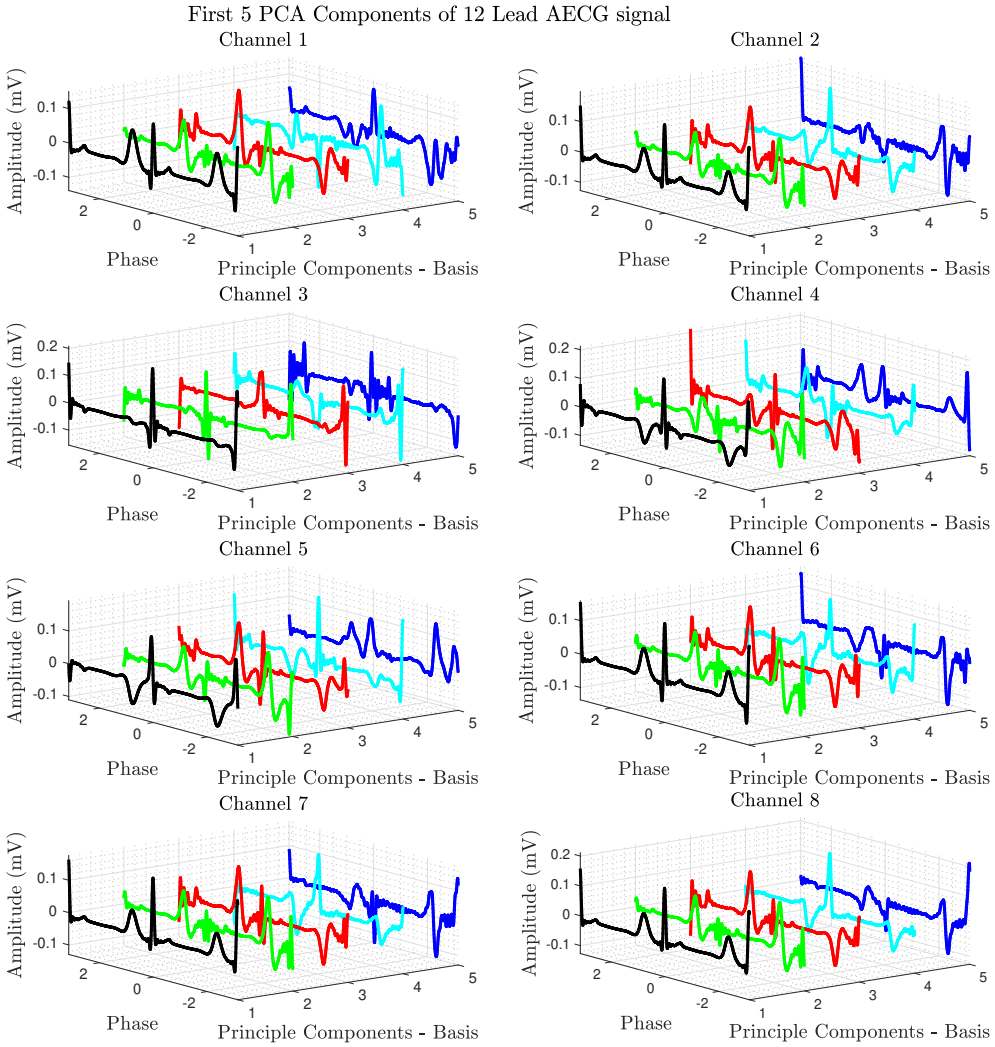


Fig. 5.11 PCA Components of 12 lead Multichannel AECG.

5.7 Results

The following sections illustrate the results of filtering normal and abnormal ECGs using DB-PCAKF. Both single and multichannel analyses are provided. Figure 5.12 presents the overview of the tests which are reported in this chapter. Single beat PCAKF framework is used as a benchmark method to compare with double beat PCAKF in this chapter. The PCA components, extracted in the previous sections, will be used in the following subsections, with DB-PCAKF framework to extract different types of cardiac signals including normal sinus rhythm, PVCs, AF, TWA and recorded AECG data with six different phases.

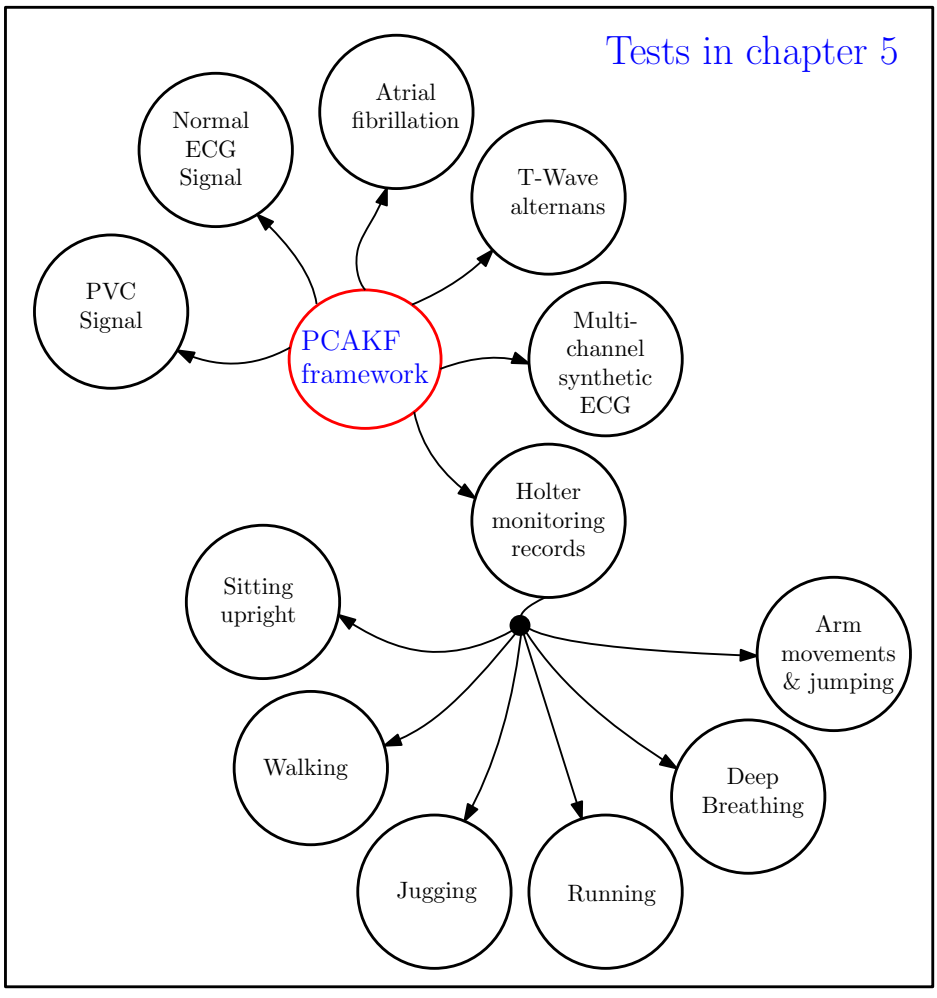


Fig. 5.12 The tests performed in this chapter using DB-PCAKF.

5.7.1 Comparison of PCAKF Methods on Normal Sinus Rhythm

In this section, synthetic ECG data is contaminated with added pink noise to compare the performance of the DB and SB-PCAKF, both applied to a single channel of data. The two minutes of synthetic ECG with a sample rate of 512 Hz and -2 dB APN is filtered using both SB-PCAKF and DB-PCAKF. Table 5.1 lists quality measures derived from two minutes of signal while the Figure 5.13 presents typical results for filtering. In this example the discontinuities introduced by the SB algorithm are clear and the DB algorithm performs significantly better than SB in all quality measures.

Methods	SNR	GoF(%)	MSEWPRD
SB-PCAKF	9.5	88.7	0.4
SB-PCAKF	13.5	95.6	0.2

Table 5.1 : Typical filtering results for two minutes of synthetic ECG data with APN with an initial SNR of -2 dB.

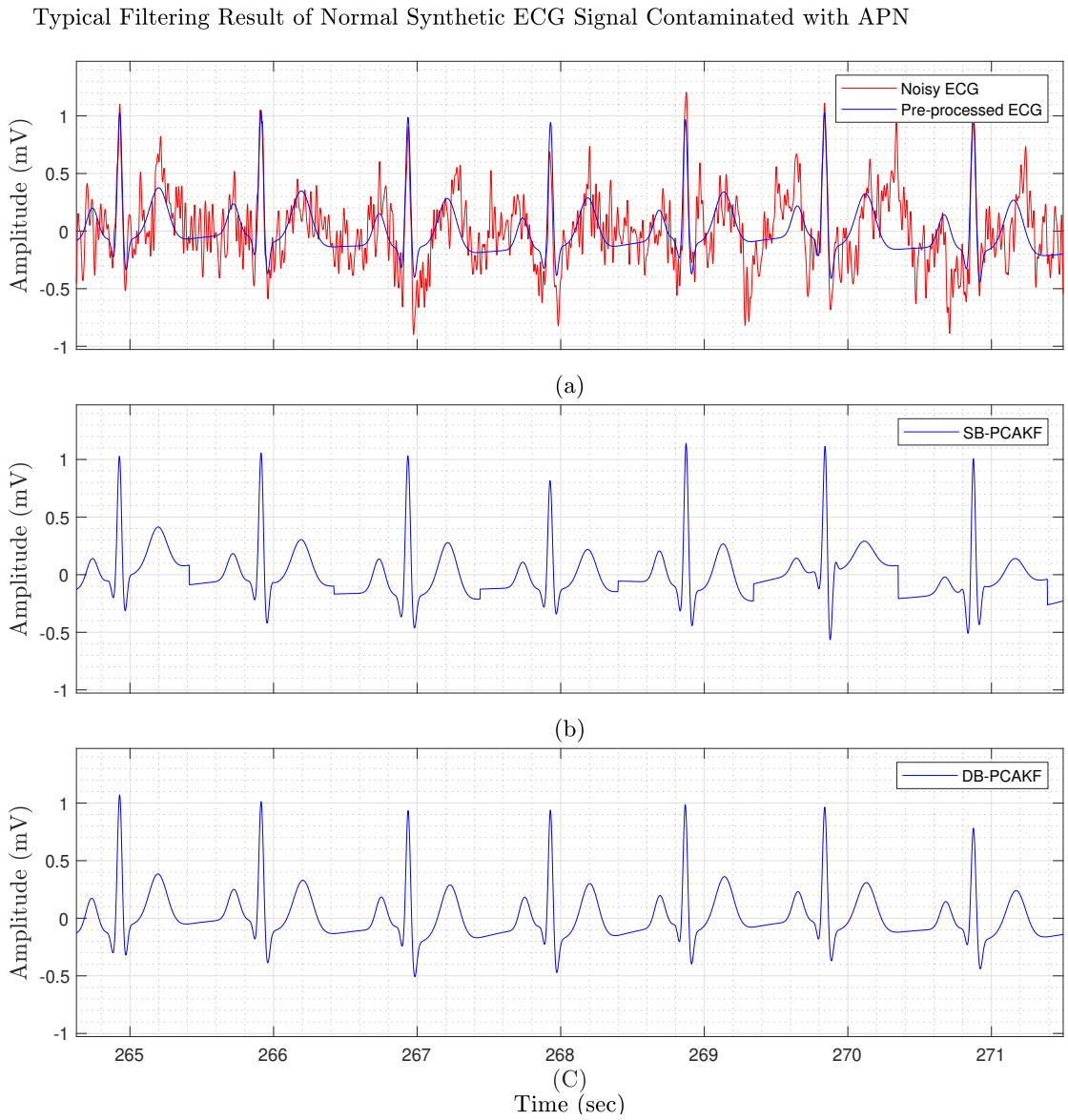


Fig. 5.13 Typical filtering results on two minutes of synthetic ECG data with APGN at -2 dB: (a) pre-processed synthetic ECG before and after addition of noise, (b) SB-PCAKF, (c) DB-PCAKF.

5.7.2 DB-PCAKF on PVCs

This section compares SB and DB-PCAKF applied to a single channel of PVC ECG signal contaminated with AWGN to yield a SNR of -4 dB. The PVC ECG data is part of the MIT-BHI arrhythmia database (Goldberger et al., 2000, Moody, 2000a). The database contains 48, 30 minute segments of two channel ambulatory ECG signals with arrhythmia. The whole half-hour of dataset 119 has been used in this test. Figure 5.14 presents a typical 30 second segment of signal before and after filtering. Again, DB yields a smoother output that is closer to the original, clean ECG signal. Both methods identify the premature ventricular contractions as heart signal rather than noise and the negative excursions are present in the filtered data. This is because the signal including segments of PVC were included in the training data. In practise, a filter may require a PCA basis augmented with an extra basis signal to span PVC anomalies.

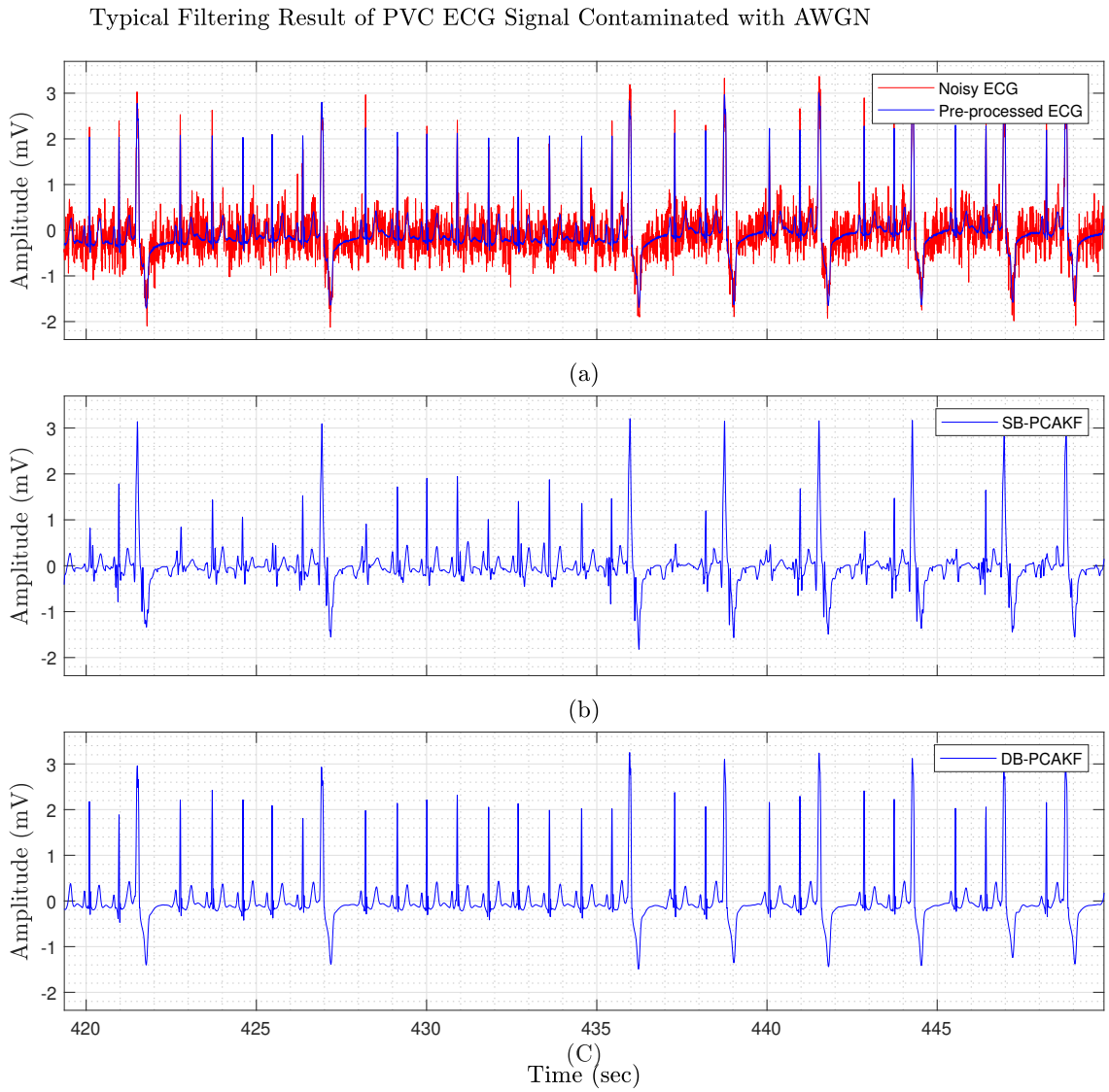


Fig. 5.14 Typical filtering results on two minutes of abnormal AECG data with AWGN at -4 dB: (a) pre-processed AECG signal before and after addition of noise, (b) SB-PCAKF, (c) DB-PCAKF.

5.7.3 DB-PCAKF on Atrial Fibrillation

This section applied SB and DB-PCAKF to a measured ECG data containing AF. The data is from the PhysioNet MIT-BHI atrial fibrillation database. The data was acquired using an AECG device with 250 samples per second. Two minutes of record 5121 with no significant noise were selected visually (Goldberger et al., 2000, Moody, 2000d). These data were contaminated with APN to yield a SNR of 3 dB. Figure 5.15 shows the results of filtering with SB and DB-PCAKF. Because the training data contained heart cycles exhibiting AF, the PCA basis recognises the anomalous signal and even reproduces the f waves. AF is characterised by a highly variable heart rate and this is visible in the variation of the inter R peak periods. The SB algorithm struggles with this due to the arbitrary heart cycle boundaries half-way between R peaks. However, the use of the R peak boundaries and the quadratic time transformation in the DB variant effectively normalises the heart cycles and yields a much more plausible filtered output.

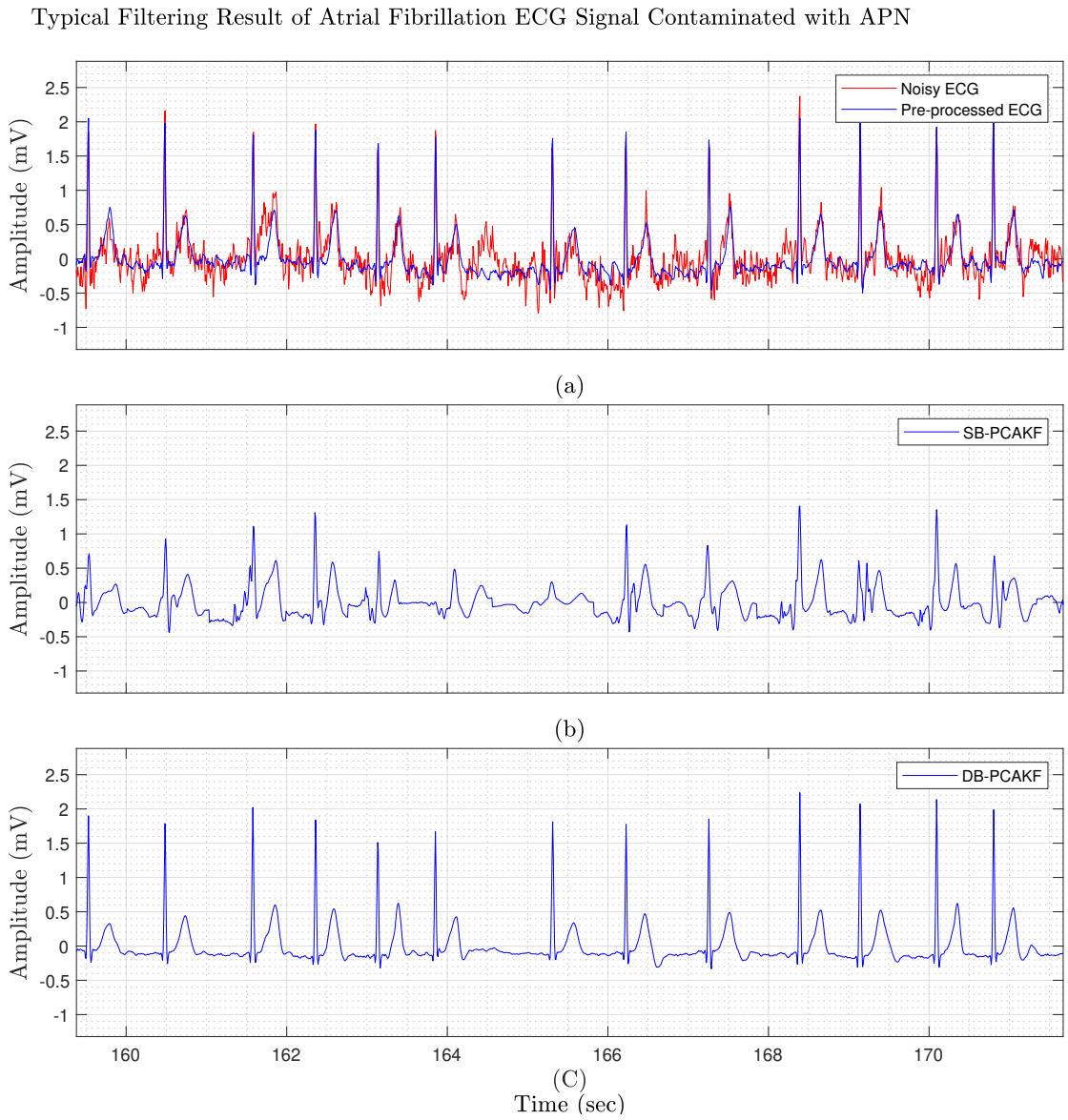


Fig. 5.15 Typical filtering results on two minutes of atrial fibrillation AECG data with AWGN at 3 dB: (a) pre-processed AECG signal before and after addition of noise, (b) SB-PCAKF, (c) DB-PCAKF.

5.7.4 DB-PCAKF on T Wave Alternans

This section compares SB and DB-PCAKF applied to a measured TWA ECG signal from the PhysioNet-TWA challenge data base which includes synthetic and real TWA ECGs (Goldberger et al., 2000, Moody, 2008). The database signal has been contaminated with APN with the same power as the signal to yield a SNR of 0 dB. Figure 5.16 shows typical filtering results for two minutes of signal and Table 5.2 lists the SNR improvements, GoF and MSEWPRD metrics.

Methods	SNR	GoF(%)	MSEWPRD
SB-PCAKF	8.2	84.8	0.3
SB-PCAKF	12.2	94.1	0.1

Table 5.2 Typical filtering results for two minutes of synthetic ECG with TWA data with APN at an initial SNR of 0 dB.

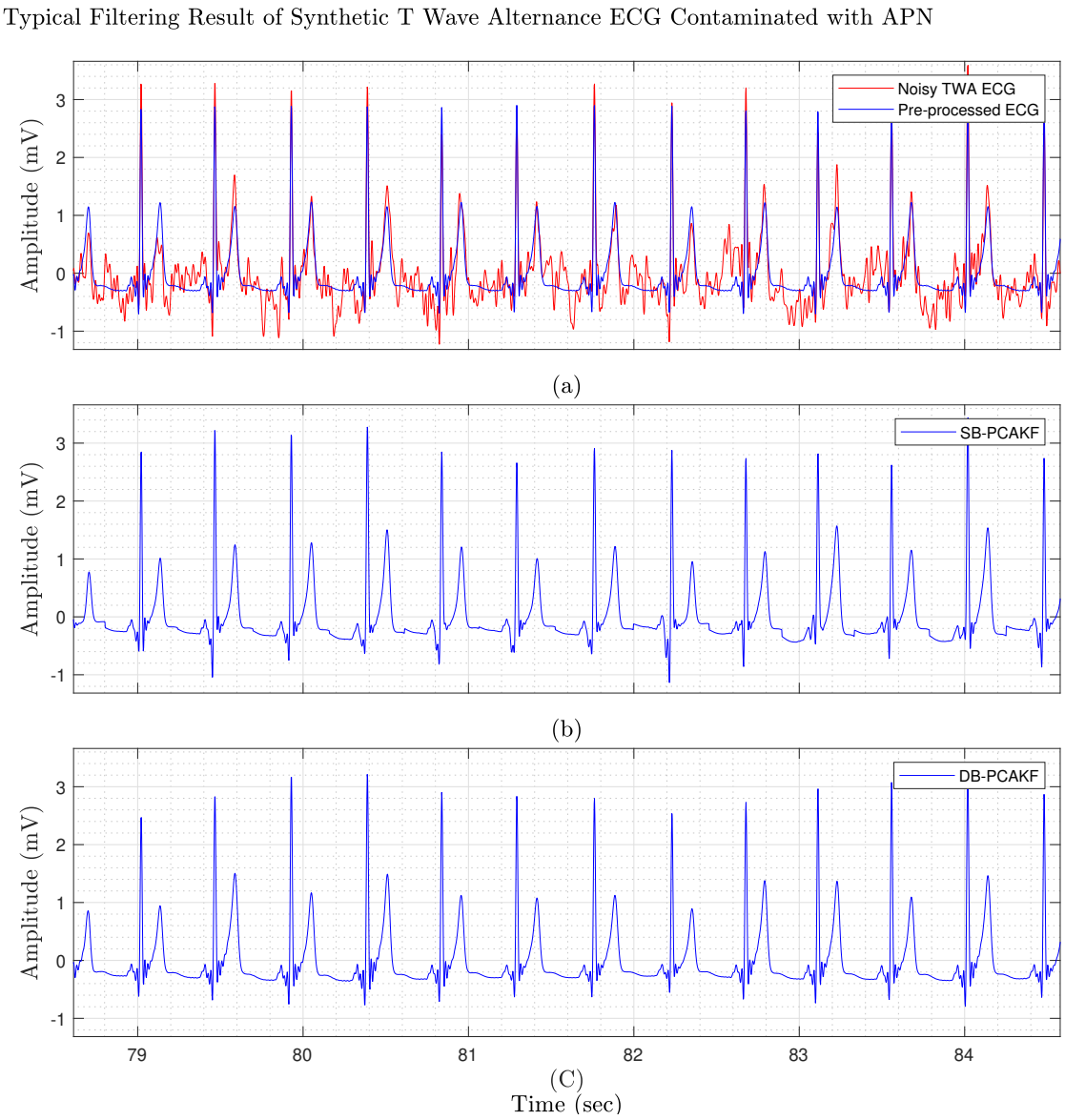


Fig. 5.16 Typical filtering results on two minutes of measured T wave alternans ECG (record TWA 91) with APN at 3 dB: (a) pre-processed AECG signal before and after addition of noise, (b) SB-PCAKF, (c) DB-PCAKF.

5.7.5 DB-PCAKF on Multi-channel Synthetic ECG Data

Previous sections have demonstrated consistent and better performance of the DB algorithm, and therefore future works will only use the DB variant. This section applied DB-PCAKF to multichannel data. Synthetic 8 channel ECG data has been produced by replicating a single channel ECG eight times. In this case the multichannel ECG signals are fully correlated and so test results are the best that could be obtained. All channels have been contaminated with additive white Gaussian noise. Noise with ten times the power of the signal have been added to channel 2 to yield a SNR of -10 dB. The other seven channels are relatively clean with SNRs of +10 dB. Figure 5.17 shows results of multichannel filtering. Applying DB-PCAKF to the channels individually increases the SNR on the clean channels by an average of 9.3 dB while Channel 2, with an initial SNR 20 dB lower, the SNR has been improved by 16.3 dB. Much of the noise rejection comes from the projection onto the PCA bases. After single channel filtering, the noisy channel 2 is still 13 dB noisier than the initially clean channels. Multichannel DB-PCAKF processes all channels simultaneously and uses the channel correlation information. After filtering, the SNR on the clean channels has increased by an average of 10.6 dB while on Channel 2, the SNR has been improved by 22.2 dB. The cross-correlation information used by multichannel filtering has significantly improved the filtering results on all channels. The clinically important features of the signals are also better revealed using multichannel analysis. Average MSEWPRD for single channel filtering is 0.08 while for multichannel filtering it is 0.05. In this case, filter outputs are so clean that MSEWPRD figures are both very small and so practically the same.

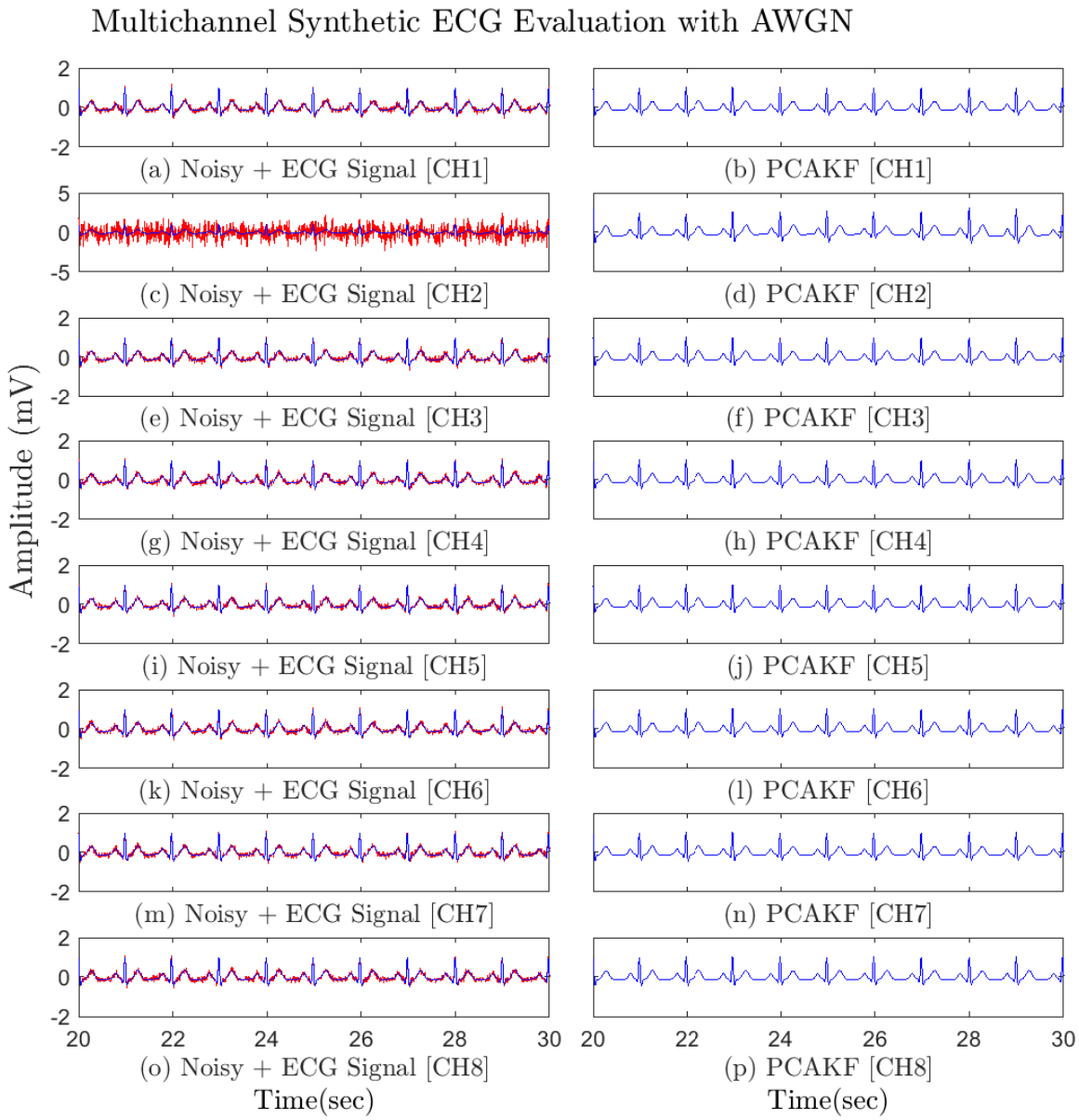


Fig. 5.17 Typical results of multi-channel analysis. Left column plots are the signal with added noise while right column plots are the filtered signals.

5.7.6 DB-PCAKF applied to AECG Data Collected during Different Physical Activities

Having presented the performance of DB-PCAKF on normal and abnormal ECG signals, this section looks at the filtering of AECG data collected during different physical activities. The sequence of activities has been chosen to yield increasing noise. The six activities yield recorded data with a range of noise sources and amplitudes. SNR improvement cannot be calculated, as the true signal is not known, so the filter output is assessed for clinical usefulness by a panel of experts who are either cardiologists or anaesthetists. The 12 lead Holter device collects 1000 samples per second. In the following section, only the first channel is presented, but the full multichannel results of this filtering are provided in Appendix B. For each individual participant, the channel PCA bases are calculated using training data selected from intervals spanning several activities. Data from the most extreme activities are typically too noisy to use, but data from walking has been included.

Sitting Upright

In this section the participant was asked to sit still and upright for 10 minutes. Figure 5.18 illustrates a typical example of an ECG signal obtained under these conditions, and the DB-MC-PCAKF filter output. In this case, the recorded data has a high initial SNR and filtering reveals very little that couldn't be seen in the recorded data.

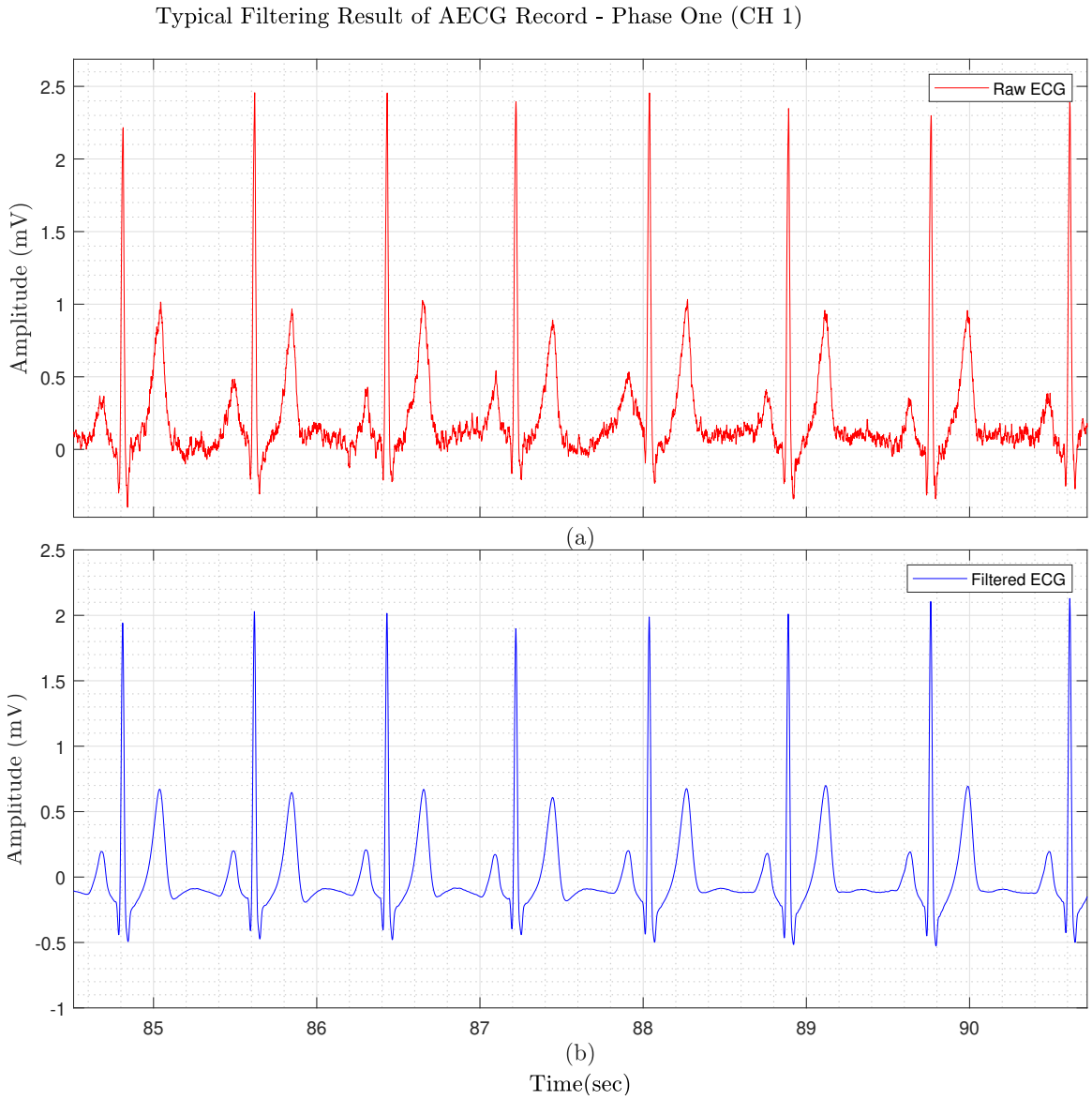


Fig. 5.18 Typical AECG recording while sitting upright: (a) recorded data, (b) filtered data.

Walking

In the second phase of data collection, the participant was asked to walk normally for 5 minutes. Figure 5.19 illustrates a typical ECG signal and filter output. The recorded data is noisier than when sitting still an upright due to electrode movement and small amounts of muscle noise. The filtered data reveals gradual changes in R peak amplitude that was not apparent in the recorded data.

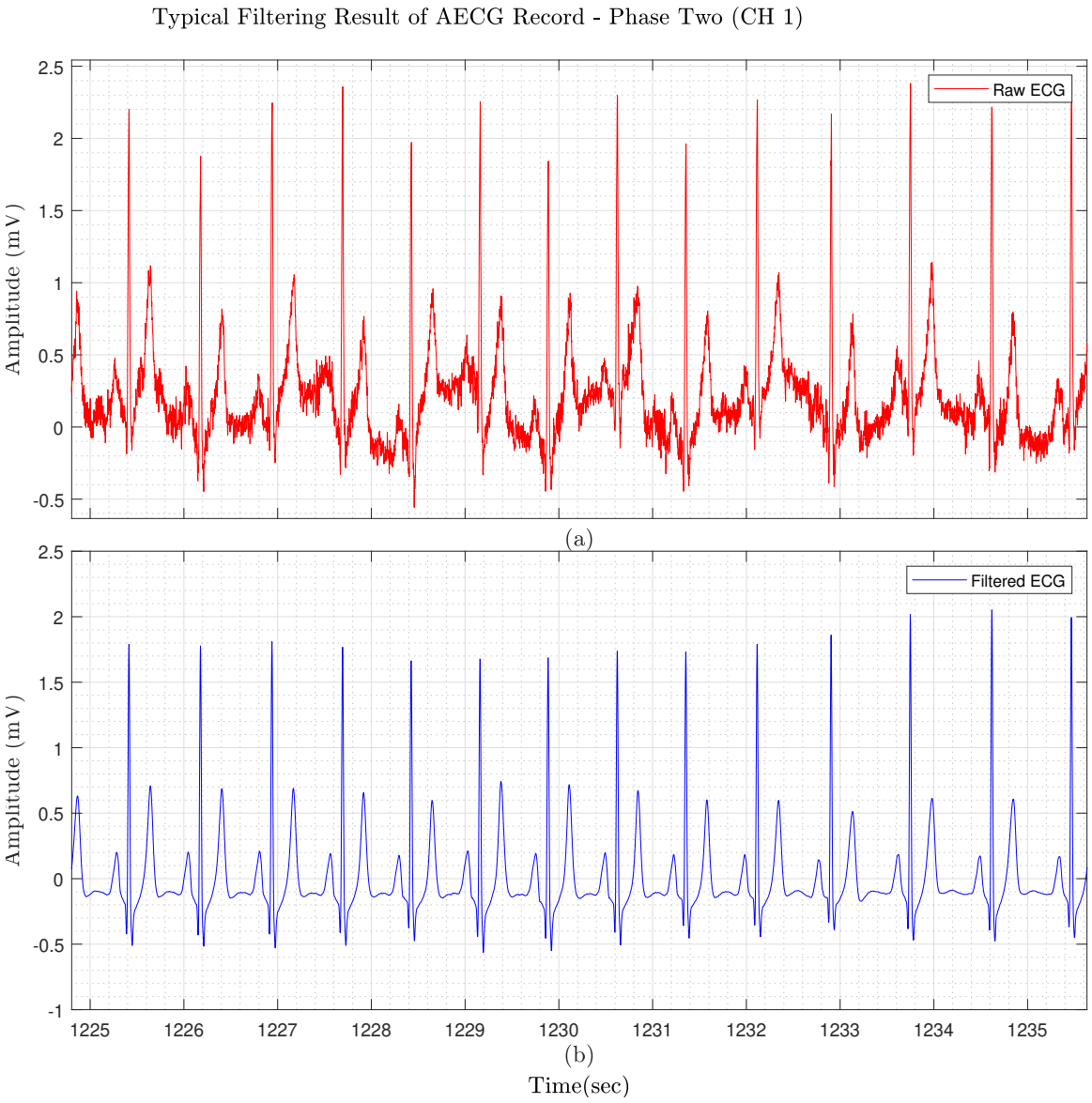


Fig. 5.19 Typical AECG recording while walking: (a) recorded data, (b) filtered data.

Jogging

Next, the participant was asked to jog to produce a signal with more noise. Figure 5.20 illustrates the typical portion of recorded and filtered signal. The signal is much noisier with features that could be mistaken for heart signals, such as the noise peak between the T and U waves of the first and second cycles, and between the third and fourth. The independence of noise on different channels allows the multichannel filter to distinguish between noise artefact and heart signal.

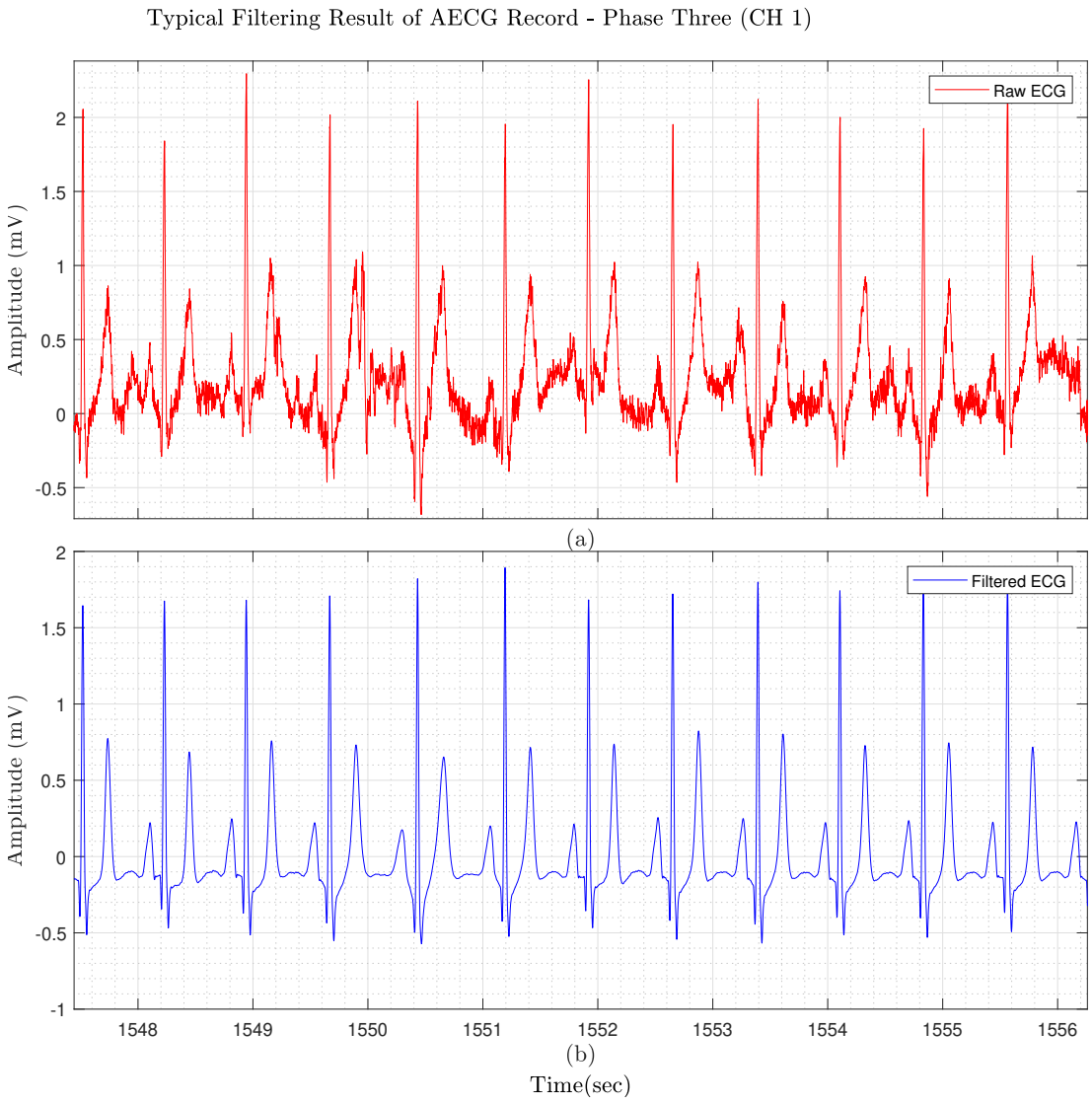


Fig. 5.20 Typical AECG recording while jogging (a) recorded data, (b) filtered data.

Running

The fourth activity was running. The participant was asked to run for 5 minutes. No other instructions were given. Figure 5.21 shows the recorded and filtered ECG signals. In these data the noise is so large that generally only the R and T peaks are identifiable in the recorded data. After filtering the smooth variation in R and T peak amplitudes is visible.

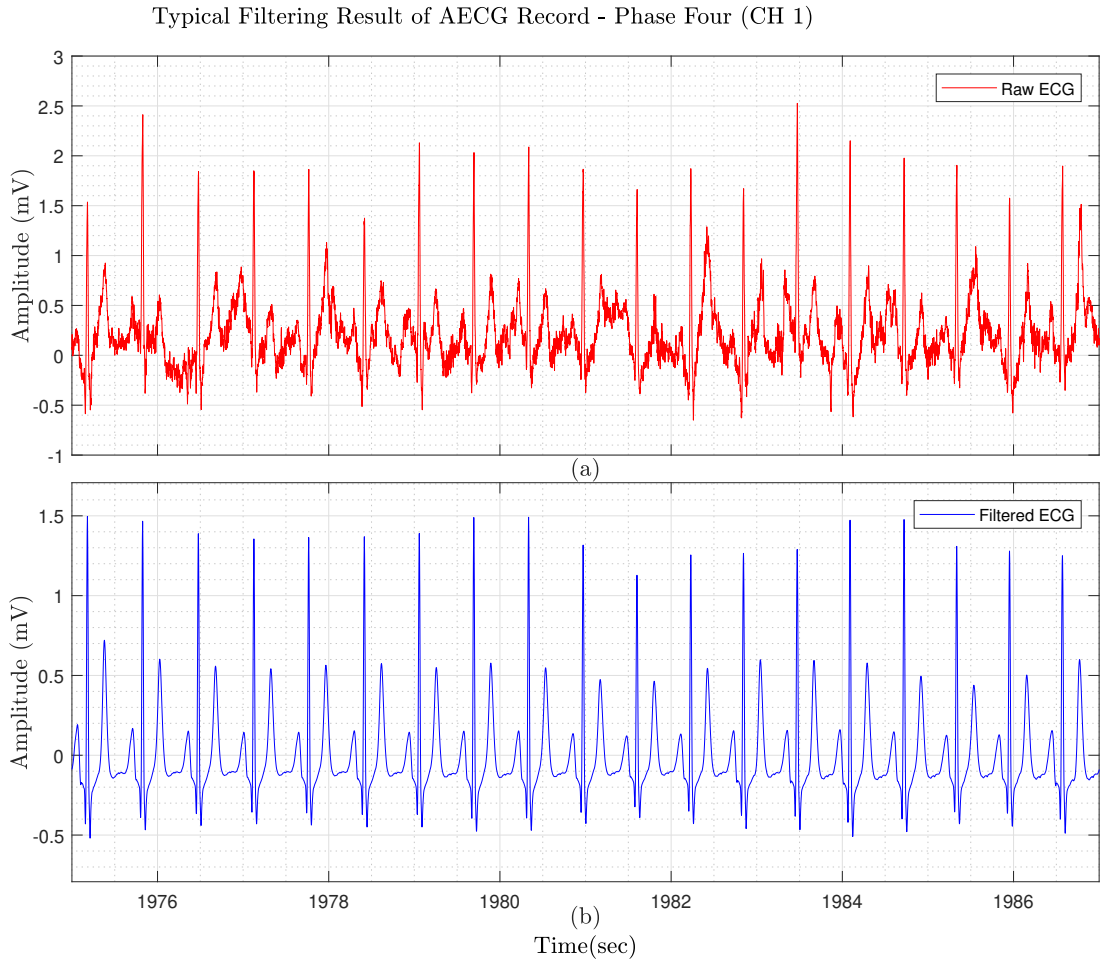


Fig. 5.21 Typical AECG recording while running: (a) recorded data, (b) filtered data.

Deep Breathing

After running, data was recorded while the participant was breathing deeply. The participant was asked to sit and take three deep breaths for three repetitions. Typical results for the breathing and filtered signals are provided in Figure 5.22. These signals are strongly contaminated with muscle noise from the diaphragm. This noise has been largely rejected by the filter. The muscle noise will be measured by all channels, but the PCA projection effectively eliminates this noise source as it was not present in the training data.

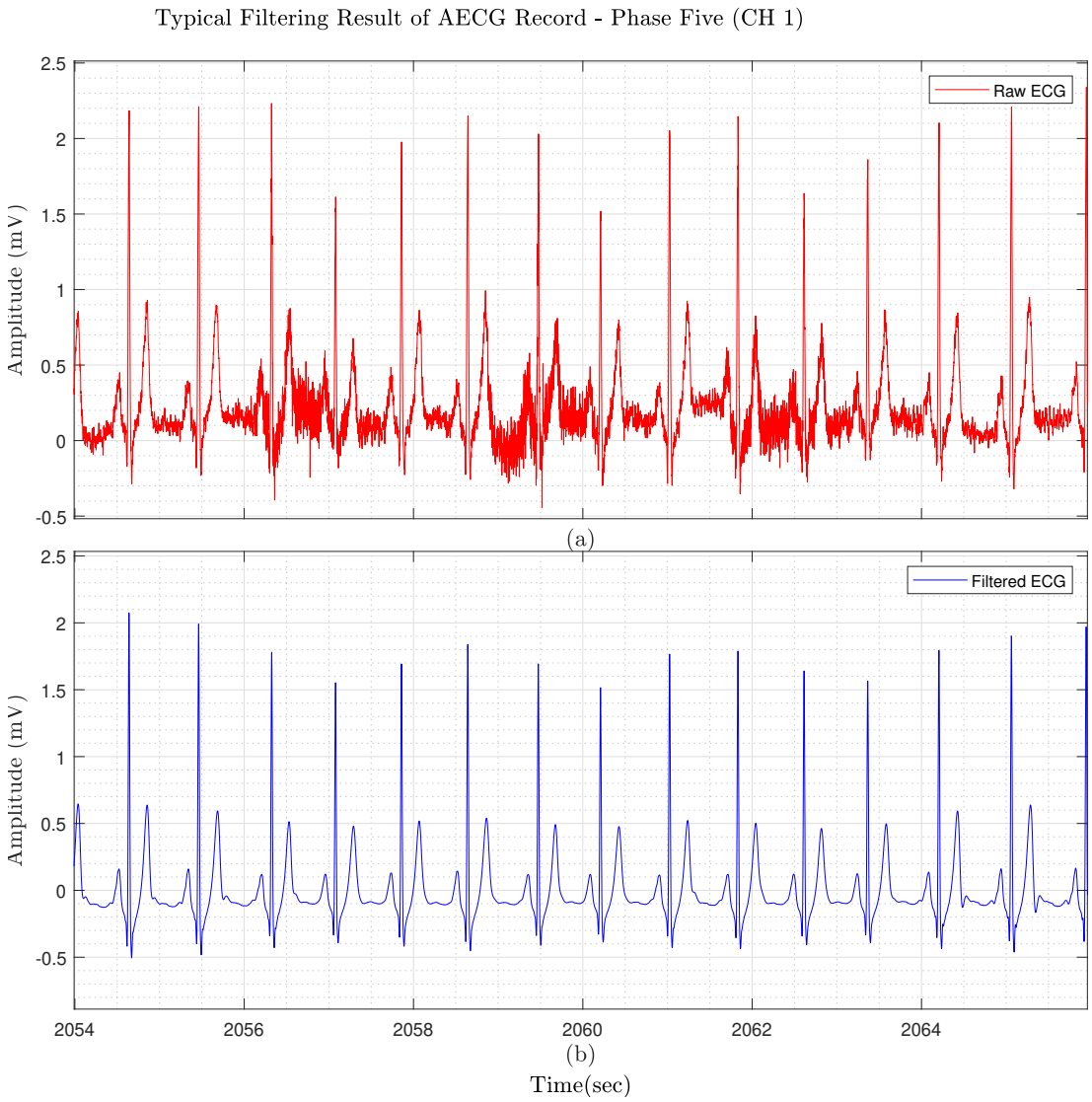


Fig. 5.22 Typical AECG recording (a) recorded data while breathing, (b) filtered data.

Arm Movements while Jumping

Additionally, participants were asked to jump and move their hands for a few seconds. This produces recordings with high amplitude and intermittent muscle and movement noise. Figure 5.23 shows the typical recorded data. Multichannel DB-PCAKF has produced consistent results despite the highly fluctuating SNR and noise spectra across the recording.

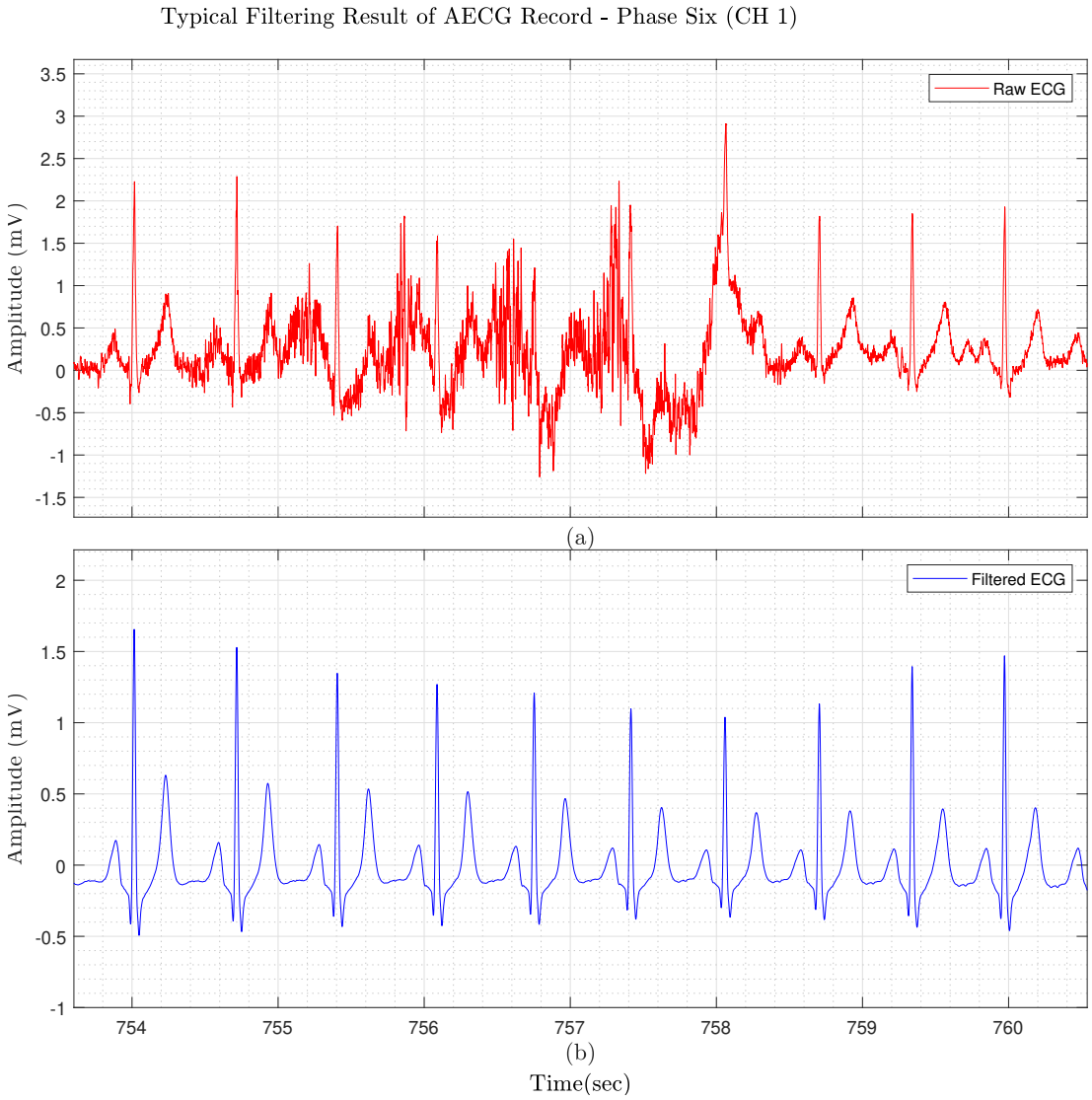


Fig. 5.23 Typical AECG recording (a) recorded data while jumping with arm movements, (b) filtered data.

5.7.7 R Peak Detections

All the ECG filtering methods presented in this project, both PCA and EKF, rely fundamentally on the detection of R peaks for their operation. Two R peak detection algorithms have been evaluated in conjunction with the filtering methods. The two methods are the standard Pan-Tompkin algorithm and the other one used by Clifford used at University of Oxford [private correspondence]. Errors in R peak detection lead to erroneous filter outputs. Typically the output signal is clearly anomalous and is unlikely to be mistaken for a real heart signal, but may obscure a real anomaly that occurs at the same time. Typical results of R peak detections and filtering are provided in Figure 5.24.

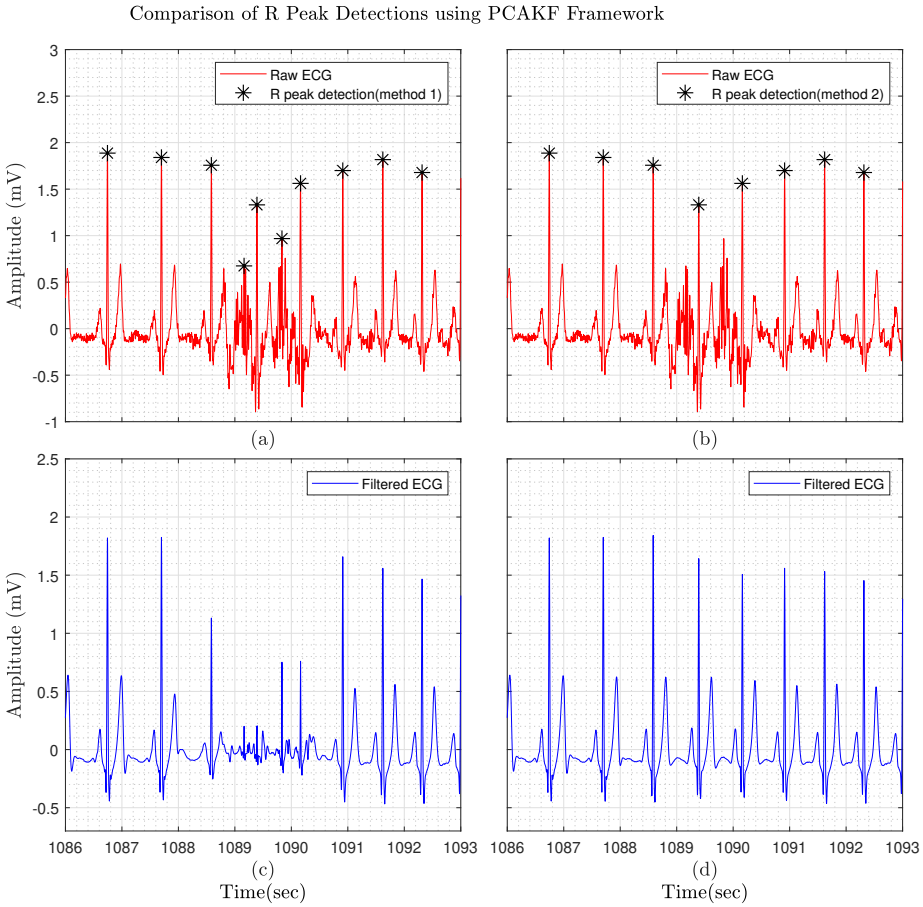


Fig. 5.24 TR peak detection comparison while using PCAKF. (a) Sameni's R peak detection algorithm (b) Pan-Tompkin R peak detection, (C) PCAKF result for Sameni's R peak detection, (d) PCAKF result for Pan-Tompkin R peak detection.

5.8 5.8 Discussion

This chapter presented the double beat (DB) variant of the PCAKF algorithm to address the problem of discontinuities in the output of single beat PCAKF when applied to very noisy signals. DB-PCAKF was found to solve this problem and also provided much better results when filtering ECG signal with common and clinically important abnormalities. Much of the improved performance was due to the quadratic time transformation used to normalise double beat intervals. The DB-PCAKF algorithm was shown to be applicable to ECG signals including abnormalities such as PVC, AF and TWL. The filter removed noise while retaining the abnormal heart signal.

A second major innovation was the multichannel PCAKF algorithm. This uses correlations between PCA projection weights on different channels, to reject noise and enhance ECG signals. Tests showed that the multichannel algorithm performed significantly better than applying the single channel algorithm to channels independently. The multichannel algorithm combines two major apriori advantages. PCA projection on each channel removes noise sources not present in the training data while the use of cross-channel correlations enhances the heart signal and further rejects noise that does not exhibit the same correlations.

Fifteen healthy men participated in this study and provided AECG data during a range of activities. The protocol was designed to yield data with a range of noise sources data range of amplitudes. The Multichannel DB-PCAKF filter has been applied to these data and typical results from one participant have been presented. The filter was able to remove most of the noise and yielded plausible heart signal variation, even when this was not apparent in the acquired data. The SNR improvement cannot be calculated for these data as the true signal is not known. However, a panel of medical experts were asked about their opinion on the quality of the filtering and the clinical usefulness of the filter output. All the comments that were returned are in Appendix B and a selection are provided below:

"According to the results, the method provided is able to remove noise and artefacts from the recorded signal, it could be very useful for us when ECG signal is contaminated with noise and ECG components (P, QRS complex and T waves) are not visible. In the filtered results ECG components are easily visible."

"The PCAKF system is powerful for heart signal detection in the presence of intermittent and high amplitude signals. The PCAKF system clears the noisy ECG signal and interpretation ECG by PCAKF is very comfortable"

"Filtering of noisy signals are clear. This could be very useful during exercise stress test, as the ST segment during the test is not visible due to the noise but it is very important to see clear ECG on the test. "

"This could be a very useful technique during the CPR to remove noise."

Hospital based ECG measurements are often performed on calm and conscious individuals while they relax on a bed. However, many situations exist where heart signals need to be acquired under far more challenging conditions. Some clinically important situations include during:

- Cardiopulmonary resuscitation (CPR) during which ECG contaminated with movement noise;
- Exercise stress tests where participants are walking, jogging and running;
- Surgery with electrical tools such as an electronic kutter. Typically these tools are turned off to check the ECG signal;
- Recording ECG from people with Parkinson's disease;
- Surgery spinal anaesthesia can cause shivering during surgery and therefore ECG signals are contaminated with muscle noise;

- Monitoring of people in intensive care units (ICU) who experience breathing difficulties. Ideally, continuous ECG monitoring is required but muscle noise due to breathing is common.

The PCAKF filter could be applied in all these situations. Finally, it was emphasised that both PCAKF and the gold standard EKF\EKS ECG filtering methods rely fundamentally on the quality of R peak detection.

5.9 Summary and Conclusions

Both the extension of SC-PCAKF to double beat (DB) and the use of cross-channel information in the multichannel version of DB-PCAKF, have shown significant improvements in the proposed algorithm. Multichannel DB-PCAKF performs significantly better than the gold standard EKF\EKS algorithms in tests with synthetic data, and real or synthetic noise. A panel of medical experts have stated that the proposed ECG filter yields signals where clinically important information have been revealed. The proposed filter is likely to have many clinical applications.

Chapter 6

Conclusions & Future Works

6.1 Conclusions

A system has been developed for the filtering of AECG signals to identify clinically important features of the heart signal in noisy data. Current gold standard methods use Extended Kalman filters with sample-by-sample extrapolation based on the McSharry sum of Gaussians ECG signal model. The methods developed as part of this project use linear Kalman filtering and extrapolation based on a sum of Principal Component basis signals. Unlike the gold standard methods, extrapolation is heart-cycle-by-heart-cycle. Several variants have been developed where basis signals span one or two heart-cycles, and are applied to single or multi-channel ECG data.

The McSharry sum of Gaussians ECG signal model was found to have major limitations. It is over-parameterised, not unique and a poor fit to many ECG signals; especially for abnormal signals. In comparison, a signal model expressed as a sum of Principal Component basis signals has many advantages: it uses the minimum number of parameters, they yield a mode with unique parameters, it provides a better fit to ECG signals and provides better separation between heart signals and noise.

Once a sum of PCA basis signals model is adopted, the Kalman filtering process becomes linear and considerably simpler than the non-linear Extended Kalman filters. This leads to significant computational gains and less complexity, which are important when signal analysis is devolved to a wearable device like a smartphone. The algorithm is also more stable and controllable. The PCA basis signals are critical and central to the filtering ability of PCAKF. It is vital that the signals span the range of heart signals present in the monitoring data. This is a source of potential problems when using the proposed system. The PCA basis signals are derived from a period of training data. This is important as signals will differ depending upon individual morphology, electrode placement and heart condition. It is particularly important when multi-channel data is to be analysed. The training data may not span abnormal heart activity that only occurs in special conditions, such as during activity. In practise, the PCA basis signals may need to be augmented with signals designed to span typical abnormalities.

McSharry introduced the concept of heart-cycle phase. Phase is a surrogate for time which increases by 2π radians over each heart-cycle. The use of phase normalises heart cycles with respect to heart rate variations. The gold standard methods use a piecewise linear transformation from time to phase with multiples of 2π radians occurring at a specific point in each heart cycle i.e. the start of the P-wave. This has several disadvantages. The start of the P-wave is poorly defined and the piecewise linear transformation is not smooth. Several alternatives were tested for the single and double beat variants of PCAKF. Initially, cubic splines were tested with phases of $(2n+1)\pi$ radians occurring at R-peaks. This yielded a smooth transformation based on the well-defined R-peaks. However, for noisy data, or where heart anomalies occur, the transformation could be non-monotonic. Double-beat uses a quadratic transformation over each pair of consecutive cycles, spanning three R-peaks. Each heart-cycle is filtered twice, as the left and right hand part of a pair, and a tapered weighted average of the two filter outputs is formed. The quadratic transformation normalises away

much of the inter-cycle variation and allows a PCA basis that spans much more of the natural variation. The transformation is always monotonic. Like the piecewise linear phase, the gradient is not continuous at heart cycle boundaries, but the filtered heart signal is.

Part of the filtering power comes from knowledge of the covariances of PCA basis weights, which are also calculated from the training data. Knowledge of covariances between PCA basis weights within heart cycles and across channels allows heart signals and noise to be distinguished. However, these correlations are derived from the training data and if the monitoring signal has different correlations then the assumed correlations may lead to poor filtering results.

The quality of R-peak detection is crucial to both the gold standard and PCAKF methods. For EKF\EKS methods, misidentified R-peaks lead to clearly wrong filter ouput in the signal portion around the misidentification. For PCAKF, misidentification in the training data is almost always identified and the heart-cycles effected are eliminated from contributing to the PCA basis function and weight covariance calculation. When processing the monitoring data, misidentification can produce an ouput that looks like an ECG signal with irregular heart rate. However, the algorithm also marks the heart-cycles effected as noisy and unreliable. R-peak detection is attracting further research attention and improvements in algorithms will feed through into better filtering output.

To compare the performance of PCAKF and EKF\EKS, a range of tests have been performed. A comprehensive set of numerical experiments have been performed using synthetic heart signals contaminated with synthetic noises. In these tests, the true heart signal is known exactly, and so SNR improvement and signal fit can be calculated and compared. Tests were also performed on clean measured ECG data with added noises. In this case the true signal is not known exactly as the clean ECG will contain some noise and artefacts. In some cases the PCAKF output was more plausible than the clean ECG signal, as the small amount of noise had been removed. For these tests, the SNR improvement and signal fit are

not exact, especially for very low amplitude additive noises. Tests were also performed on databases of ECG data exhibiting common heart anomalies. Finally, the methods were tested on real measured AECG data collected during a range of activities. In this case, the filter output has been scrutinised by experts who have provided opinions on the clinical usefulness of the filtered data.

Initial tests on the single-beat single-channel PCAKF method were encouraging but identified the discontinuity between heart cycles as an issue. It was also found to perform poorly on ECG data exhibiting anomalous heart signals. Based on these results the double-beat single-channel PCAKF method was developed. This algorithm eliminated the discontinuity and also performed much better on signals containing anomalies such as atrial fibrillation and T wave alternance.

Finally, multichannel versions of both single-beat and double beat PCAKF were developed. The big advantage of the multichannel variants is the use of cross-channel PCA weight covariances. During the calculation of the PCA basis signals using the trial data, the covariance between weights in the sum of PCA basis signals fit to each channel are calculated. These covariances provide important information to distinguish artefacts from heart signals. As all channels are measuring a different view of the evolution of the heart electric field, the signals measured on each channel have significant similarities encapsulated by the PCA weights covariances. For example, if an R-peak is delayed or of increased amplitude, this should be reflected in all channel measurements. Typically, artefacts will not exhibit the same covariances. In practise, the advantage provided my multichannel processing was less than expected. In the case where all channels experience independent additive noise with the same SNR, the SNR improvement was similar to that yielded by single channel filtering. The advantage was much clearer where a few channels had significantly worse SNR. In this case, the multichannel variants greatly increased the SNR on the noisy channels. In the extreme, if a channel is completely missing, such as when an electrode becomes disconnected, then

the multichannel variants can recreate the missing data using the other channels and the cross-channel covariance information.

6.2 Future Work

This project has developed a completely new ECG filtering tool with the specific goal of extracting clinically useful data from noisy AECG data. Future work can develop this idea further and apply the PCAKF method to other problems.

- The PCAKF method has been thoroughly laboratory tested. The next stage is to test the method when integrated into an AECG system and worn by participants for prolonged periods. This is likely to raise some new issues, both practical and signal processing, that will need to be addressed for a robust systems to be completed.

- The double-beat multichannel PCAKF method has been shown to be applicable to anomalous data including atrial fibrillation and Twave alternance. Other conditions may require small adjustments, such as the addition of basis signals so that anomalies are spanned by basis signals. Possibly a set of anomalous basis signals could be included as standard so that all standard anomalies can be identified

- The system automatically labels heart-cycles where the filter output is significantly different from the measured ECG signals. Significant difference indicates that the filter output less reliable. Further processing could be added to distinguish between anomalous heart signals not spanned by the PCA basis, and noise. An entropy measure may be able to do this.

- Many EKF\EKS variants attempt to extract heart signal portions, such as just the P-wave, QRS complex or T wave. It is likely that this could be achieved using PCAKF if the basis signals were edited to contain just the segment of interest.

- The PCA basis has the advantage of minimising the number of basis signals required to fit a given proportion of the variance in the training data. However, other orthogonal bases may have advantageous properties in some applications. For example, independent component analysis (ICA) has been applied to fetal ECG to separate mother and fetal heart signals. Using bases that span variation in these signals separately may yield improved filtering for fetal ECG.

- The current algorithms spate the processing into a training period and a monitoring period. If the monitoring period produced heart signals quite different from those encountered in the training data, then the method will yield a poor fit and label the output as unreliable. An alternative approach would be to iteratively improve the PCA basis as the monitoring data is processed. Algorithms exist that assimilate each new measurement and update the PCA basis calculation. Such algorithms have a memory parameter that put greater weight on the most recent measurements. A training period would still be required to initialise the process, but iterative refinement of the PCA basis signals, and weight covariances, could yield an algorithm that is more robust over long periods and which adapts to changes in the heart signals during, such as between sleeping and awake.

There are many research avenues yet to be explored. However, PCAKF is a significant improvement on the current gold standard algorithms and so it is important that the method is adopted by AECG manufacturers and users to achieve the maximum benefit for mankind. It is expected that the algorithm will be commercialised by a large medical equipment company, such as Phillips or General Electric (GE), for rapid introduction into standard practices globally.

References

Agante, P. M. and De Sá, J. P. M. (1999). ECG noise filtering using wavelets with soft-thresholding methods. In *Computers in Cardiology, 1999*, pages 535–538. IEEE.

Akhbari, M., Ghahjaverestan, N. M., Shamsollahi, M. B., and Jutten, C. (2018). ECG fiducial point extraction using switching Kalman filter. *Computer Methods and Programs in Biomedicine*, 157:129–136.

Akhbari, M., Shamsollahi, M. B., and Jutten, C. (2013a). ECG fiducial points extraction by extended Kalman filtering. In *2013 36th International Conference on Telecommunications and Signal Processing (TSP)*, pages 628–632. IEEE.

Akhbari, M., Shamsollahi, M. B., and Jutten, C. (2013b). Fiducial points extraction and characteristicwaves detection in ecg signal using a model-based bayesian framework. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 1257–1261. IEEE.

Akhbari, M., Shamsollahi, M. B., and Jutten, C. (2014). Twave alternans detection in ecg using extended kalman filter and dualrate ekf. In *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, pages 2500–2504. IEEE.

Bai, Y., Wang, Y.-L., Shantsila, A., and Lip, G. Y. (2017). The global burden of atrial fibrillation and stroke: a systematic review of the clinical epidemiology of atrial fibrillation in asia. *Chest*, 152(4):810–820.

Barros, A. K., Mansour, A., and Ohnishi, N. (1998). Removing artifacts from electrocardiographic signals using independent components analysis. *Neurocomputing*, 22(1-3):173–186.

Behar, J., Oster, J., Li, Q., and Clifford, G. D. (2013). ECG signal quality during arrhythmia and its application to false alarm reduction. *IEEE transactions on biomedical engineering*, 60(6):1660–1666.

Braunwald, E. and Zipes, L. (2001). Heart disease 6th ed. *WB saunders*, pages 1301–8.

British Heart Foundation (2018). CVD STATISTICS – BHF UK FACTSHEET.

Chawla, M., Verma, H., and Kumar, V. (2006). Ecg modeling and qrs detection using principal component analysis.

Clifford, G., Tarassenko, L., and Townsend, N. (2001). One-pass training of optimal architecture auto-associative neural network for detecting ectopic beats. *Electronics letters*, 37(18):1126–1127.

Cohen, A. (1983). Biomedical Signal Processing: Time and Frequency Domains Analysis. Boca Raton, Fla.

Di Marco, L. Y., Duan, W., Bojarnejad, M., Zheng, D., King, S., Murray, A., and Langley, P. (2012). Evaluation of an algorithm based on single-condition decision rules for binary classification of 12-lead ambulatory ECG recording quality. *Physiological Measurement*, 33(9):1435–1448.

Garcia, T. B. (2014). *Introduction to 12-lead ECG: the art of interpretation*. Jones & Bartlett Publishers.

Garcia, T. B. and Miller, G. T. (2009). *Arrhythmia recognition: The art of interpretation*. Jones & Bartlett Publishers.

Ghahjaverestan, N. M., Shamsollahi, M. B., Ge, D., and Hernández, A. I. (2015). Switching kalman filter based methods for apnea bradycardia detection from ecg signals. *Physiological measurement*, 36(9):1763.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., and Mark, R. G. (1995). The PTB Diagnostic ECG Database.

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220. Circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

Hall, J. E. (2015). *Guyton and Hall textbook of medical physiology e-Book*. Elsevier Health Sciences.

Hampton, J. (2013). *The ECG Made Easy E-Book*. Elsevier Health Sciences.

Harvard-Medical School (2014). What do irregular heartbeats mean? - Harvard Health.

He, T., Clifford, G., and Tarassenko, L. (2006). Application of independent component analysis in removing artefacts from the electrocardiogram. *Neural Computing & Applications*, 15(2):105–116.

Hesar, H. D. and Mohebbi, M. (2017a). An Adaptive Particle Weighting Strategy for ECG Denoising Using Marginalized Particle Extended Kalman Filter: An Evaluation in Arrhythmia Contexts. *IEEE Journal of Biomedical and Health Informatics*, 21(6):1581–1592.

Hesar, H. D. and Mohebbi, M. (2017b). Ecg denoising using marginalized particle extended kalman filter with an automatic particle weighting strategy. *IEEE journal of biomedical and health informatics*, 21(3):635–644.

Hurst, J. W. (1998). Naming of the waves in the ecg, with a brief account of their genesis. *Circulation*, 98(18):1937–1942.

Jousilahti, P., Laatikainen, T., Peltonen, M., Borodulin, K., Männistö, S., Jula, A., Salomaa, V., Harald, K., Puska, P., and Vartiainen, E. (2016). Primary prevention and risk factor reduction in coronary heart disease mortality among working aged men and women in eastern finland over 40 years: population based observational study. *bmj*, 352:i721.

Kestler, H. A., Haschka, M., Kratz, W., Schwenker, F., Palm, G., Hombach, V., and Hoher, M. (1998). De-noising of high-resolution ecg signals by combining the discrete wavelet transform with the wiener filter. pages 233–236. IEEE.

Kher, R., Pawar, T., and Thakar, V. (2014). Comparative analysis of pca and wavelet based motion artifact detection and spectral characterization in w-ecg. *WSEAS Transactions on Signal Processing*, 10:116–123.

Kohler, B.-U., Hennig, C., and Orglmeister, R. (2002). The principles of software qrs detection. *IEEE Engineering in Medicine and Biology Magazine*, 21(1):42–57.

Kotas, M. (2006). Application of projection pursuit based robust principal component analysis to ecg enhancement. *Biomedical signal processing and control*, 1(4):289–298.

Laguna, P., Jane, R., Meste, O., Poon, P. W., Caminal, P., Rix, H., and Thakor, N. V. (1992). Adaptive filter for event-related bioelectric signals using an impulse correlated reference input: comparison with signal averaging techniques. *IEEE transactions on biomedical engineering*, 39(10):1032–1044.

Lander, P. and Berbari, E. J. (1997). Time-frequency plane wiener filtering of the high-resolution ecg: development and application. *IEEE transactions on biomedical engineering*, 44(4):256–265.

Le, J. H., Otmani, A., Marijon, E., Waintraub, X., Lepillier, A., Chachoua, K., Lavergne, T., and Pornin, M. (2008). Atrial fibrillation: the most common arrhythmia. *Presse medicale (Paris, France: 1983)*, 37(5 Pt 2):821–826.

Lin, C., Mailhes, C., and Tourneret, J.-Y. (2010). P-and t-wave delineation in ecg signals using a bayesian approach and a partially collapsed gibbs sampler. *IEEE transactions on biomedical engineering*, 57(12):2840–2849.

Maglaveras, N., Stamkopoulos, T., Diamantaras, K., Pappas, C., and Strintzis, M. (1998). Ecg pattern recognition and classification using non-linear transformations and neural networks: a review. *International journal of medical informatics*, 52(1-3):191–208.

Manikandan, M. S. and Dandapat, S. (2007). Wavelet energy based diagnostic distortion measure for ECG. *Biomedical Signal Processing and Control*, 2(2):80–96.

Manikandan, M. S. and Dandapat, S. (2008). Multiscale Entropy-Based Weighted Distortion Measure for ECG Coding. *IEEE Signal Processing Letters*, 15:829–832.

Martis, R. J., Acharya, U. R., and Min, L. C. (2013). Ecg beat classification using pca, lda, ica and discrete wavelet transform. *Biomedical Signal Processing and Control*, 8(5):437–448.

MayoClinic (2016). Atrial fibrillation. www.mayoclinic.org/diseases-conditions/atrial-fibrillation/multimedia/img-20096412/. Last checked on Oct 01, 2018.

McSharry, P., Clifford, G., Tarassenko, L., and Smith, L. (2003). A dynamical model for generating synthetic electrocardiogram signals. *IEEE Transactions on Biomedical Engineering*, 50(3):289–294.

Moody (2008). T-wave alternans challenge database. https://physionet.org/pn3/twadb/. Last checked on Oct 01, 2018.

Moody, G. (2000a). MIT-BIH arrhythmia database. https://physionet.org/physiobank/database/mitdb/. Last checked on Oct 01, 2018.

Moody, G. (2000b). The MIT-BIH noise stress test database. https://www.physionet.org/physiobank/database/nsrdb/. Last checked on Oct 01, 2018.

Moody, G. (2000c). The mit-bih noise stress test database. https://physionet.org/physiobank/database/nstdb/. Last checked on Oct 01, 2018.

Moody, M. R. (2000d). The MIT-BIH atrial fibrillation database. https://physionet.org/physiobank/database/afdb/. Last checked on Oct 01, 2018.

Olmos, S., Garcia, J., Jane, R., and Laguna, P. (1999). Ecg signal compression plus noise filtering with truncated orthogonal expansions. *Signal Processing*, 79(1):97–115.

OpenStax, C. (2013). Conduction system of heart. Last checked on Oct 01, 2018.

Palaniappan, R. and Khoon, T. E. (2004). Uni-channel PCA for noise reduction from ECG signals. In *Proc 1st International Bioengineering Conference*, pages 436–439.

Paradis, N. A., Halperin, H. R., Kern, K. B., Wenzel, V., and Chamberlain, D. A. (2007). *Cardiac Arrest: The Science and Practice of Resuscitation Medicine*. Cambridge University Press, 2 edition.

Petrenas, A., Marozas, V., Sörnmo, L., and Lukosevicius, A. (2012). An echo state neural network for qrst cancellation during atrial fibrillation. *IEEE Transactions on Biomedical Engineering*, 59(10):2950.

Pollock, M. L. (1986). *Heart disease and rehabilitation*. Churchill Livingstone.

Rodríguez, R., Mexicano, A., Bila, J., Cervantes, S., and Ponce, R. (2015). Feature extraction of electrocardiogram signals by applying adaptive threshold and principal component analysis. *Journal of applied research and technology*, 13(2):261–269.

Romero, I. (2010). PCA-based noise reduction in ambulatory ECGs. In *Computing in Cardiology, 2010*, pages 677–680. IEEE.

Roonizi, E. K. and Sassi, R. (2016). Dominant atrial fibrillatory frequency estimation using an extended kalman smoother. In *Computing in Cardiology Conference (CinC), 2016*, pages 989–992. IEEE.

Roonizi, E. K. and Sassi, R. (2017). An extended bayesian framework for atrial and ventricular activity separation in atrial fibrillation. *IEEE journal of biomedical and health informatics*, 21(6):1573–1580.

Sameni, R. (2008a). *Extraction of Fetal Cardiac Signals from an Array of Maternal Abdominal Recordings*. Phd thesis, Sharif University of Technology.

Sameni, R. (2008b). Multi-channel synthetic ecg generator. http://www.mit.edu/~gari/CODE/ECGSYN/MATLAB/MULTI/Multi-Channel%20Synthetic%20ECG%20Generator.htm. Last checked on Oct 01, 2018.

Sameni, R., Clifford, G. D., Jutten, C., and Shamsollahi, M. B. (2007a). Multichannel ECG and Noise Modeling: Application to Maternal and Fetal ECG Signals. *EURASIP Journal on Advances in Signal Processing*, 2007(1):043407.

Sameni, R., Shamsollahi, M. B., and Jutten, C. (2008a). Model-based bayesian filtering of cardiac contaminants from biomedical recordings. *Physiological Measurement*, 29(5):595.

Sameni, R., Shamsollahi, M. B., and Jutten, C. (2008b). Model-based bayesian filtering of cardiac contaminants from biomedical recordings. *Physiological Measurement*, 29(5):595.

Sameni, R., Shamsollahi, M. B., Jutten, C., and Clifford, G. D. (2007b). A nonlinear bayesian filtering framework for ecg denoising. *IEEE Transactions on Biomedical Engineering*, 54(12):2172–2185.

Sarfraz, M., Li, F. F., and Khan, A. A. (2015). Independent component analysis methods to improve electrocardiogram patterns recognition in the presence of non-trivial artifacts. *Journal of Medical and Bioengineering Vol*, 4(3).

Sayadi, O. and Shamsollahi, M. B. (2007). Multiadaptive bionic wavelet transform: Application to ecg denoising and baseline wandering reduction. *EURASIP Journal on Advances in Signal Processing*, 2007(1):041274.

Sayadi, O. and Shamsollahi, M. B. (2008). Ecg denoising and compression using a modified extended kalman filter structure. *IEEE Transactions on Biomedical Engineering*, 55(9):2240–2248.

Sayadi, O. and Shamsollahi, M. B. (2009). A model-based Bayesian framework for ECG beat segmentation. *Physiological Measurement*, 30(3):335–352.

Sayadi, O., Shamsollahi, M. B., and Clifford, G. D. (2010). Robust detection of premature ventricular contractions using a wave-based bayesian framework. *IEEE Transactions on Biomedical Engineering*, 57(2):353–362.

Shafqat, S., Kelly, P., and Furie, K. L. (2004). Holter monitoring in the diagnosis of stroke mechanism. *Internal medicine journal*, 34(6):305–309.

Sharma, L. N., Dandapat, S., and Mahanta, A. (2010). Multiscale principal component analysis to denoise multichannel ECG signals. In *2010 5th Cairo International Biomedical Engineering Conference*, pages 17–20. IEEE.

Simon, D. (2006). *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley-Interscience, New York, NY, USA.

Stahrenberg, R., Weber-Kruger, M., Seegers, J., Edelmann, F., Lahno, R., Haase, B., Mende, M., Wohlfahrt, J., Kermer, P., Vollmann, D., et al. (2010). Enhanced detection of paroxysmal atrial fibrillation by early and prolonged continuous holter monitoring in patients with cerebral ischemia presenting in sinus rhythm. *Stroke*, 41(12):2884–2888.

Stanhope, K. L., Medici, V., Bremer, A. A., Lee, V., Lam, H. D., Nunez, M. V., Chen, G. X., Keim, N. L., and Havel, P. J. (2015). A dose-response study of consuming high-fructose corn syrup–sweetened beverages on lipid/lipoprotein risk factors for cardiovascular disease in young adults–. *The American journal of clinical nutrition*, 101(6):1144–1154.

Stridh, M. and Sornmo, L. (2001). Spatiotemporal qrst cancellation techniques for analysis of atrial fibrillation. *IEEE Transactions on Biomedical Engineering*, 48(1):105–111.

World Health Organazasion (2018). WHO | Cardiovascular diseases (CVDs).

Yan, H. and Li, Y. (2010). Electrocardiogram analysis based on the karhunen-loève transform. In *Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on*, volume 2, pages 887–890. IEEE.

Youseffi, M. and Achilleos, A. (2015). *Design and manufacturing of a single channel medical ECG device*. LAP LAMBERT Academic Publishing. ISBN : 978-3-659-78086-8.

Zigei, Y., Cohen, A., and Katz, A. (2000). The weighted diagnostic distortion (WDD) measure for ECG signal compression. *IEEE Transactions on Biomedical Engineering*, 47(11):1422–1430.

Zipes, D. and Libby, Bonow, M. T. G. (2018). *Braunwald's Heart Disease E-Book: A Textbook of Cardiovascular Medicine*. Elsevier Health Sciences.

Appendix A

Multi-Channel AECG Filtering

This chapter provides multi-channel filtering results using DB-PCAKF for different phases.

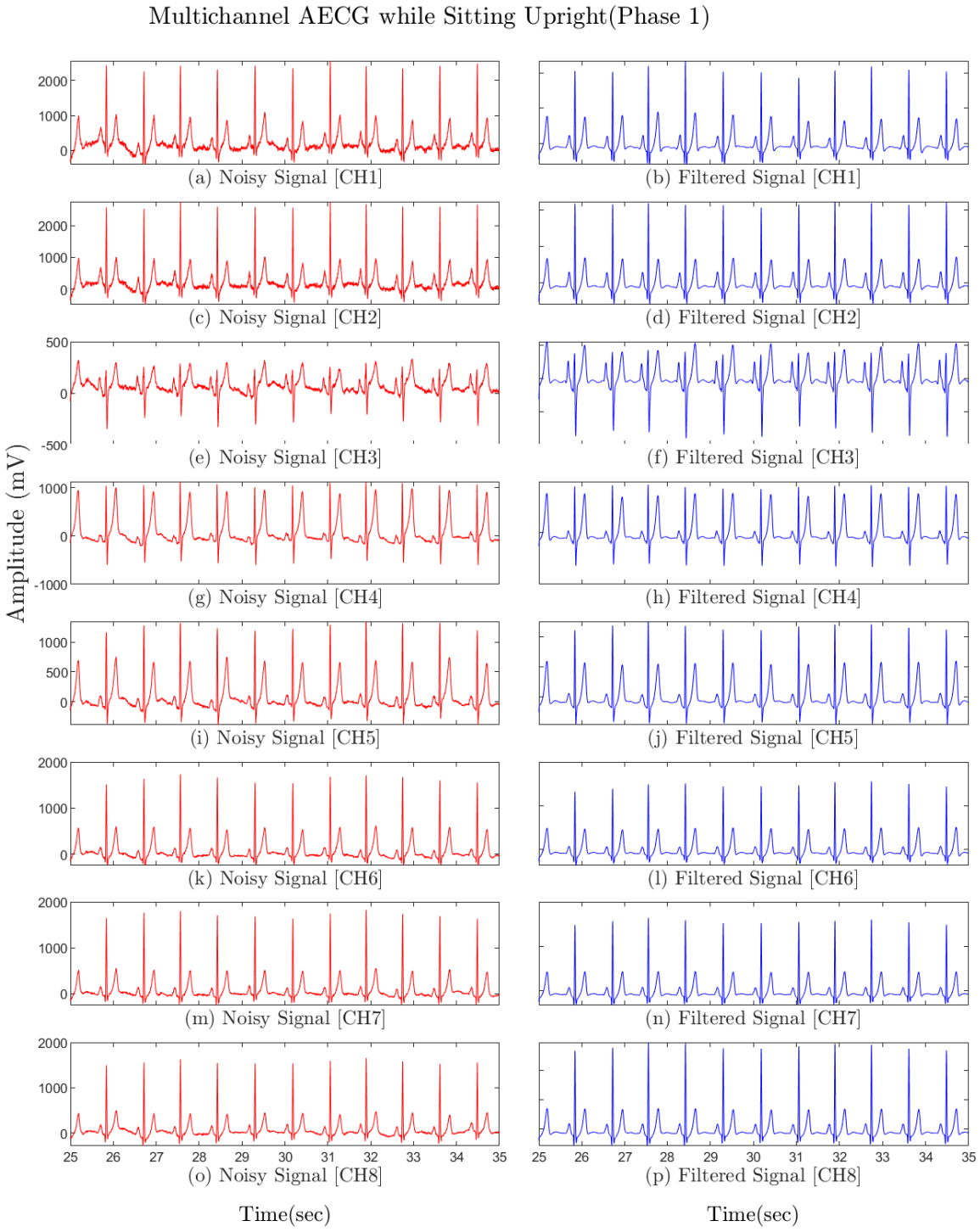


Fig. A.1 Typical results of multi-channel AECG recordings on phase one. Left column plots are the raw ECGs while right column plots are filtered signals.

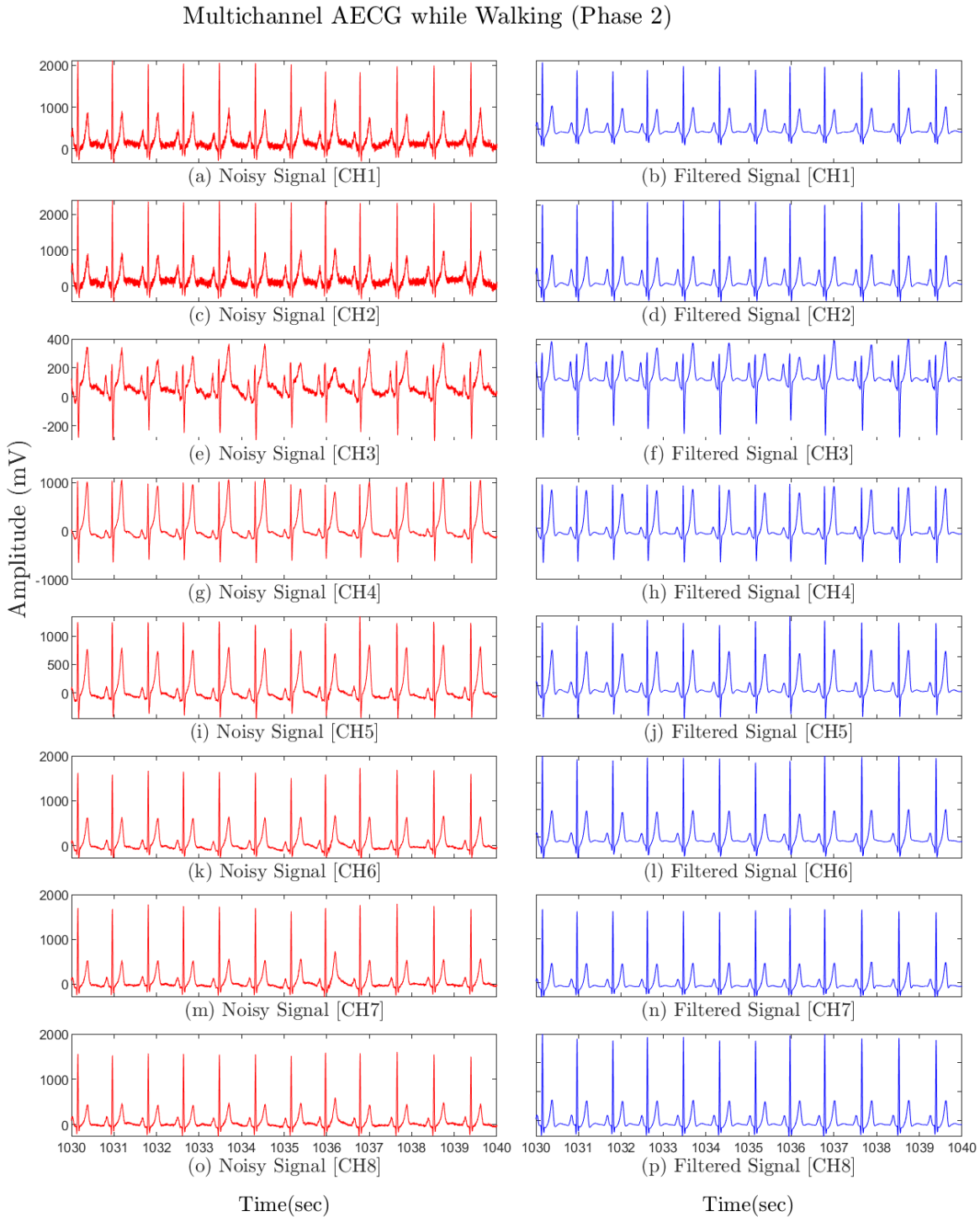


Fig. A.2 Typical results of multi-channel AECG recordings on phase two. Left column plots are the raw ECGs while right column plots are filtered signals

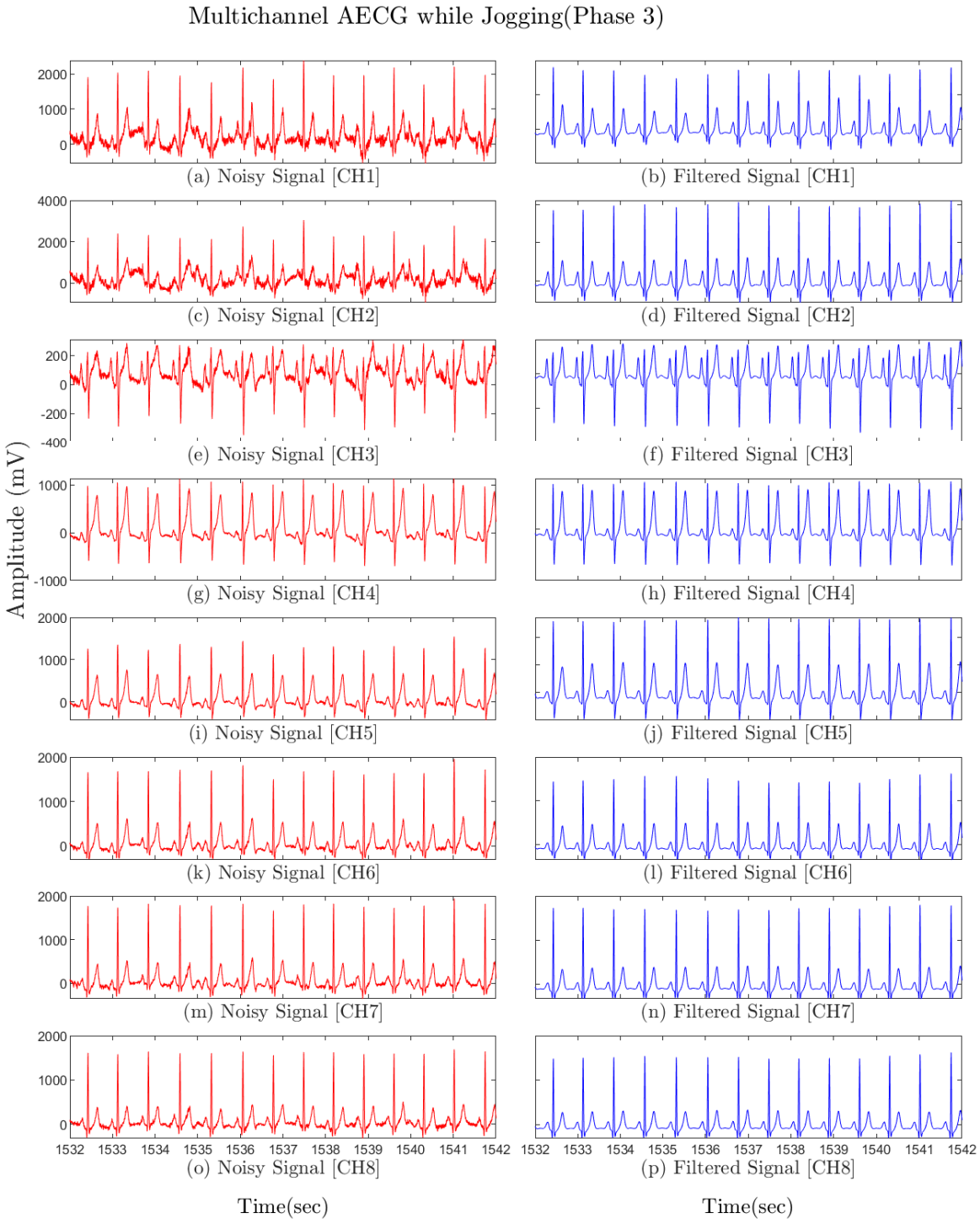


Fig. A.3 Typical results of multi-channel AECG recordings on phase three. Left column plots are the raw ECGs while right column plots are filtered signals

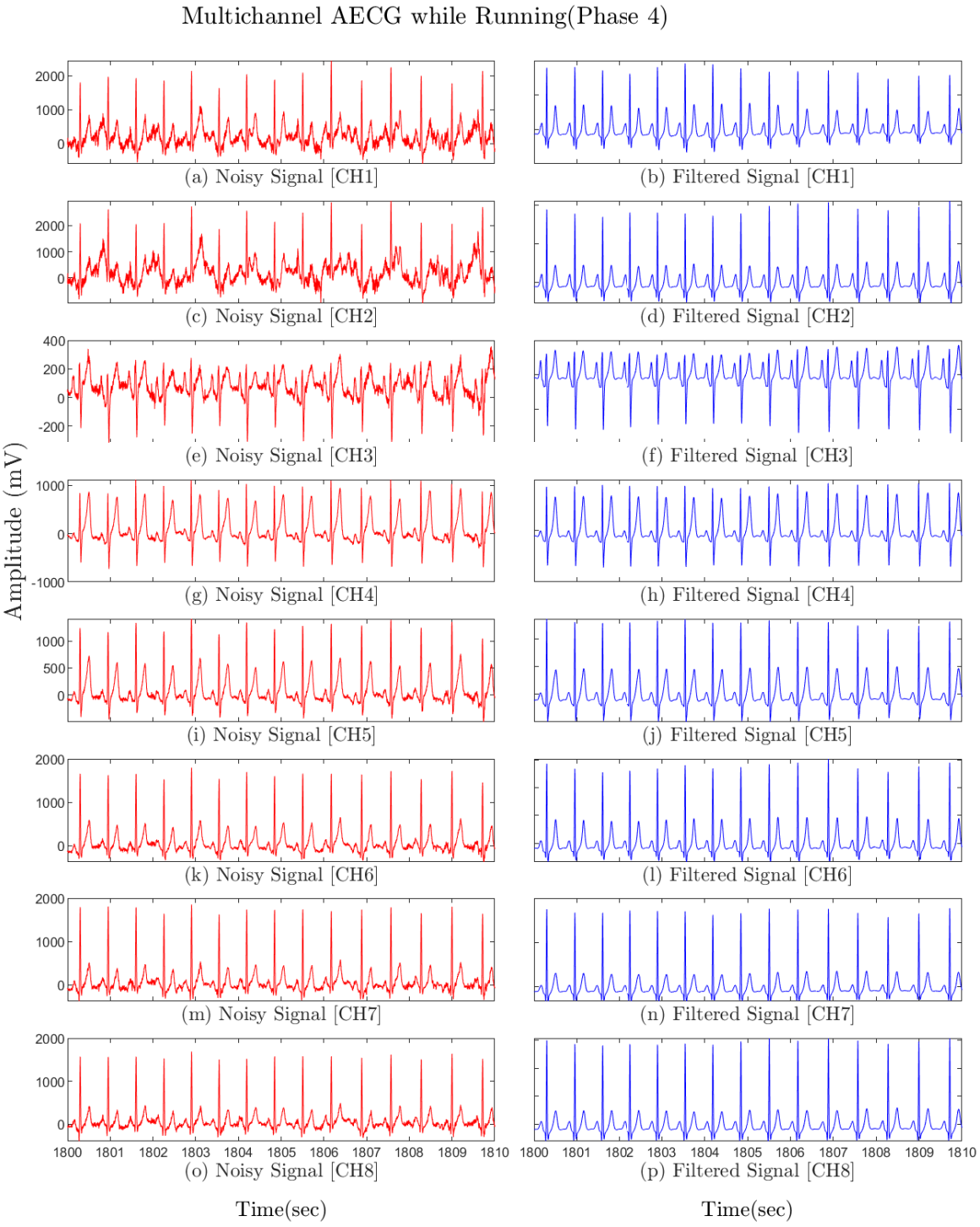


Fig. A.4 Typical results of multi-channel AECG recordings on phase four. Left column plots are the raw ECGs while right column plots are filtered signals

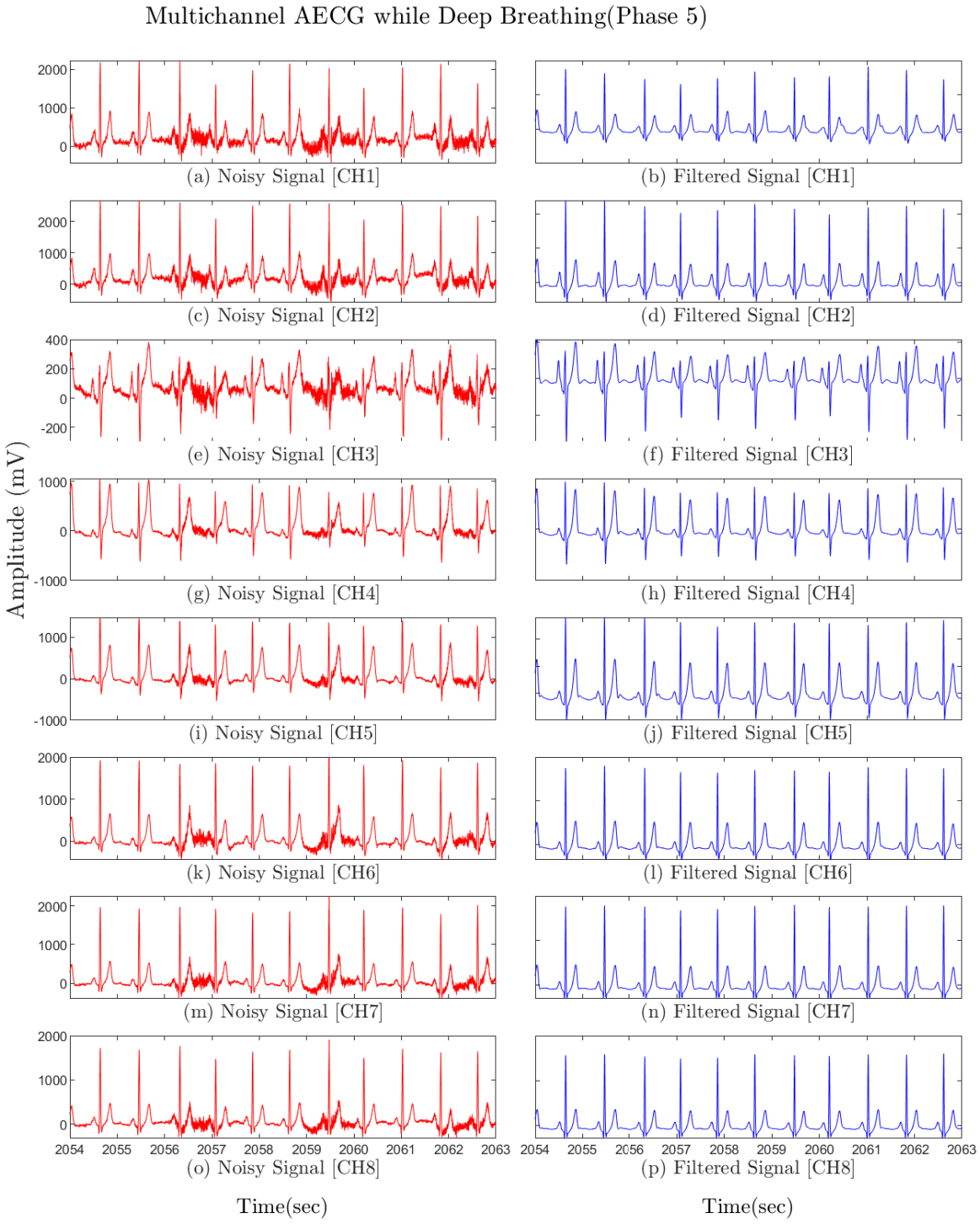


Fig. A.5 Typical results of multi-channel AECG recordings on phase five. Left column plots are the raw ECGs while right column plots are filtered signals

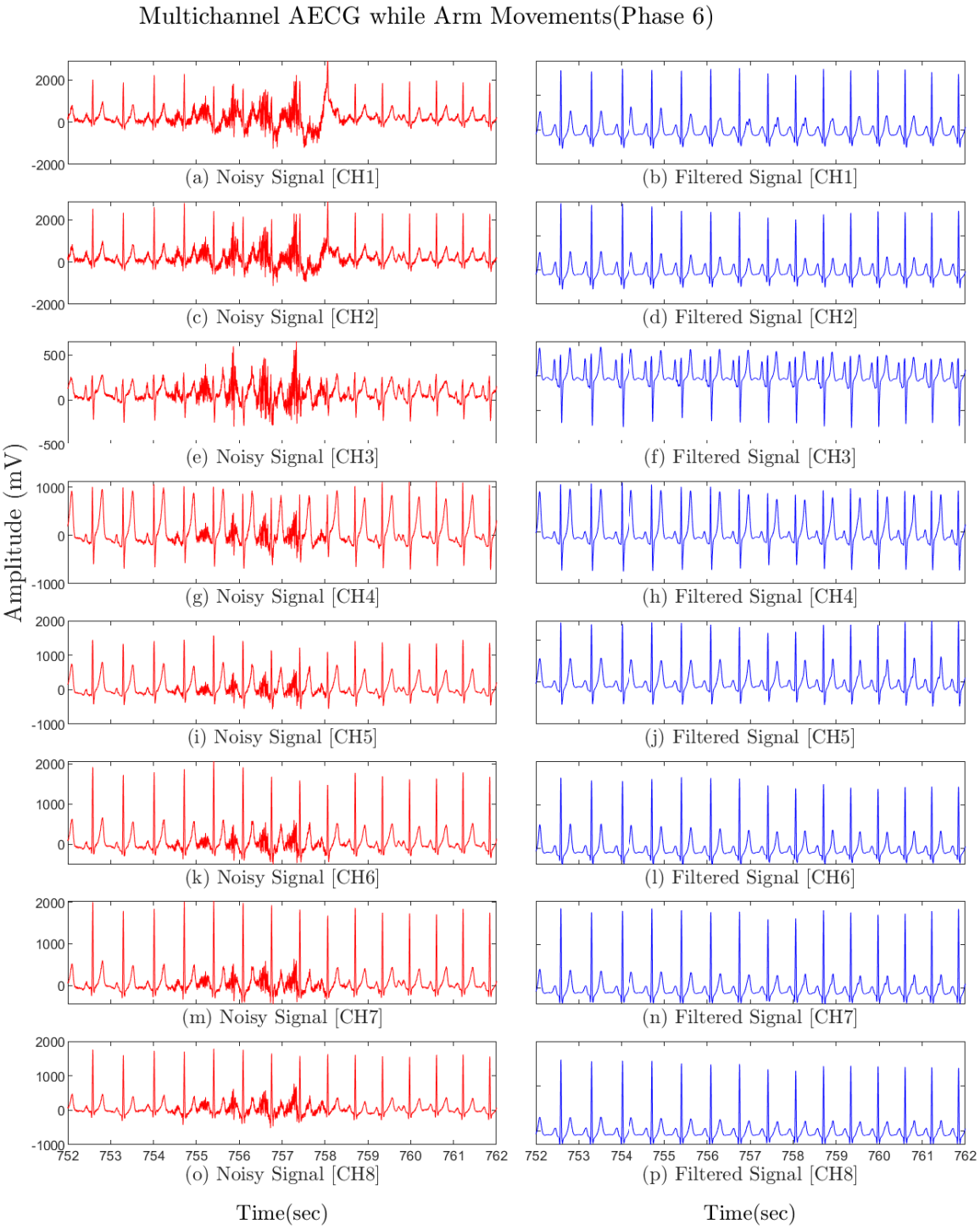


Fig. A.6 Typical results of multi-channel AECG recordings on phase six. Left column plots are the raw ECGs while right column plots are filtered signals

Appendix B

Experts opinions

I am a PhD student at the University of Hull and I have been working on the filtering of multichannel ambulatory ECG (AECG) signals contaminated with intermittent and high amplitude noise. The algorithm that has been developed is based on a combination of statistical and adaptive methods. These methods are Principal Component Analysis and Kalman filtering, which we called (PCAKF). A period of training data is required where information such as inter and infra-channel correlations in PQRST peak heights is collected. This information makes the PCAKF very powerful.

The method has been tested using different types of synthetic ECG and synthetic noises, such as band limited, additive white and pink Gaussian noise. An example is given below where white Gaussian noise, band limited to the spectrum spanned by ECG signals, has been added to a synthetic ECG signal with 10% inter-beat variation in parameters to yield an SNR of -2 dB. Where the signal is known the benefits of filtering can be quantatively measured using SNR and the error in clinically important parameters. Figure 1a illustrates the signal and signal plus noise, while Fig 1b shows the output of the filter. PCAKF yields a SNR improvement of almost 21 dB.

Using synthetic data, the PCAKF method has outperformed all gold standard filtering methods, including Extended Kalman Filters (EKF2 and EKF17).

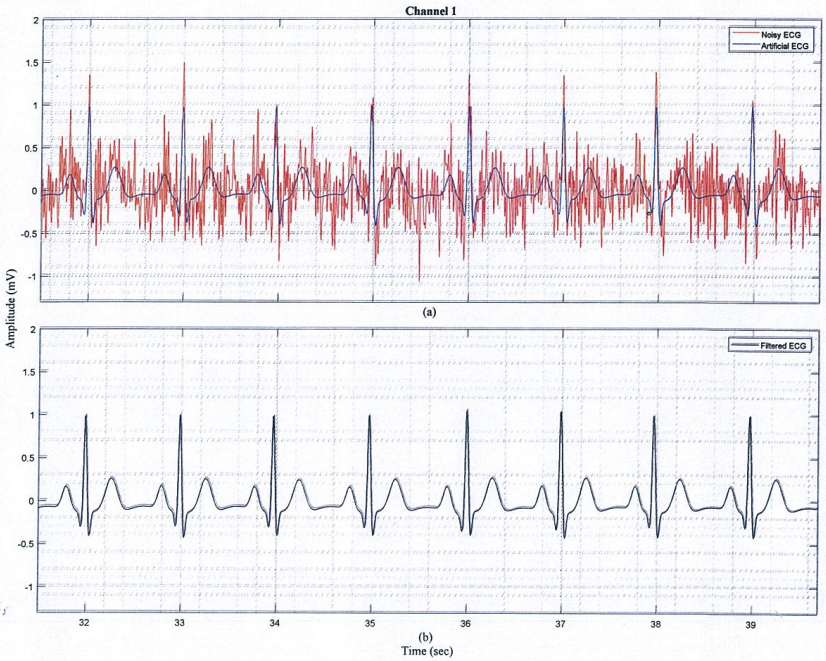
Where your help would be greatly appreciated is in assessing the performance of the method applied to AECG signals. In this case, the signal is not known and so SNR improvement cannot be calculated. The important measure is how well an expert thinks that filtering has revealed clinically important features of the signal. Six sets of results are presented below where participants undertake a range of activities including sitting, walking, jogging and running. Additionally, abnormal and AF ECG also used to test the PCAKF framework. I would be very grateful if you would comment on how well filtering performed.

I think PCAKF is a significant step forward and would be very keen to see it applied in clinical situations, or bound into future grant applications.

Dear Mehdi

The PCAKF system is powerful for deletion of intermittent and high amplitude signales. The PCAKF system clear the ECG and interpretation ECG by PCAKF is very comfortable

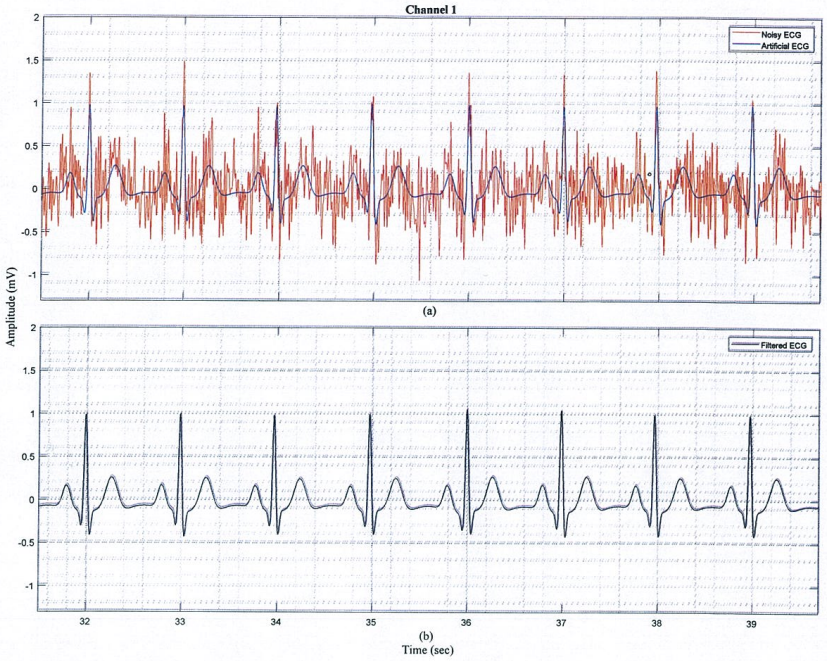
Example 1: Synthetic ECG with Additive White Noise Initial SNR = -2 dB.



Cardiologist opinion:

Has filtering added to the clinical value of the data?

Example 1: Synthetic ECG with Additive White Noise Initial SNR = -2 dB.



Cardiologist opinion:

Has filtering added to the clinical value of the data?

باسلام

تمام مثالهای ارائه شده از ECG در فازهای مختلف مشاهده شد.

فیلترینگ پیشنهادی بطور واضح توانسته است noise های موجود

در حرکت های مختلف حذف کند، که حذف noise ها کاربرد فراوانی

در هولترمونیتورینگ میتواند داشته باشد چه وجود noise ها

هنگام CPR که noise های ناشی از حرکت ماهیچه ها و

قفسه سینه مانع تشخیص واضح امواج P, QRS و T در ECG می شود.

و یا میتوان استفاده از الکترودکوتر هنگام جراحی که بدلیل تداخل الکتریکی

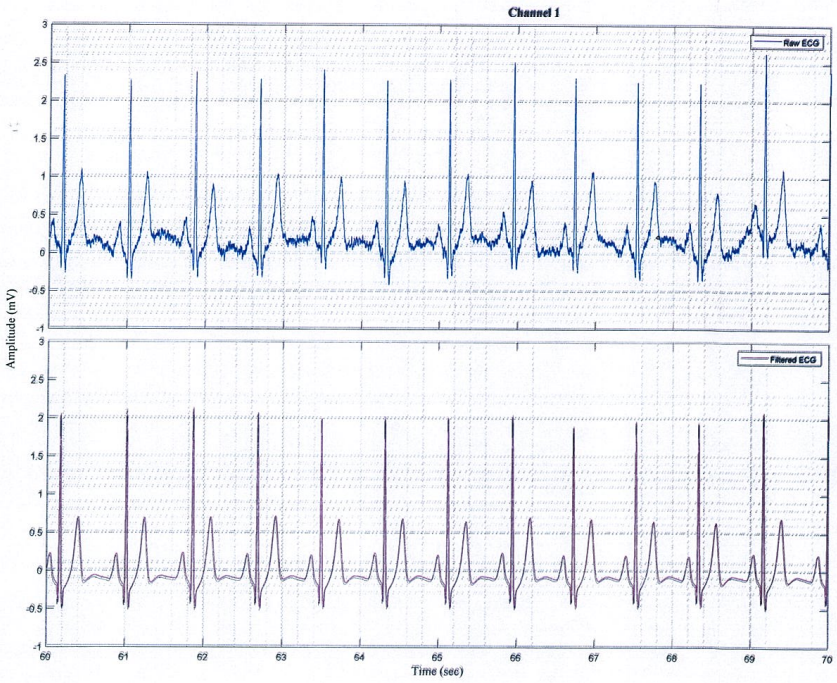
و آرتیفکت تشخیص ECG مشکل است اگر روش پیشنهادی میتواند

سیگنالهای قابل قبولی بدهد واقعاً حیاتی است.

مرکز آموزشی درمانی و تحقیقاتی امام رضا (ع) تبریز
دکتر صمد اسلام جمال گزازی
متخصص بیهوشی و مراقبتهای ویژه
استادیار تمام وقت جراحی ن پ ۹۳۲۵۸

2 /B

Activity 1: Participant sitting upright.



Student description:

Low initial noise but PCAKF provides some smoothing and makes visible some variation not immediately apparent in the unfiltered data and baseline drift removed.

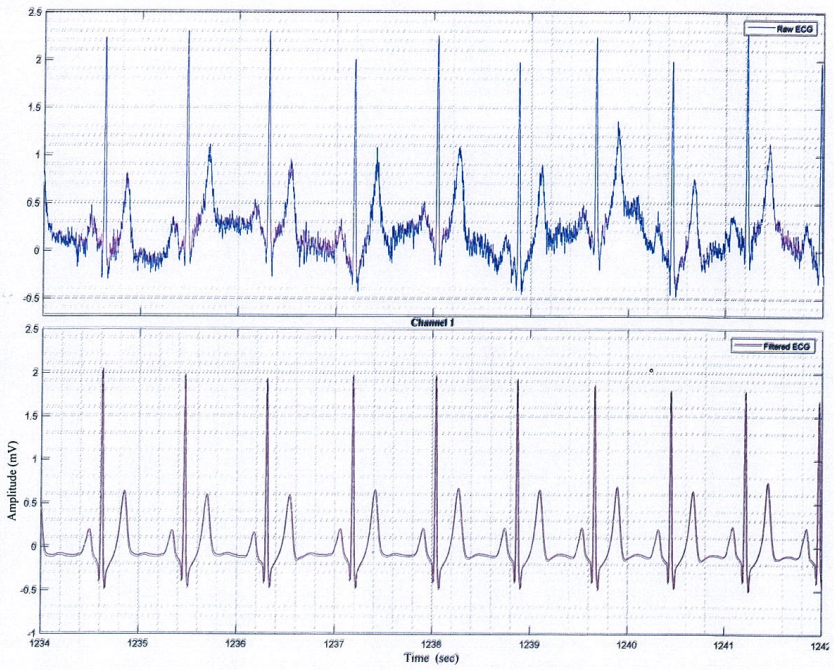
Cardiologist opinion:

Has filtering added to the clinical value of the data?

تمام موارد ارائه شده در هولتر مونیتورینگ رویت شده. فیلترینگ مذکور قادر است آرتیفکت و noise ها را حذف کند و در موارد اسپاینال آنستزیا (SA) یا جنرال آنستزیا (GA) بیمار ممکن است دچار لرز یا شیورینگ shivering شود که تشخیص امواج ECG بدلیل لرز بیمار مشکل است. یا در بیماران پارکینسونی که noise های زیادی در حین مونیتورینگ ECG دارند این noise ها شبیه noise های ناشی از ECG حین هولتر مونیتورینگ است که به دلیل وجود حرکت های متناوب ماهیچه ها ایجاد می شود. اگر روش فوق نتواند آرتیفکت های بوجود آمده را حذف کند کمک بزرگی در زمینهٔ مونیتورینگ ECG می کند.

مرکز آموزشی و درمانی طالقانی تبریز
دکتر عفت ترابی
متخصص بیهوشی و مراقبتهای ویژه
(دارای بورد تخصصی) ت.ب ۲۵۶۳۸

Activity 2: Participant Walking.



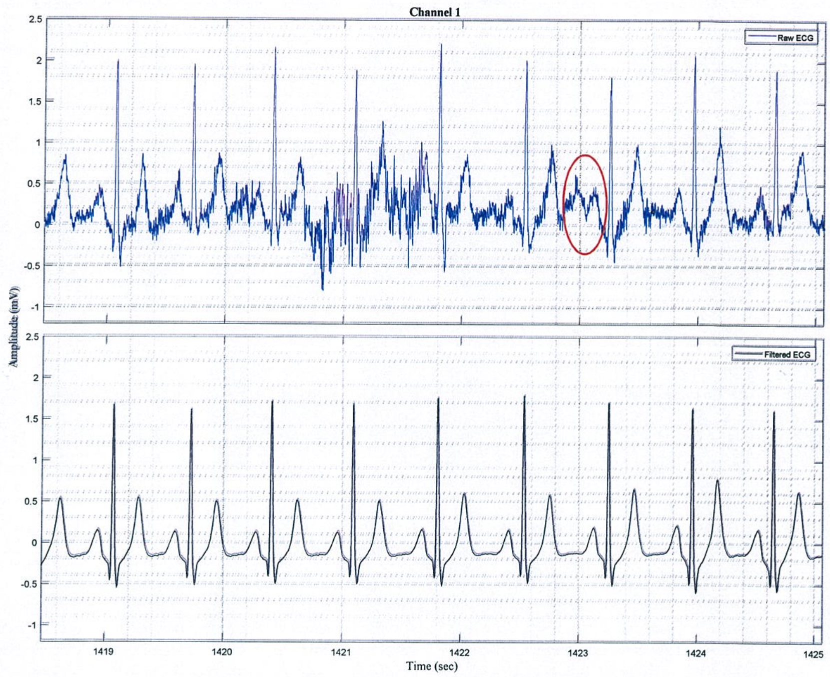
Student description:

Slightly larger noise. Inter-beat variation made clearer by PCAKF

Cardiologist opinion:

Has filtering added to the clinical value of the data?

Activity 3: Participant Jogging.



Student description:

Muscle noise and electrode movement artefacts distort the cardiac signal, especially the ECG components at around the times 1420 to 1422 or T wave at 1423 marked by red circle. PCAKF reconstructs plausible cardiac signals.

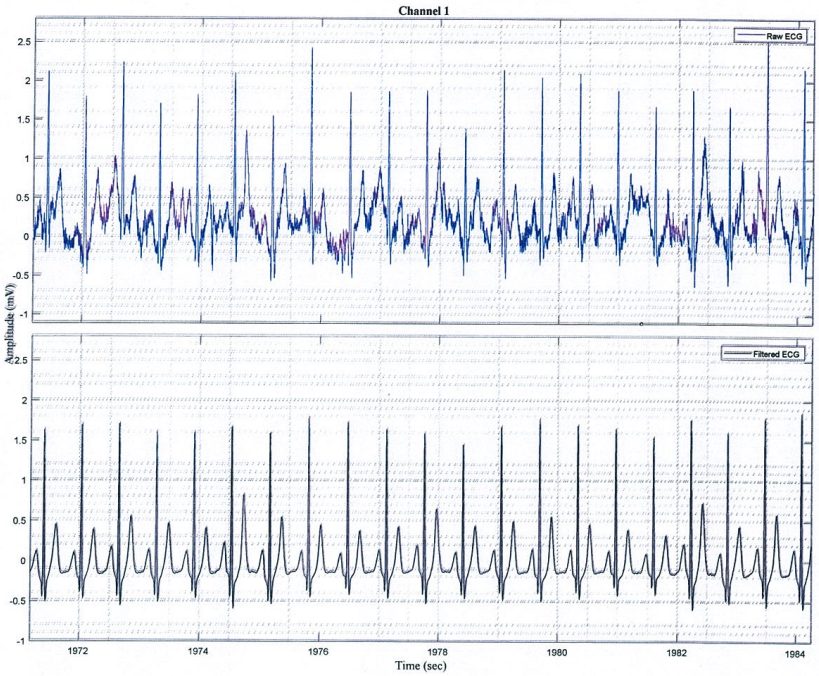
Cardiologist opinion:

Has filtering added to the clinical value of the data?

در اقدام سوم فیلترینگ باعث بهبود شد. امکان اجرای این فیلترینگ در
حین تست ورزشی که بیمار در حال حرکت و دویدن است بسیار کمک
کننده است زیرا افت قطعه ST حتی 1mm در rateهای بالا
جهت تشخیص مهم و حائز است که متاسفانه بدلیل آرتیفکت گاهاً
nondiagnostic می‌شود. برای تعیین ریتم زمینه‌ای هم
بسیار کمک کننده است

دکتر مهری محمدپور
متخصص قلب و عروق
بورد تخصصی از دانشگاه شهید بهشتی تهران
ن.پ ۹۰۲۶۰

Activity 4: Participant Running.



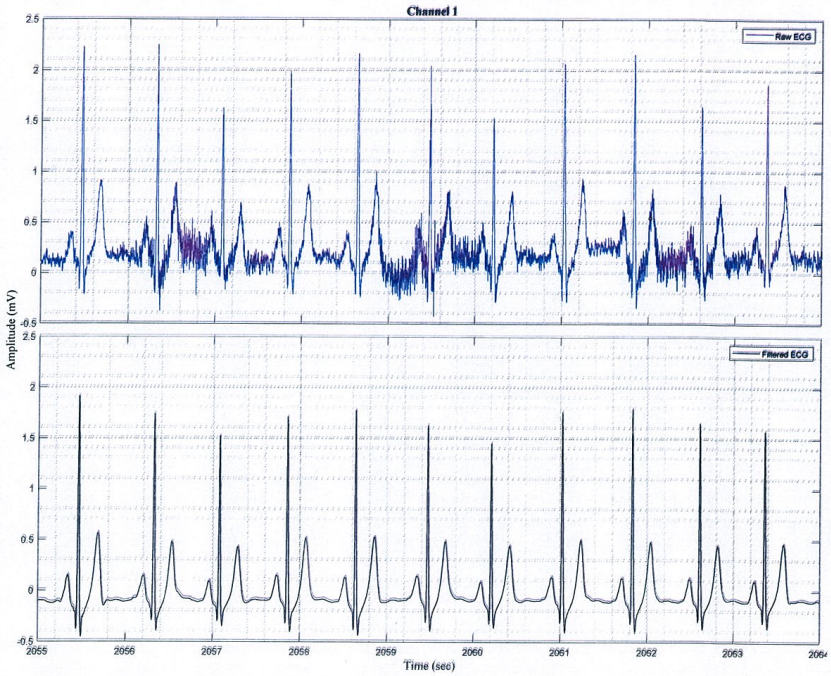
Student description:

Cardiac signal swamped by noise artefacts but PCAKF reconstructs a plausible signal.

Cardiologist opinion:

Has filtering added to the clinical value of the data?

Activity 5: Participant Deep Breathing.



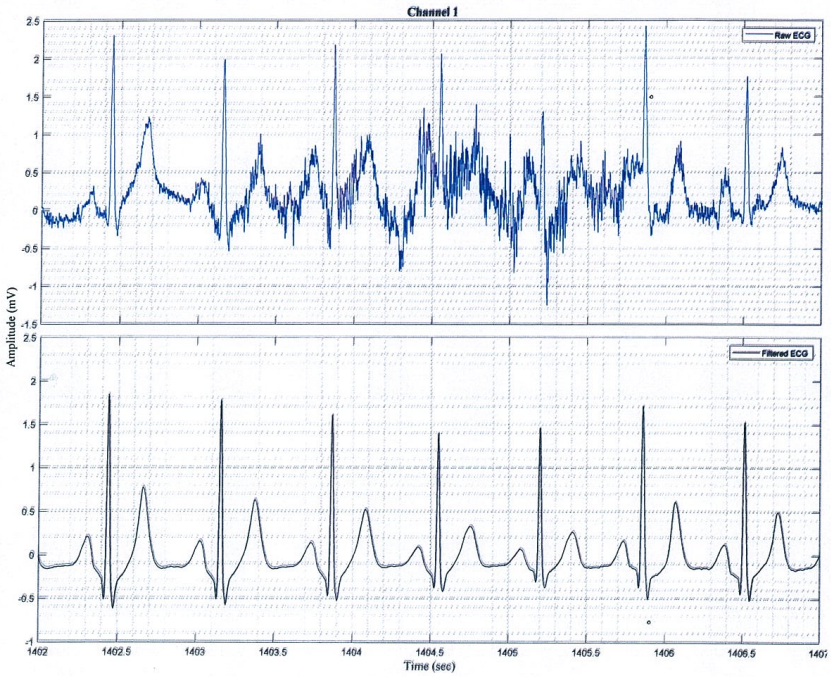
Student description:

The figure shows the noise caused by tree deep breathing and the proposed method could able to remove noise.

Cardiologist opinion:

Has filtering added to the clinical value of the data?

Activity 6: Participant Running and Jogging with Arm Movements.



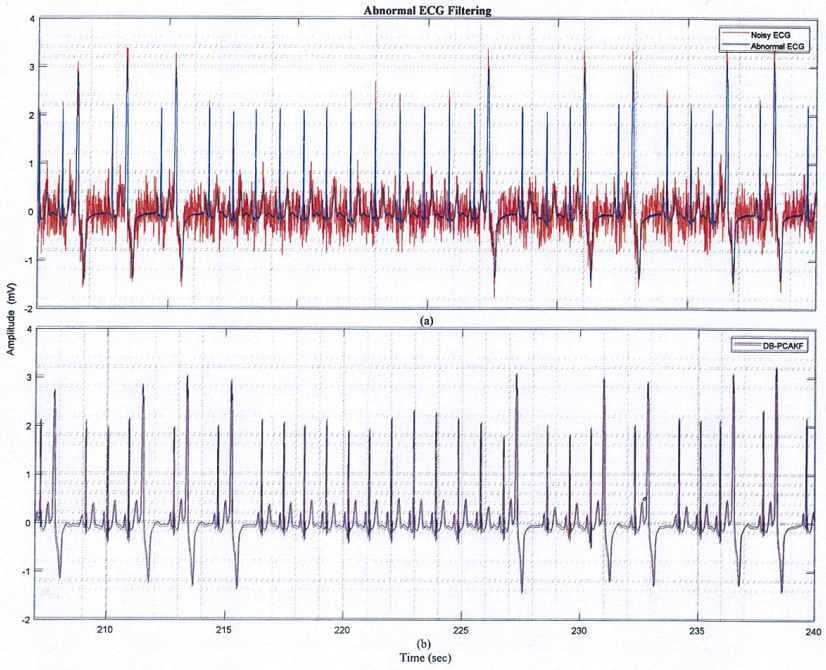
Student description:

Although R-peaks are identifiable, other features are lost in the noise, particularly in the middle portion.

Cardiologist opinion:

Has filtering added to the clinical value of the data?

Abnormal ECG contaminated with Additive White Noise Initial SNR = 5 dB .



Student description:

Abnormal ECG collected from &&& contaminated with white Gaussian noise and PCA-KF able to remove noise from the noisy signal.

Cardiologist opinion:

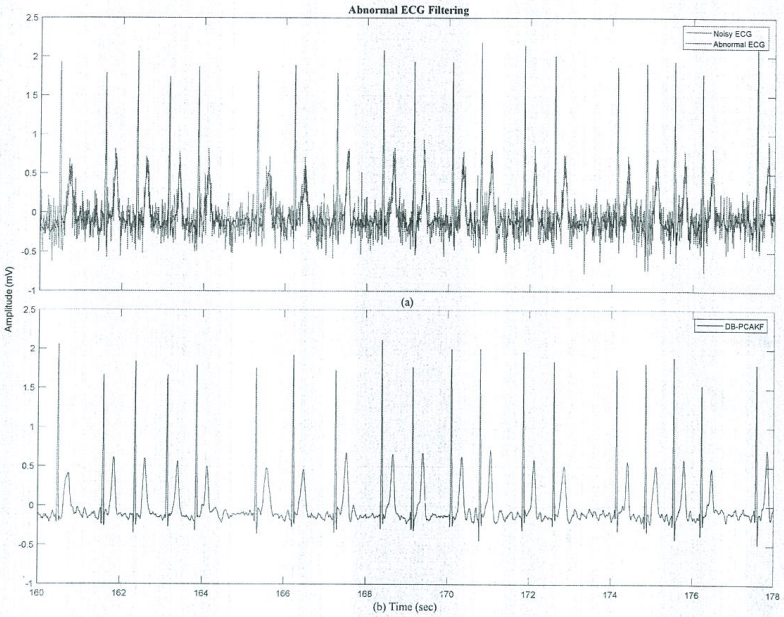
Has filtering added to the clinical value of the data?

حین بدست CPR بدلیل حرکت قفسه سینه خربان ، خوف بارامت

بیمار کمک کننده است

دکتر مهری محمد پور
متخصص قلب و عروق
بورد تخصصی از دانشگاه شهید بهشتی تهران
۹۰۲۶۰

Atrial Fibrillation ECG contaminated with Additive White Noise Initial SNR = 5 dB .



Student description:

The performance of the proposed method studied using atrial fibrillation data while added with white Gaussian noise at 5 dB. The PCAKF able to remove noise from the noisy signal. R-R intervals, QRS-T waves and f waves are visible at the filtered signal.

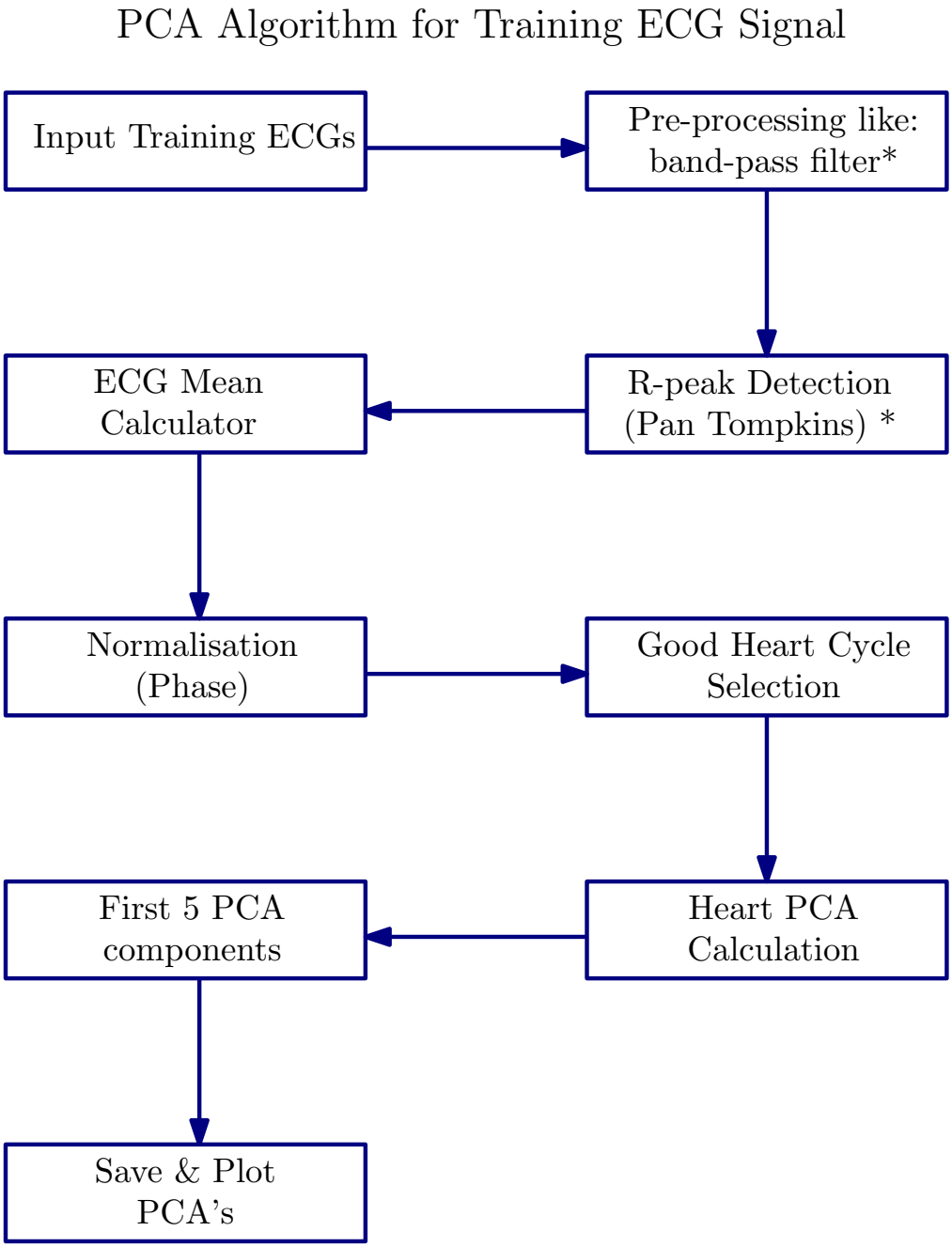
Cardiologist opinion:

Has filtering added to the clinical value of the data?

Appendix C

PCAKF Flowchart

This section provides flowcharts for PCA extraction and PCAKF algorithm.



$*DiMarco, L.Yet.al., 2012.Evaluationofanalgorithmbasedon single-conditiondecisionrulesforbinaryclassificationof12-lead ambulatoryECGrecordingquality.Physiologicalmeasurement.$

$*Pan, J, andTompkins, W.J.(1985).$
$Areal-timeQRSdetectionalgorithm.IEEETrans.Biomed.$

Fig. C.1 The PCA extraction Flowchart.

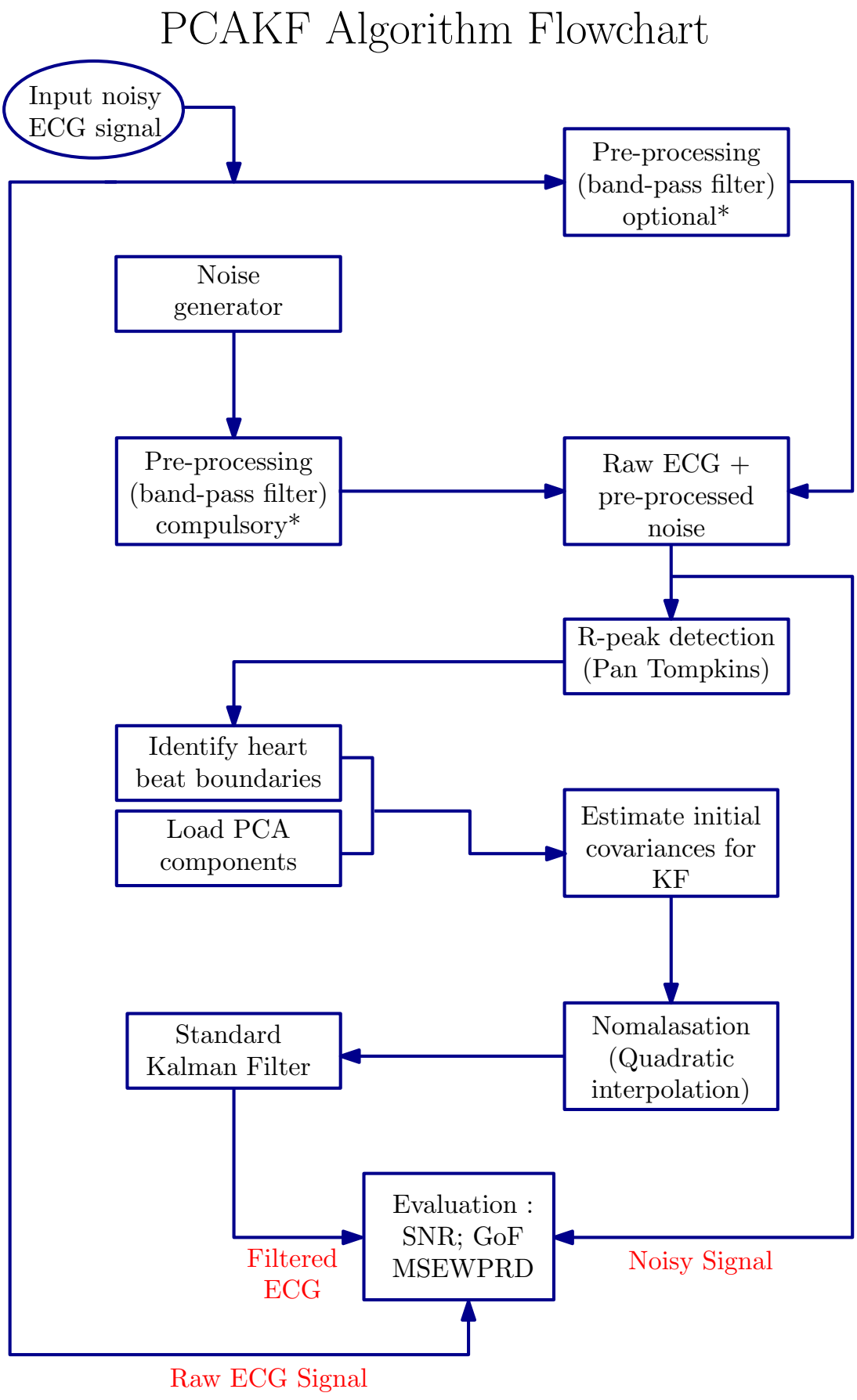


Fig. C.2 The PCAKF algorithm Flowchart.

Appendix D

Single Beat Single Channel PCAKF Algorithm Matlab Code

This section provides an example of Matlab code for the single beat single channel PCA calculation and PCAKF algorithm.

%% Single Beat Single Channel PCA Calculator

% % Contact email address: **mahdi.torabi84@gmail.com**

```
% %  ****************************
%   ** Set parameters used in filter   **
% %  ****************************

% Define the clean data as used for PCA calculation and for determining
% channel polarity

NumPCABeat = 10;                %This is number of Clean Beats which used for PCA & Polarity
NumSamplePerBeat = 1000;        %This is NOT fs of Data - Only used for PCA and R peaks
load ('E:ECG_Data')             % load ECG data

% % Set Number of channels to filter

NumChannel = size(ECG_Data,1);
NumSamples = size(ECG_Data,2);
FirstBeat = 1;
fs = 1000;                          %Frequency of the signal
t = (0:NumSamples-1)/fs;

% % ******************************
%    ** Do some pre-processing **
% % ******************************
% % The pre-processing MUST be same as PCAKF algorithm one

nMedianSamp = 400;              % The median is calculated over data spanning this number of samples
nMedianStep =  100;             % Number of samples between median calculations

% loop over the median intervals calulating the reference level
MedianCentres = 1+nMedianSamp/2 : nMedianStep : NumSamples-nMedianSamp/2;
nMedian = length( MedianCentres );
ReferenceLevel = zeros(nMedian,1);

% loop over channels
for iChan = 1:NumChannel
```

```
    for iMedian = 1:nMedian
        MedianIntervalData = ECG_Data( iChan, MedianCentres(iMedian) - nMedianSamp/2 :
MedianCentres(iMedian) + nMedianSamp/2 );
        ReferenceLevel(iMedian) = median( MedianIntervalData ); %(Index) );
    end

    % Calculate reference level by interpolation
    Reference = interp1(MedianCentres , ReferenceLevel , (1:NumSamples) ,'linear','extrap');
    % subtract reference from ECG signal
    ECG_Data(iChan,:) = ECG_Data(iChan,:) - Reference;

end

% %  **********************
%   ** Peak Detection    **
% %  **********************

%% You can use any other R peak detection method

[qrs_amp_raw,qrs_i_raw,delay] = pan_tompkin(ECG_Data(1,:),fs,1);
IndexPeaks = qrs_i_raw;                                  %  Peak Index from Peak detection
NumPeak = numel(IndexPeaks);

%   ************************************
%   **  Identify beat boundaries      **
%   ************************************

% determine the start and end of each heart cycle
% NOTE: the first and last partial heart cycles are not used.
HeartCycleBoundaries = round( ( IndexPeaks(1:end-1) + IndexPeaks(2:end) )/2 );
NumHeartCycle = numel( HeartCycleBoundaries ) - 1;

%   ************************************
%   **  Identify PCA Basis            **
%   ************************************

NumPCABasis = 7;
% [PCA_Basis,Mean] =
Heart_PCA(ECG_Data_New,HeartCycleBoundaries,NumPCABeat,FirstBeat,NumSamplePerBeat,NumP
CABasis);
PCA_Basis =
Heart_PCA(ECG_Data,HeartCycleBoundaries,NumPCABeat,FirstBeat,NumSamplePerBeat,NumPCABa
sis); % This is PCA Calculation

save('PCA_Cutted_Data.mat','PCA_Basis');
```

%% Heart PCA Function

```
% %  ******************
%   ** Uncentered PCA **
% %  ******************

% This code can use for both single and multi-channel analysis

function PCA_Basis =
Heart_PCA(ECG_Data,HeartCycleBoundaries,NumPCABeat,FirstBeat,NumSamplePerBeat,NumPCABa
sis)

NumChannel = size(ECG_Data,1);

 % define data structure to hold data required to calculate the
% covariance of time-resampled beat signal
PlotFlag = true;

% declare array to hold heart beat samples
X = zeros(NumPCABeat,NumSamplePerBeat,NumChannel);

% loop over beats resampling to normalised time
NormalisedSampleTime = (1:NumSamplePerBeat)/NumSamplePerBeat;

% This is the time 0->1 used to resample beats of different duration

SumNormBeat = zeros(NumSamplePerBeat,NumChannel);    % vector used to calculate mean beat

% This is the number of good beats to initiate the good beat statistics
NumInitiateBeat = 3;

 % loop over two beats to estimate the interbeat variation
figure(12);
hold on
for iHeartCycle = FirstBeat:FirstBeat+NumInitiateBeat-1

   % Extract data for one heart cycle
   StartHeartCycle = HeartCycleBoundaries(iHeartCycle  );
   EndHeartCycle   = HeartCycleBoundaries(iHeartCycle+1);
   HeartCycle_n    = EndHeartCycle-StartHeartCycle;
   NormalsedTimeInBeat = (0:HeartCycle_n)/HeartCycle_n;

   for iChan = 1:NumChannel
      HeartCycleData  = ECG_Data(iChan,StartHeartCycle:EndHeartCycle);
      NormalisedBeat = interp1(NormalsedTimeInBeat,HeartCycleData,NormalisedSampleTime);
      % accumulate data to calculate covariance
      X(iHeartCycle,:,iChan) = NormalisedBeat;
      % save data to calculate mean beat
      SumNormBeat(:,iChan) = SumNormBeat(:,iChan) + NormalisedBeat';
```

```
        if iChan==1
            plot(NormalisedSampleTime,NormalisedBeat,'b')
        end

    end

end

% Calculate typical interbeat variance
NumHeartCyclesUsed = NumInitiateBeat;
BeatVariance = zeros(NumInitiateBeat,NumChannel);          % beat variances
for iChan = 1:NumChannel

    MeanBeat = SumNormBeat(:,iChan)/NumHeartCyclesUsed;

    for iHeartCycle = 1:NumInitiateBeat
        BeatVariance(iHeartCycle,iChan) = sum( (X(iHeartCycle,:,iChan)-MeanBeat').^2 );
    end

    if iChan==1
        plot(NormalisedSampleTime,MeanBeat,'k','LineWidth',2)
    end

end
MeanVariance = mean( BeatVariance )';

% loop over beats resampling to normalised time
ThisBeatVariance = zeros(NumChannel,1);
for iHeartCycle = FirstBeat+NumInitiateBeat:FirstBeat+NumPCABeat

    % Extract data for one heart cycle
    StartHeartCycle = HeartCycleBoundaries(iHeartCycle  );
    EndHeartCycle   = HeartCycleBoundaries(iHeartCycle+1);
    HeartCycle_n    = EndHeartCycle-StartHeartCycle;
    NormalsedTimeInBeat = (0:HeartCycle_n)/HeartCycle_n;

    for iChan = 1:NumChannel
        HeartCycleData  = ECG_Data(iChan,StartHeartCycle:EndHeartCycle);
        NormalisedBeat = interp1(NormalsedTimeInBeat,HeartCycleData,NormalisedSampleTime);
        % accumulate data to calculate covariance
        X(NumHeartCyclesUsed+1,:,iChan) = NormalisedBeat;
        MeanBeat = SumNormBeat(:,iChan)/NumHeartCyclesUsed;
        ThisBeatVariance(iChan) = sum( (NormalisedBeat-MeanBeat').^2 );
    end

    % check this beat is OK and keep if it is
    if max( ThisBeatVariance./MeanVariance ) < 4^2       %By changing the power we can change the
senstivity of the bad and good signals
        NumHeartCyclesUsed = NumHeartCyclesUsed+1;
        SumNormBeat = SumNormBeat + squeeze(X(NumHeartCyclesUsed,:,:));
        plot(NormalisedSampleTime,X(NumHeartCyclesUsed,:,1),'b')
```

```
    else
      plot(NormalisedSampleTime,X(NumHeartCyclesUsed+1,:,1),'r')
      title ('Heart Cycles which removing heart beats which contaminated highly with noise')
    end

end

 % % NOW Do PCA

    PCA_Basis = zeros(NumSamplePerBeat,NumPCABasis);

    Coeff = pca(X(1:NumHeartCyclesUsed,:),'NumComponents',NumPCABasis,'Centered',false);
%     [Coeff,score,latent,tsquared,explained] =
pca(X(1:NumHeartCyclesUsed,:),'NumComponents',NumPCABasis,'Centered',false);
       PCA_Basis(:,:) = Coeff(:,1:NumPCABasis);
```

%% Single Beat Single Channel PCAKF Algorithm

```
% % ****************************
%   ** Set parameters used in filter   **
% % ****************************

NumSamplePerBeat = 1000;

% % ********************
%   ** Loading ECG Data **
% % ********************

load ('ECG_Data')

% % Set Number of channels to filter
NumChannel = size(ECG_Data,1);
NumSamples = size(ECG_Data,2);
FirstBeat = 1;
fs = 1000;
t = (0:NumSamples-1)/fs;
ECG_Data_New = ECG_Data(1,:);

% % ************************
%   ** Peak Detection    **
% % ************************
% same R peak detection of PCA calculation must be used

[qrs_amp_raw,qrs_i_raw,delay] = pan_tompkin(ECG_Data(1,:),fs,1);        % peak detection using
Pam_Tompkin
IndexPeaks = qrs_i_raw;                                  %  Peak Index from Peak detection
NumPeak = numel(IndexPeaks);

%% Start Loop

NumLoop = 1;
SNR_Noisy_dB = zeros(NumLoop,1);
for i_loop = 1:NumLoop


% % ************************
% % %Adding Noise to Data %
% % ************************

ECG_Data_Noise = ECG_Data(1,:);

for iChan=1,NumChannel;
   Index = find( abs(t-500)<100 );
   sd = 0.2;
   Len = length(Index);
   fs = 512;
```

```
   beta = 1 ;
   Noise = ColoredNoise(sd,Len,fs,beta);
   ECG_Data_Noise(iChan,Index) = (ECG_Data_New(iChan,Index)+ Noise);
end

% % ************************
%    ** Do some pre-processing **
% % ************************
% % The pre-processing should be same as PCA calculation one

nMedianSamp = 400;                % the median is calculated over data spanning this number of samples
nMedianStep =  100;               % number of samples between median calculations

% loop over the median intervals calulating the reference level
MedianCentres = 1+nMedianSamp/2 : nMedianStep : NumSamples-nMedianSamp/2;
nMedian = length( MedianCentres );
ReferenceLevel = zeros(nMedian,1);

% loop over channels
for iChan = 1:NumChannel

   for iMedian = 1:nMedian
      MedianIntervalData = ECG_Data_Noise( iChan, MedianCentres(iMedian) - nMedianSamp/2 :
MedianCentres(iMedian) + nMedianSamp/2 );
%       Index = find( MedianIntervalData>Bot_of_Ref_Range &
MedianIntervalData<Top_of_Ref_Range );
      ReferenceLevel(iMedian) = median( MedianIntervalData ); %(Index) );
   end

   % Calculate reference level by interpolation
   Reference = interp1(MedianCentres , ReferenceLevel , (1:NumSamples) ,'linear','extrap');
   % subtract reference from ECG signal
   ECG_Data_Noise(iChan,:) = ECG_Data_Noise(iChan,:) - Reference;

end

% Number of Peaks come from Peak Detection

NumPeak = numel(IndexPeaks);

%   ************************************
%   **  Identify beat boundaries      **
%   ************************************
% determine the start and end of each heart cycle
% NOTE: the first and last partial heart cycles are not used.
HeartCycleBoundaries = round( interp1(1:NumPeak,IndexPeaks,(1:NumPeak-1)+0.5,'spline') );
NumHeartCycle = numel( HeartCycleBoundaries ) - 1;

% %  **************
%    **  Input PCA   **
% %  **************
```

```
load('PCA_Cutted_Data.mat','PCA_Basis')
NumPCABasis = 5; % Different PCA numbers can be used

%   ************************************
%   **  Estimate Initial Covariances  **
%   **  For the Kalman Filter         **
%   ************************************
 nTrainingBeats = 20;
% Initialise storage for best fit Gaussian models for training beats
X_Training_Fit = zeros(nTrainingBeats,NumPCABasis);
% Initialise first estimate Gaussian Parameters
X = zeros(NumPCABasis,1); X(1) = 1;

% Loop over training beats fitting PCA models
for iHeartCycle = 1:nTrainingBeats

   % Extract data for one heart cycle
   StartHeartCycle = HeartCycleBoundaries(iHeartCycle  );
   EndHeartCycle   = HeartCycleBoundaries(iHeartCycle+1);

   HeartCycle_n    = EndHeartCycle-StartHeartCycle+1;
   HeartCyclePhase = linspace(0,2*pi,HeartCycle_n);
   HeartCycleData  = ECG_Data_Noise(StartHeartCycle:EndHeartCycle);


   % fit PCA model
   [X1, R] = LSQ_Fit_PCA_Model(HeartCyclePhase,HeartCycleData,PCA_Basis);
   X_Training_Fit(iHeartCycle,:) = X1;
end
% initial mean Gaussian Parameters
X_bar = mean( X_Training_Fit , 1 )';
% initial Gaussian Parameter covariance matrix
P = cov( X_Training_Fit );
% initial Gaussian Parameter time-step transition covariance matrix
Q = cov( X_Training_Fit(2:end,:) - X_Training_Fit(1:end-1,:));

%   ************************************
%   **  Loop over each heart cycle    **
%   ************************************

% declare storage to save filtered parameters
X            = zeros(NumPCABasis,NumHeartCycle);
ECG_Filtered = zeros(size(ECG_Data_Noise));

for iHeartCycle = 1:NumHeartCycle

   % Extract data from one heart cycle
   StartHeartCycle = HeartCycleBoundaries(iHeartCycle  );
   EndHeartCycle   = HeartCycleBoundaries(iHeartCycle+1);
```

```
    HeartCycle_n    = EndHeartCycle-StartHeartCycle+1;
    HeartCyclePhase = linspace(0,2*pi,HeartCycle_n);
    HeartCycleData  = ECG_Data_Noise(StartHeartCycle:EndHeartCycle);
    HeartCycleTime  = t(StartHeartCycle:EndHeartCycle);

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    % Kalman Beat Filter Heart Cycle  %
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    % Predict next beat parameters
    Pred_X_bar = X_bar;
    Pred_P = P + Q;

    % Calculate measured parameters by fitting Gaussians to Measurements.
    % The fitted parameters are Z_Measured and the uncertainty covariance
    % is R.
    [Z_measured, R] = LSQ_Fit_PCA_Model(HeartCyclePhase,HeartCycleData,PCA_Basis);
    % Calculate Kalman correction factor based on relative uncertainties in
    % prediction and measurement
    K = Pred_P / ( Pred_P + R );
    % Next add Kalman correction based on measurement
    X_bar = Pred_X_bar + K*( Z_measured - Pred_X_bar  );
    % Update uncertainty in Gaussian parameter vector
    P = (eye(NumPCABasis)-K)*Pred_P*(eye(NumPCABasis)-K)' + K*R*K';   % Josepth Form

    % Save the filtered parameters and beat
    X(:,iHeartCycle) = X_bar;
    ECG_Filtered(StartHeartCycle:EndHeartCycle) =
interp1(linspace(0,2*pi,NumSamplePerBeat),Heart_PCA_Model(PCA_Basis,X_bar),HeartCyclePhase);
end

%%%%%%%%%%%%%
%Calculate SNR%
%%%%%%%%%%%%%

SNR_Noisy_dB(i_loop) = SNR_Calculation(t,ECG_Data_New,ECG_Data_Noise,ECG_Filtered,Index);
% Any other SNR calculation can be used.
end

Mean_SNR_over100Loop = mean(SNR_Noisy_dB)

figure;
plot(t,ECG_Data_Noise/10,'b')
hold on
plot(t,ECG_Filtered/10,'k')
% title('Whiet Gaussian noise')
grid on
legend('Noisy','Single-Channel-PCAKF')
xlabel('Time(s)')
ylabel('Amplitude(mV)')
```