Adaptive Parameter Estimation, Modeling and Patient-Specific Classification of Electrocardiogram Signals

by

Shwetha Reddy Edla

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2012 by the
Graduate Supervisory Committee:

Antonia Papandreou-Suppappola, Chair
Chaitali Chakrabarti
Narayan Kovvali
Cihan Tepedelenlioğlu

ARIZONA STATE UNIVERSITY

December 2012

ABSTRACT

Adaptive processing and classification of electrocardiogram (ECG) signals are important in eliminating the strenuous process of manually annotating ECG recordings for clinical use. Such algorithms require robust models whose parameters can adequately describe the ECG signals. Although different dynamic statistical models describing ECG signals currently exist, they depend considerably on *a priori* information and user-specified model parameters. Also, ECG beat morphologies, which vary greatly across patients and disease states, cannot be uniquely characterized by a single model.

In this work, sequential Bayesian based methods are used to appropriately model and adaptively select the corresponding model parameters of ECG signals. An adaptive framework based on a sequential Bayesian tracking method is proposed to adaptively select the cardiac parameters that minimize the estimation error, thus precluding the need for pre-processing. Simulations using real ECG data from the online Physionet database demonstrate the improvement in performance of the proposed algorithm in accurately estimating critical heart disease parameters. In addition, two new approaches to ECG modeling are presented using the interacting multiple model and the sequential Markov chain Monte Carlo technique with adaptive model selection. Both these methods can adaptively choose between different models for various ECG beat morphologies without requiring prior ECG information, as demonstrated by using real ECG signals.

A supervised Bayesian maximum-likelihood (ML) based classifier uses the estimated model parameters to classify different types of cardiac arrhythmias. However, the non-availability of sufficient amounts of representative training data and the large inter-patient variability pose a challenge to the existing supervised learning algorithms, resulting in a poor classification performance. In addition,

recently developed unsupervised learning methods require *a priori* knowledge on the number of diseases to cluster the ECG data, which often evolves over time. In order to address these issues, an adaptive learning ECG classification method that uses Dirichlet process Gaussian mixture models is proposed. This approach does not place any restriction on the number of disease classes, nor does it require any training data. This algorithm is adapted to be patient-specific by labeling or identifying the generated mixtures using the Bayesian ML method, assuming the availability of labeled training data.

To *Papa*, *Mummy* and my brother *Sunny*.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the guidance and backing of several people who in one way or another contributed and extended their valuable assistance to me throughout the course of my graduate career. The journey has been long and bumpy at times, but it was all made worthwhile with the help and encouragement that I have received.

First and foremost, I would like to thank my advisor and committee chair, Dr. Antonia Papandreou-Suppapola for her advice and unfailing support. Her encouragement, guidance and kindness is something I will always be grateful for. My long association with her throughout the course of my graduate studies has shaped me into what I am today. She has always supported me to the fullest, encouraged me in every step of the way, and instilled me with confidence by recognizing and trusting my abilities.

I sincerely acknowledge Dr. Narayan Kovvali for his informative ideas and discussions. His patience in answering all my questions and accessibility have been a great benefit throughout this project.

I would like to thank Dr. Chaitali Chakrabarti for being a part of my committee from the beginning of my PhD, and for her insightful and useful input from time to time. I would also like to acknowledge Dr. Cihan Tepedelenlioğlu for graciously accepting to be part of my dissertation defense committee.

I would also like to thank my family for their encouragement, love and patience throughout my career; my father for infusing the value of education in me and for always having words of encouragement and support in whatever I chose to pursue, my mother for all she did for me, from teaching me my first alphabet to pushing me to always do my best, and my little brother for inspiring me through his own achievements. I simply wouldn't have been here without them.

Last but not the least, I would like to thank Venkat for his constant support. When the chips were down, his words of advice and backing helped the most. This dissertation would not have been possible, both literally and figuratively, without him.

TABLE OF CONTENTS

Page

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER

1 INTRODUCTION 1

1.1 ECG Signal Modeling 1

1.2 ECG Modeling Contributions 4

1.3 ECG Classification 5

1.4 ECG Classification and Clustering Contributions 11

1.5 Organization 12

2 BAYESIAN APPROACH TO PARAMETER ESTIMATION 15

2.1 General Bayesian state-space framework for parameter estimation 15

2.2 General Bayesian state-space framework for parameter estimation using multiple models 18

2.3 Application of Bayesian approach to ECG signal modeling 20

3 ADAPTIVE PARAMETER ESTIMATION USING MULTI-HARMONIC ECG MODEL 21

3.1 State-space Model using Multi-harmonic Components 21

3.2 Framework for Adaptive Signal Parameter Estimation 22

3.3 Simulations and Discussion 25

4 ECG MODELING USING INTERACTING MULTIPLE MODELS . . 29

4.1 IMM Algorithm 29

4.2 State-space Model using IMMs 31

4.3 Framework for ECG Modeling with IMM 32

4.4 Simulations and Discussion 34

5 ECG MODELING USING SEQUENTIAL MARKOV CHAIN MONTE CARLO METHOD WITH SIMULTANEOUS MODEL SELECTION . 38

CHAPTER Page
5.1 SMCMC Filtering and Simultaneous Model Selection 39
5.2 State-space Model for the SMCMC Filter with Simultaneous Model Selection 41
5.3 Framework for ECG Modeling with SMCMC Filtering and Simultaneous Model Selection 43
5.4 Simulations and Discussion 45
6 ECG ARRHYTHMIA CLASSIFICATION USING BAYESIAN MAXIMUM LIKELIHOOD CLASSIFIER 56
6.1 Electrical Activity of the Heart 57
6.2 Bayes Maximum-Likelihood Classification Method 58
6.3 Classification with the Multi-harmonic ECG Model 59
6.3.1 Classification Results 60
6.4 Classification with the IMM-KF ECG Model 60
6.4.1 Classification Results 62
6.5 Classification with the SMCMC ECG Model 62
6.5.1 Classification Results 62
7 PATIENT-SPECIFIC ECG ARRHYTHMIA CLUSTERING USING BAYESIAN NON-PARAMETRIC METHODS 64
7.1 Dirichlet Process Mixture Modeling 65
7.1.1 DP Model Formulation 66
7.1.2 Estimation of DP Model Parameters using Blocked Gibbs Sampling 70
7.2 Framework for ECG Beat Clustering using the DP GMM 73
7.2.1 Feature Design for ECG Beat Clustering 75
7.3 Cluster Labeling using Bayes ML Method 78
7.4 Simulations and Discussion 79
8 CONCLUSIONS AND FUTURE WORK 91

CHAPTER Page
8.1 Conclusions 91
8.2 Future work 95
REFERENCES 98

LIST OF TABLES

Table		Page
1.1	List of acronyms.	13
1.2	List of MIT-BIH arrhythmia notation.	14
5.1	Comparison of average RMSE (x 10^{-2}).	55
6.1	Confusion matrix showing classification results using the multi-harmonic model with adaptive parameter estimation.	60
6.2	Confusion matrix showing classification results using the IMM-KF ECG model.	62
6.3	Confusion matrix showing classification results using the SMCMC ECG model.	63
6.4	Comparison of classification results.	63
7.1	Confusion matrix showing clustering results using the DP algorithm.	81
7.2	Confusion matrix showing cluster labeling results using the Bayes ML method.	82
7.3	Grouping of MIT-BIH arrhythmia database beat types into AAMI recommended beat types.	85
7.4	Confusion matrix showing clustering results for the DP algorithm using AAMI recommended practice.	86
7.5	Confusion matrix showing cluster labeling results with the Bayes ML method using AAMI recommended practice.	86
7.6	Comparison of clustering results.	89

LIST OF FIGURES

Figure Page

3.1 Surface plot of the MSE for the entire chosen range of the adaptive parameters for a segment from a normal sinus rhythm ECG signal. The arrow indicates the minimum MSE point for one adaptive parameter pair. 24

3.2 MSE of the multi-harmonic ECG modeling algorithm with fixed and adaptive model parameters, and variation in the optimum number of harmonics for different types of ECG signals. 28

4.1 Original and reconstructed ECG signals using the IMM-KF algorithm for different beat types. The letters in the square boxes are the beat labels. 35

4.2 Original and reconstructed ECG signals using the IMM-KF algorithm for different beat types. The letters in the square boxes are the beat labels. 36

5.1 (a) Synthetically generated noisy signal from a quadratic model and the reconstructed signal with 500 data samples; (b) Model probabilities evaluated using equal prior model probabilities (top) and unequal prior model probabilities (bottom) using SMCMC algorithm with 500 data samples. 48

5.2 (a) Synthetically generated noisy signal from a quadratic model and the reconstructed signal with 2000 data samples; (b) Model probabilities evaluated using equal prior model probabilities (top) and unequal prior model probabilities (bottom) using SMCMC algorithm with 2000 data samples. 49

5.3 Original and reconstructed ECG signals from the SMCMC filtering algorithm with simultaneous model selection for different beat types. The letters in the square boxes are the beat labels. 50

Figure Page

5.4 Original and reconstructed ECG signals from the SMCMC filtering algorithm with simultaneous model selection for different beat types. The letters in the square boxes are the beat labels. 51

5.5 Illustration of model selection using typical ECG beats. The black asterisks indicate the end-times of the adaptive segments over which the model parameters were assumed static. (a), (c) and (e) N type ECG beat; (b), (d) and (f) L type ECG beat. 52

5.6 Original and reconstructed ECG signal from a typical Monte Carlo run. 54

6.1 Depiction of electrical activity in the heart for a single cardiac cycle (ECG beat) [1]. 57

7.1 Illustration of beat types with morphological similarity and temporal differences using N and A type beats from Record 101 of MIT-BIH arrhythmia database. 75

7.2 Illustration of the features used to form the feature set. The dots denote the selected 5 points in the QRS complex for a typical case. . 77

7.3 Example showing evaluation of performance of the DP algorithm using the different labeling schemes discussed. The labels inside the circle which denotes the cluster, are the true labels and the labels outside the circle in bold are the assigned labels. 80

7.4 Illustration of morphological similarity of F type beats to both N and V beat types. 83

7.5 Example of clustering and labeling performance using Record 207 of the MIT-BIH arrhythmia database. 90

Chapter 1

INTRODUCTION

Electrocardiogram (ECG) signals represent the temporal recordings of electrical activity caused by the constant depolarization and repolarization of cardiac cells during each heart beat. Analysis of ECG recordings is the principal tool used in the diagnosis of cardiac abnormalities and disorders. A variety of signal processing techniques have been used over the years to directly help with classification of ECG signals by constructing modeling and classification algorithms, or to indirectly aid this process by denoising the ECG signal, enhancing signal quality etc. Some of these techniques include eliminating parasitic signals, detecting cardiac cycles, identifying significant complexes or waves, and selecting distinctive feature sets to classify cardiac abnormalities [2].

1.1 ECG Signal Modeling

A crucial step in the classification of ECG signals is the modeling and extraction of information about the signals using the model parameters (feature extraction). The task of ECG modeling, in particular, has resulted in various parametric representations. Early precedents for such representations of ECG signals using signal processing techniques were set by employing orthonormal basis functions [3]. Karhunen-Loeve basis functions which form a set of orthonormal basis functions were used in [4, 5] to optimally represent the ECG signals. Chebyshev polynomials of the first class were used as the orthonormal basis functions in [6] for ECG signal representation. Other orthonormal basis function representations of ECG signals included the use of Hermite polynomials and time-warped polynomials. In [7–9], the similarity of Hermite polynomials to the shape of the QRS complexes in an ECG signal was exploited and the shape of each ECG beat was characterized using the coefficients of the Hermite basis functions. In [10], the differences between the morphologies present in an ECG waveform were exploited

by delineating the signal and modeling the different segments of the signal using Hermite functions of different orders, depending on the presence of smooth waves or sharp peaks. A new class of orthogonal basis functions based on time-warped polynomials were introduced in [11], wherein, the interbeat interval in the ECG signals was represented using an optimally time-warped polynomial.

In addition, ECG signals were studied using autoregressive models in [12] and linear prediction techniques in [13]. Other ECG models utilized the concept of data flow graphs that depend upon the time intervals of the different segments in the ECG signals (P wave, QRS wave, T wave etc.) [14], and principal component analysis that extracts the QRS complex of the ECG signal as the component with the largest variance [15].

Several of the later works on ECG signal modeling have used mathematical representations and fitted these functions to the different fiducial points in ECG signals. In [16], Gaussian Mesa functions and Bi-Gaussian functions were used to fit ECG signals. In [17–19], ECG data was delineated and the various ECG complexes were modeled either by using straight lines or parabolas. In all of these aforementioned works, preprocessing of the ECG signals was necessary to delineate them into the different waves and complexes, such as the P wave, QRS complex etc., and model them using different functions.

Recently, the advent of statistical signal processing techniques has led to the application of sequential Bayesian methods, such as the Kalman filter (KF) [20], extended Kalman filter (EKF) [21], particle filter (PF) [21,22] etc., to several ECG signal processing algorithms, with analysis not being confined to modeling and parameter estimation alone. In particular, noise artifacts and interference effects have been eliminated from the ECG data in [23, 24] and [25] using the KF and EKF, in order to enhance the quality of the recorded ECG data. In addition, Bayesian methods have also been used for estimating the heart rate from

ECG signals, since abnormalities in the heart rate often indicate the presence of arrhythmias. For example, in [26, 27], a KF was used to estimate the heart rate from noisy measurements obtained from multiple fused sensors. In [28], the heart rate variability was calculated from the parameters of a non-linear autoregressive model, estimated using a Kalman smoother.

The dynamical nature of the ECG parameters can be best exploited by the use of statistical models and sequential Bayesian estimation techniques for ECG signal modeling. A joint statistical framework was used in [29] to estimate cardiovascular parameters from multiple signal sources including ECG, arterial blood pressure, intracranial pressure and pulse oximetry signals. In particular, in [29], an EKF was used to estimate the parameters of the cardiovascular signals that were decomposed into cardiac and respiratory components by the algorithm. The ECG signal was represented as a linear combination of weighted frequency harmonics, and the parameters such as the harmonic weights, frequencies and phases were estimated using the EKF. The same formulation was also employed in [30], wherein relevant cardiac parameters and weights of the harmonic model were tracked using both an EKF and a marginalized PF. This model relied heavily on user-specified model paramaters, such as the noise variances, as well as *a priori* information, such as the number of multi-harmonic components present in the data and the mean cardiac frequency.

In [31], the ECG beats were modeled as a trajectory moving around a unit circle in a three-dimensional (3-D) coordinate plane. The trajectory was represented by a system of coupled ordinary differential equations, the solution to which resulted in modeling each ECG beat as a sum of five Gaussian functions. This model was integrated in [23, 24, 32] with a non-linear Bayesian filtering framework such as the EKF to perform ECG denoising by estimating the parameters of the Gaussian functions and reconstructing the original ECG signal. This approach

also depended upon *a priori* information about the underlying dynamics of the ECG signal and needed non-linear solvers in the preprocessing stage to initialize the tracking filter. It was observed that the model was not robust to initialization errors and was also unable to track parameters for ECG signals with abnormalities that appear intermittently in only a few ECG cycles.

1.2 ECG Modeling Contributions

In this work, the focus is on ECG signal modeling methods that:

(a) are robust to initialization errors,

(b) do not require pre-processing steps, *a priori* information or user-defined parameters, and

(c) do not use a single representation to describe ECG signals, which can differ greatly across individuals and disease states.

First, using the multi-harmonic ECG model, an adaptive parameter estimation technique to dynamically select key parameters, such as the number of harmonics and mean frequency, by minimizing the mean-squared error (MSE) between the actual and reconstructed signals is proposed. This adaptivity leads to an improvement in the estimation accuracy of critical heart disease parameters. The designed algorithm uses the EKF to track multi-harmonic ECG model parameters while adaptively selecting the number of harmonics and mean frequency at each time step. Numerical results obtained using real ECG data corroborate the fact that the performance of the adaptive algorithm working under dynamic selection is superior to that of the algorithm using fixed parameters [33].

In addition, two novel ECG modeling methods that enable the representation of ECG fiducial points using multiple models that can account for variations in morphology across different individuals and cardiac abnormalities, are also pro-

posed. The first modeling method leverages the multiple-mode flexibility of the interacting multiple model (IMM) [34] framework to adaptively model ECG signals by using polynomial-order representations that best fit them. The evolving ECG signal dynamics are allowed to switch between three different modes of operation (linear, quadratic, and cubic polynomials) according to first-order Markovian transition probabilities and without requiring prior information or pre-processing initialization steps [35]. The IMM-based ECG model assumes that the model parameters vary slowly between time steps. In high noise scenarios, however, tracking performance can degrade due to the sensitivity of the IMM algorithm to the choice of the transition probability matrix. The second modeling method is based on a sequential Bayesian parameter estimation and simultaneous model selection [36] framework, which does not rely on Markovian mode transition probabilities. For the ECG modeling algorithm in this work, sequential Markov chain Monte Carlo (SMCMC) filtering is used to estimate ECG parameters that are assumed to be static over a time segment; the length of this time segment is determined adaptively using the model likelihood function. Multiple models are used to characterize different ECG signal morphologies with polynomials of different orders. Using real ECG data, it is demonstrated that different ECG signals types can be tracked effectively with this algorithm and exhibit the preference of distinct models depending on the ECG morphology [37]. Both ECG modeling methods allow for model parameters that can be used to distinguish between different types of ECG signals.

1.3 ECG Classification

Cardiovascular disease is considered to be the principal source of death and disability around the world [38]. ECG signals provide a powerful and non-invasive tool for the diagnosis of cardiac diseases, majority of which are preventable and non-life threatening upon timely diagnosis. In order to eliminate the strenuous

process of manual annotation of large amounts of ECG data by cardiologists, automated and computerized analysis of ECG data has gained a lot of importance. For example, in order to diagnose certain infrequently occurring arrhythmias, Holter ECG monitors are used to record upto a week of ECG activity, the manual analysis of which is obviously arduous.

Several automated ECG classification algorithms have been proposed by researchers using a number of different features to represent the ECG signals, and numerous classification methods. Some of the features include heuristic features based on ECG morphology, such as QRS amplitudes and durations [39, 40], time samples selected from the QRS interval [41–43] and time samples from the QRS complex coupled with ECG temporal features [44–48]. Reduced QRS morphology feature sets obtained using principal component analysis of samples from the QRS complex [49] coupled with temporal parameters [50] were also employed. Other features include frequency based features [51], hermite polynomial coefficients [8, 9, 52], Lyapunov exponents [53], autoregressive model coefficients [54], higher order cumulant functions [9], wavelet transform coefficients [55–60], ECG signals expressed as data-process streams [61], ECG signal polarograms [62, 63], fidelity measures of cardiac parameter estimates generated using Bayesian filters [63] and vectocardiogram maximal vector and angle [64],

Classification methods utilized include self organizing maps (SOM) [65], SOM with learning vector quantization [47, 50, 66], self organizing networks [52], back-propogation neural networks [8, 50], multilayer perceptron neural networks [56], block-based neural networks [67], artificial neural networks with particle swarm optimization [57], cross-distance analysis [43], linear discriminants [45, 46, 51, 60, 64], generalized linear model [54], support vector machines (SVMs) [9, 66], kth nearest neighbors [68], active learning techniques [48,58,59], packet-processing concepts such as counter units and hashing functions [61], and fuzzy SVMs [49].

In addition, Bayesian techniques such as Markov modeling [69], Kalman filtering [70] and hidden Markov models [71] have also been used to perform ECG signal classification.

A very common approach to designing ECG classifiers has been to employ classification methods based on supervised learning techniques, which are trained on large ECG datasets and thus do not require any input from an expert to perform labeling (identification) of the different classes [8, 9, 45, 51, 68]. However, ECG signals exhibit large variations across individuals as well as diseases. The shape of the ECG signals and the timing of the various waveforms that comprise the ECG signal depend on the underlying physical conditions of an individual's heart [72]. Thus, such algorithms do not perform well since they do not properly account for inter-patient variation of morphologies, making the system trained on a given set of data ineffective when tested using data that was not represented in the training set [46].

The challenges faced by existing automatic ECG classification algorithms are hence twofold:

(a) On one hand, the algorithms should be able to identify different classes of diseases, given the ECG data that exhibits large variations in morphology. For example, the morphology of a normal ECG beat in one individual will differ greatly from the normal beat morphology of another individual, and if the algorithm is not trained to identify such inter-patient differences in morphology, misclassification can take place.

(b) The second challenge faced by ECG classification algorithms is the availability of large amounts of training and testing data, and also the choice of training data that can attempt to encompass different conditions and morphologies. This presents difficulties in a clinical setting where speedy diagnosis is of the

highest importance, and one cannot rely on the availability of enough training data that covers all the waveform morphologies that might be encountered in the ECG of all individuals.

In order to deal with these challenges, patient-specific (also known as patient-adaptable) classifiers that can account for inter-patient variability have been proposed. These algorithms attempt to adaptively diagnose the cardiac conditions of an individual, rather than being based on diagnosing the cardiac conditions of the general population. One approach to designing patient-specific classifiers is the use of a global and local classifier approach [46, 50], wherein a global classifier is trained on a large set of available data, while the local classifier is trained using data from a specific patient and attempts to profile the nature of an individual's ECG data. In [50], a mixture-of-experts approach was utilized to adapt the algorithm to each individual. This was done by employing a global classifier, which is used to classify an individual's ECG signal based on a large existing database of ECG signals, and a local classifier that is trained specifically using only the individual's ECG record. The results from these two classifiers are combined using the mixture-of-experts (MOE) approach. Although fairly good results were achieved using this method, the method relied heavily on the presence of a large database containing ECG data from many patients, which might not be feasible in general. In [46], a similar global and local classifier approach was used to propose a patient adapting ECG beat classifier. In this work, the global classifier produced a set of beat annotations given an individual's record, which were validated first by an expert. These corrected annotations were then used to train the local classifier, the output of which was then combined with that of the global classifier to produce the final classification output. The training and testing datasets were obtained by equally dividing the available ECG dataset. This might lead to the absence of certain unique morphologies that could belong to a certain

disease, from the training dataset, which could possibly result in misclassification if the disease is present in the testing dataset.

Another approach to designing patient-specific classifiers is through the use of neural networks whose structure can be adaptively varied depending on the different operating environments that might be caused by inter-patient variations in the ECG signal. In [42], a patient-independent neural network was modulated using a three-parameter patient model to achieve adaptability. Optimization of neural network structure for each patient was achieved using block-based neural networks in [67], and particle swarm optimization in [57]. These algorithms used a training set that consisted of two parts; a common set that is the same for all patients and a patient-specific set. Although these methods achieved a good classification performance, they still depend on the availability of a representative common training dataset which can contain patterns that are not included in the patient-specific training set.

Using unsupervised learning algorithms that do not rely on the availability of separate training and testing data sets presents another method for designing patient-specific classifiers. Clustering of the different beats present in an ECG record using features derived from the Hermite polynomial representation of the QRS complex of each beat was presented in [52]. In this work, unsupervised self-organizing neural networks were employed to cluster the data into 25 groups, and it was assumed that expert knowledge is available to perform labeling of these clusters. In [43], a k-means clustering algorithm was first used to determine the clusters in the given data and then a classifier based on cross-distance analysis was used to label the ECG beats in each cluster as normal or abnormal. Unbalanced clustering and a fuzzy SVM were used respectively to cluster and classify ECG data in [49]. The use of a classifier to label the clustered data relies either on the availability of labeled data or an expert such as a cardiologist

who can manually label each cluster. In [60], a Gaussian mixture model whose parameters were determined using an expectation-maximization (EM) algorithm was used to cluster ECG data. The clusters were later labeled automatically using a linear-discriminant classifier, an expert performing the task manually or a combination of both depending on the outcome of a voting process. Although the aforementioned clustering algorithms provide a good means to retain patient-specific information from the ECG data, the number of clusters in the data is assumed to be constant. If the number of clusters is fixed, it might either create more clusters than necessary, or lead to lesser number of clusters than those required to properly separate beats with different shapes. In [73], an attempt was made to select the optimal number of clusters for a Gaussian mixture model using a Bayesian selection criterion in Holter ECG signals. However, since ECG signals are constantly evolving, the number of selected clusters might no longer be optimal if new cardiac conditions arise.

Another patient-adaptive scheme was proposed in [61] to profile an individual's normal ECG behavior based on packet-processing concepts. Although this technique provides a good tool to identify a person's normal ECG behavior it cannot provide further identification of the type of abnormalities that an individual might exhibit. Also, if an person's normal behavior exhibits a change due physical activity etc., the profile of the normal behavior derived previously might cause misclassification of normal behavior as a type of abnormality.

Using active learning and transductive transfer learning, patient-specific classifiers were proposed in [58] and [59], respectively, to classify between normal and ectopic ECG beats. However, these methods were only able to perform two-class classification and although they used no patient-specific training data, a careful choice of examples is required to build an initial training set, which might not be possible in clinical settings. Also, other algorithms that worked with

a limited amount of patient-specific training data were proposed using active learning techniques in [48].

1.4 ECG Classification and Clustering Contributions

In this work, the aim was to develop:

(a) Clustering algorithms for ECG data that do not depend on availability of a large training database, and

(b) Patient-adaptable (patient-specific) ECG classification algorithms so that variations in ECG signals across individuals can be accounted for.

First, a Bayesian maximum-likelihood (ML) classification algorithm that uses features based on the estimated model parameters from the proposed ECG signal models to classify between normal sinus rhythm and different types of cardiac arrhythmias is developed. This was done to demonstrate that the proposed model parameters can be used for ECG arrhythmia classification. The algorithm achieves a correct classification rate of almost 0.9 (90%) for classifying between normal sinus rhythm and three different types of arrhythmia using the adaptively estimated parameters from the multi-harmonic ECG model as features. ECG morphology based features extracted from the parameters of the IMM and SMCMC based ECG models are used to classify between normal and abnormal beats with four different types of arrhythmia, and achieve an average correct classification rate of 0.98 (98%).

Later an adaptive learning method based on the Bayesian nonparametric method known as the Dirichlet process (DP) [74] is used to adaptively cluster the ECG beats corresponding to normal sinus rhythm and different types of arrhythmias. The DP is an unsupervised learning technique which does not require separate training and testing data sets and achieves adaptability due to the fact that it places no restrictions on the number of clusters in the ECG data, i.e.,

the number of models that can represent the given dataset and their parameters. ECG morphology based features from the proposed SMCMC model and temporal information based on the RR-intervals (distance between the R-peaks of two successive ECG beats) are used as features that form the input data set to the DP algorithm. Assuming the availability of expert knowledge in labeling the clusters in the ECG data, results show that on an average 98% of the major beat types are clustered correctly. Correct classification is defined when beats are assigned to a cluster in which a similar type of beat is dominant. In addition, in order to deal with scenarios wherein expert knowledge to annotate ECG beats is not available, the Bayes ML method used to perform cluster labeling is also described, and is shown to correctly label 98.3% of the beat types for which sufficient training data was available.

1.5 Organization

This dissertation is organized as follows. Since, the techniques employed in this work fall under the general framework of Bayesian methods, in Chapter 2, the framework for parameter estimation using noisy measurements is briefly discussed. In Chapter 3, the multi-harmonic ECG model is presented and an algorithm to adaptively estimate the model parameters is proposed. In Chapter 4, IMM-based modeling method is discussed and the state-space framework for ECG signal modeling is outlined; the SMCMC based ECG modeling method with simultaneous model selection is described in Chapter 5. In Chapter 6, the ML classification of different types of ECG signals using the proposed models is discussed, along with a description of the features from the corresponding model parameter estimates that are used by the classifier. The DP algorithm for patient-specific ECG beat clustering along with the ECG signal features used are described in Chapter 7. Finally, Chapter 8 outlines the major results obtained in this work and presents avenues for future research.

Table 1.1: List of acronyms.

Acronym	**Description**
AR	Autoregressive process
DP	Dirichlet process
ECG	Electrocardiogram
EKF	Extended Kalman filter
EM	Expectation-maximization
GMM	Gaussian mixture model
GPB	Generalized pseudo-Bayesian
IF	Instantaneous frequency
IMH	Independent Metropolis-Hastings
IMM	Interacting multiple model
IMM-KF	Interacting multiple model combined with the Kalman filter
KF	Kalman filter
KL	Kullback-Leibler
MC	Monte Carlo
MCMC	Markov chain Monte Carlo
ML	Maximum likelihood
MSE	Mean-squared error
pdf	Probability density function
PF	Particle filter
PVC	Premature ventricular contraction
RMSE	Root mean-squared error
SIS	Sequential importance sampling
SMCMC	Sequential Markov chain Monte Carlo

Table 1.2: List of MIT-BIH arrhythmia notation.

Notation	**Description**
A	atrial premature beat
E	Ventricular escape beats
F	Fusion of normal and ventricular beats
L	Left bundle branch block beats
N	Normal sinus rhythm beats
P	Paced beats
Q	Unclassifiable beats
R	Right bundle branch block beats
S	Supraventrical premature beat
V	Premature ventricular contraction beats
a	Aberrated atrial premature beat
f	Fusion of normal and paced beats
j	Junctional or nodal escape beats

Chapter 2

BAYESIAN APPROACH TO PARAMETER ESTIMATION

Bayesian estimation techniques are a group of algorithms that can extract information about a set of parameters, given noisy measurements and some prior knowledge. The idea is to recursively estimate the unknown parameters that describe the state of a dynamic system using noisy observations made on the system in conjunction with information about the evolution of the system [21]. This kind of analysis of dynamic systems is made possible by employing the state-space approach, which consists of two mathematical models describing the system. The first model, known as the state model (also known as the system model or dynamic model), describes the evolution of the state with time. The second model relates the noisy measurements to the state of the system and is called the measurement model.

2.1 General Bayesian state-space framework for parameter estimation

The state and measurement models form the general state-space framework and can be expressed for discrete-time systems as,

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1}\,, \tag{2.1}$$

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k\,. \tag{2.2}$$

Equation (2.1) describes a Markov state model, where the state vector $\mathbf{x}_k$ represents the unknown state (set of parameters) of the system at time k, the state transition function $\mathbf{f}(\cdot)$ represents the evolution of the unknown state vector parameters with time, and $\mathbf{v}_k$ is the process noise, that is here Gaussian with zero mean and covariance matrix $\mathbf{Q}$. The measurement model is described by (2.2), where $\mathbf{z}_k$ denotes the measurements at time k, $\mathbf{h}(\cdot)$ is the measurement function, and $\mathbf{v}_k$ is the Gaussian observation noise with zero mean and covariance matrix $\mathbf{R}$. The state vector thus contains all the relevant information about the system.

The Bayesian approach assumes that the state and measurement model are available in a probabilistic form and continuously updates the parameters' posterior probability with reception of new measurements [21]. The parameter estimates are constructed recursively using the posterior probability density function (pdf) of the state, given the sequentially obtained measurements. Thus, at a time step k, the posterior pdf of the state $\mathbf{x}_k$ given by $p(\mathbf{x}_k|\mathbf{Z}_k)$, is obtained using the set of measurements $\mathbf{Z}_k = \{\mathbf{z}_1, \ldots, \mathbf{z}_k\}$. It is assumed that the initial pdf or the prior of the state vector $p(\mathbf{x}_0|\mathbf{z}_0) = p(\mathbf{x}_0)$, is available. The first step in obtaining the posterior pdf of the state is the prediction stage that involves using the state model shown in (2.1) and the pdf from the previous time $k-1$ to obtain the prior pdf as [22],

$$p(\mathbf{x}_k|\mathbf{Z}_{k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1}) d\mathbf{x}_{k-1}. \tag{2.3}$$

In the above equation, $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ represents the probabilistic model of the state and is given by the state model in (2.1).

Using Bayes' theorem the sequential update for the posterior pdf is then obtained by incorporating the measurement $\mathbf{z}_k$ that becomes available at time k, and updating the prior pdf as [22],

$$p(\mathbf{x}_k|\mathbf{Z}_k) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)\, p(\mathbf{x}_k|\mathbf{Z}_{k-1})}{\int p(\mathbf{z}_k|\mathbf{x}_k)\, p(\mathbf{x}_k|\mathbf{Z}_{k-1})}. \tag{2.4}$$

This means that the posterior pdf at time k is calculated using the posterior pdf from the previous time step $k-1$, by taking into account the new information obtained from the current measurement using the likelihood function $p(\mathbf{z}_k|\mathbf{x}_k)$ that is defined using the measurement model in (2.2). In addition, the denominator in (2.4) is a normalizing constant denoted by $p(\mathbf{z}_k|\mathbf{Z}_{k-1})$.

An analytical solution to solving (2.4) and obtaining the posterior pdf is available only if the system and measurement models satisfy a restrictive set of conditions. Since in a practical situation, these restrictive conditions are not

always satisfied, analytical approximations and suboptimal Bayesian methods can be employed to obtain the posterior pdf [21]. In some cases, however, optimal solutions to this problem are available under the following scenarios:

(a) The state and measurement functions of the system are linear and the noise is Gaussian. In this case, the Kalman filter can used to recursively solve for the posterior pdf by computing its mean and covariance, and is the optimal estimator in the least mean-square sense [20].

(b) The state-space of the system is a finite, discrete-valued sequence, the solution to which can be obtained using grid based search methods [21].

(c) The posterior pdf of a nonlinear dynamic system has a sufficient statistic. In this case, Beneš [75] and Daum filters [76] can be applied to find solve for the pdf.

But if such scenarios do not exist, especially since many systems are nonlinear in nature, the following approximations to finding the posterior pdf can be employed.

(a) Analytical solutions such as the EKF [77], which approximate the nonlinear functions using the Taylor series expansion, thus approximating the posterior pdf as a Gaussian whose mean and covariance can be then computed recursively as in the KF.

(b) Numerical methods that are approximate grid-search methods based on discretizing the state-space and approximating the posterior pdf as a summation over the entire grid [21].

(c) Gaussian sum filters which approximate the posterior pdf as a weighted sum of Gaussians instead of as a single Gaussian [78].

(d) Unscented KF [79] that approximates the posterior pdf using a Gaussian density which is represented using a set of deterministically chosen samples that specify its mean and covariance.

(e) More generally, when the models are nonlinear and the noise in the models is non-Gaussian, the posterior pdf is estimated using sequential Monte Carlo methods [80], which include the particle filter [22], and are based on representing the posterior pdf using a set of particles and their corresponding weights.

2.2 General Bayesian state-space framework for parameter estimation using multiple models

In addition, certain dynamic systems are best described using multiple operating modes. This is based on the fact that the behavior of these systems cannot be characterized at all times using a single model, but a finite number of models are required to describe its behavior at different times. Such systems are referred to as hybrid systems, whose state and measurement models exhibit an additional dependance upon the operating model at a given time k. Parameter estimation for such systems is done in a manner similar to recursive Bayesian estimation by constructing the posterior pdf of the estimates from each corresponding model, using a two-step prediction and update approach. However, the only difference is that in addition to estimating the parameters, the model that best describes the system at a given time also has to be estimated.

Since at each time instant k, the system depends on an additional parameter, which is the mode m_k of the system, the general state-space equations in (2.1) and (2.2) can be re-written for a multiple mode setup as [21],

$$\mathbf{x}_k = f_{m_k}(\mathbf{x}_{k-1}) + g_{m_k}(\mathbf{w}_{k-1}), \tag{2.5}$$

$$\mathbf{z}_k = h_{m_k}(\mathbf{x}_k) + q_{m_k}(\mathbf{v}_k). \tag{2.6}$$

In the above equations, $f_{m_k}(\mathbf{x}_{k-1})$ is the mode-dependent state transition function, $g_{m_k}(\mathbf{w}_{k-1})$ is the modeling error function, $h_{m_k}(\mathbf{x}_k)$ is the measurement function and $q_{m_k}(\mathbf{v}_k)$ is the measurement noise function under mode m_k at time k. The system mode can be modeled by an M-state first-order Markov chain with transition probabilities $\pi_{ij} = \Pr(m_k = j \mid m_{k-1} = i)$, given that the models i and j were in effect at times $k-1$ and k respectively, for $i, j = 1, \ldots, M$, where M is the total number of models required to describe the system. The initial mode probabilities are given by $\mu_0^i = \Pr(m_0 = i)$ [21]. It is also assumed that the initial pdf of the state vector $p(\mathbf{x}_0, m_0 = i|\mathbf{z}_0) = p(\mathbf{x}_0, m_0 = i)$, is available for $i = 1, \ldots, M$. The prediction step which yields the prior pdf is now given by [21],

$$p(\mathbf{x}_k, m_k = j|\mathbf{Z}_{k-1}) = \sum_{i=1}^{M} \pi_{ij} \int p(\mathbf{x}_k|\mathbf{x}_{k-1}, m_k = j) p(\mathbf{x}_{k-1}, m_{k-1} = i|\mathbf{Z}_{k-1}) d\mathbf{x}_{k-1} \quad (2.7)$$

where, $p(\mathbf{x}_k|\mathbf{x}_{k-1}, m_k = j)$ represents the probabilistic state model for model j defined in (2.5) and $p(\mathbf{x}_{k-1}, m_{k-1} = i|\mathbf{Z}_{k-1})$ is the posterior pdf for model i that was in effect at time $k-1$.

Using the prior pdf, the posterior pdf is now computed in the update stage using the most recent measurement $\mathbf{z}_k$ as [21]

$$p(\mathbf{x}_k, m_k = j|\mathbf{Z}_k) = \frac{p(\mathbf{z}_k|\mathbf{x}_k, m_k = j)\, p(\mathbf{x}_k, m_k = j|\mathbf{Z}_{k-1})}{\sum_{j=1}^{M} \int p(\mathbf{z}_k|\mathbf{x}_k, m_k = j)\, p(\mathbf{x}_k|\mathbf{Z}_{k-1})}. \quad (2.8)$$

In the above equation, $p(\mathbf{z}_k|\mathbf{x}_k, m_k = j)$ is the likelihood function for model j calculated using the corresponding measurement model defined in (2.6) and the denominator is a normalizing constant known as the model-conditioned likelihood denoted by $p(\mathbf{z}_k|\mathbf{Z}_{k-1}, m_k = j)$.

Hybrid state estimation algorithms consist of a bank of model-matched filters each of which yields the posterior pdfs of the corresponding models as described in (2.7) and (2.8). These model-matched filters can correspond to a KF [34, 81, 82], EKF [34], PF [83], etc. depending on the nature of the state and

measurement models. In addition, hybrid state estimation algorithms also need some approach to determine the method of cooperation between each of the model-matched filters, in order to determine the final estimate of the system that takes into account the contribution from the estimates of each model. Some of these approaches include the Generalized Pseudo-Bayesian (GPB) method [84–86], the IMM algorithm [34, 81, 82], reversible jump Markov Chain Monte Carlo (MCMC) [87], multiple model pruning [88] etc.

2.3 Application of Bayesian approach to ECG signal modeling

In this work, Bayesian methods are applied to model the ECG signals and estimate the cardiac parameters of interest. In this case, the state vector represents the cardiac parameters, which might vary depending on the state-space framework used. The measurement model represents the ECG signal as a function of the corresponding cardiac state parameters. The KF, EKF and SMCMC methods are used to estimate the posterior pdf of the state given the real ECG data (measurements), depending on the type of state-space framework. The KF and SMCMC methods are also combined with the IMM and reversible jump MCMC, respectively, to perform simultaneous model selection for modeling the ECG signals using multiple models without being restricted to a single mathematical model (representation). In addition, the MCMC method method known as Gibbs sampling is employed to estimate the posterior pdf of the model parameters of the different mixtures for a DP mixture model that is used to cluster the ECG data.

Chapter 3

ADAPTIVE PARAMETER ESTIMATION USING MULTI-HARMONIC ECG MODEL

Adaptive parameter estimation techniques provide the flexibility to optimize the estimator over a conceivable parameter space by minimizing the appropriate cost function, making them extremely useful for the ECG parameter estimation problem. The shape and nature of the ECG signals vary greatly across different types of cardiac diseases. Also, ECG signals are unique to each individual. Thus, statistical models of ECG signals, that are constructed to estimate cardiac parameters, should take these variations into account and refrain from making overly restrictive assumptions about these signals without requiring *a priori* knowledge about the ECG signals. In order to preclude their dependence on user-defined parameters and *a priori* information, statistical models of ECG signals can take advantage of the adaptive parameter estimation framework by adaptively varying certain parameters of the model such that a cost function is minimized.

In this chapter, adaptive parameter estimation of cardiac signal parameters is demonstrated using the multi-harmonic components model of the ECG signal proposed in [29,30]. The state-space model for ECG signals is first provided using the multi-harmonic components framework. The method for adaptive parameter estimation is outlined next, followed by simulation results which demonstrate that the adaptive algorithm achieves better MSE performance compared to the non-adaptive method, thus improving the estimation accuracy of the algorithm.

3.1 State-space Model using Multi-harmonic Components

Cardiac signals can be considered quasi-periodic and have been described using a sinusoidal measurement model with multiple harmonically-related components that vary slowly in frequency and amplitude [29,30]. Consider an ECG signal $z(t)$

sampled at the rate T_s to obtain the measurement $z_k = z(k\,T_s)$ at time k. The multi-harmonic measurement model for z_k can be represented as [30]

$$z_k = \bar{z}_k + \left[\sum_{h=1}^{N_h} \Big(b_{k,h}\, \cos(h\,\theta_k) + c_{k,h}\, \sin(h\,\theta_k) \Big)\right] + v_k, \tag{3.1}$$

where $\bar{z}_k$ is a low-frequency signal trend, N_h is the number of harmonics, $b_{k,h}$ and $c_{k,h}$ are slowly-varying amplitudes corresponding to the hth harmonic, θ_k is the instantaneous phase of the fundamental component at time k, and v_k is zero-mean white Gaussian measurement noise.

In order to estimate dynamic parameters of interest in (3.1) using sequential Bayesian estimation methods, a state space formulation is needed. So in addition to the measurement equation in (3.1), a state equation for the cardiac parameter vector of clinical significance $\mathbf{x}_k = [\theta_k\ f_k\ \mathbf{b}_k\ \mathbf{c}_k\ \bar{z}_k]^T$, is formulated. Here, $\mathbf{b}_k = [b_{k,1}\ \ldots\ b_{k,N_h}]^T$, $\mathbf{c}_k = [c_{k,1}\ \ldots\ c_{k,N_h}]^T$, f_k is the instantaneous frequency (IF), and T denotes vector transpose. The proposed state model in [30] models fluctuations in the instantaneous phase as a first-order approximation of an integral of the IF; and the IF is modeled as a first-order autoregressive (AR) process with mean cardiac frequency $\bar{f}$ and AR coefficient α. The remaining state variables are modeled by a random walk model. Thus, the state equation is given by [30],

$$\mathbf{x}_k = [\mathrm{mod}_{2\pi}\{\theta_{k-1} + 2\pi\,T_s\,f_{k-1}\}\ \ \alpha(f_{k-1} - \bar{f}) + \bar{f}\ \ \mathbf{b}_{k-1}\ \ \mathbf{c}_{k-1}\ \ \bar{z}_{k-1}]^T + \mathbf{u}_{k-1} \tag{3.2}$$

In the above equation, $\mathbf{u}_{k-1} = [u_{\theta_{k-1}}\ u_{f_{k-1}}\ u_{b_{k-1,1}}\ \ldots\ u_{b_{k-1,N_h}}\ u_{c_{k-1,1}}\ \ldots\ \ u_{c_{k-1,N_h}}\ u_{\bar{z}_{k-1}}]^T$ is a zero-mean white Gaussian modeling error process, and $\mathrm{mod}_{2\pi}$ is used to keep θ_{k-1} within the range $[0, 2\pi)$.

3.2 Framework for Adaptive Signal Parameter Estimation

As seen from the state and measurement equations for the multi-harmonic components model of ECG signals given in Section 3.1, the model depends on user-defined parameters, such as mean cardiac frequency and process noise variances,

and *a priori* information, such as the number of harmonic components, that has to be obtained through spectral analysis. However, the user-defined parameters and *a priori* information may vary greatly from person to person, as well as across various cardiac diseases. In the previous works of [29] and [30], only fixed values of these parameters were used in the EKF or PF recursion when tracking the unknown state parameters. To reduce the dependence of the system on user-defined parameters and pre-processing steps, we employ a framework that enables adaptive selection of the optimum number of harmonics N_h and mean cardiac frequency $\bar{f}$ in (3.2). These two parameters are considered as the adaptive parameters of the system.

The system parameters are adaptively optimized via selection of the parameter pair that minimizes a cost function metric. In this work, the cost function is chosen as the MSE between the actual and reconstructed signal. The number of harmonics and mean cardiac frequency are allowed to vary in a range $[N_h^{(\min)}, N_h^{(\max)}]$ and $[\bar{f}^{(\min)}, \bar{f}^{(\max)}]$, respectively. The ranges are chosen to represent a reasonable span of the number of multi-harmonic components and a conceivable selection of mean frequencies. In addition, as the state and measurement equations in (3.1) and (3.2) are nonlinear, the EKF tracker is utilized to estimate the state vector over the specified ranges of the adaptive parameters.

The ECG signal is divided into segments of fixed length and adaptive parameter selection is performed separately over each segment. With the segment length taken as approximately equal to the beat length, this approach is feasible for those types of ECG signals which do not usually exhibit sudden changes in their morphologies within a beat. For each candidate parameter pair, the MSE is calculated for the segment as the energy of the error signal normalized by the ECG signal energy. In particular, the reconstruction MSE for the lth segment

with N_l samples can be approximated as

$$\text{MSE}_l = \frac{\sum_{k=1}^{N_l} (z_{k,l} - \hat{z}_{k,l})^2}{\sum_{k=1}^{N_l} z_{k,l}^2}. \tag{3.3}$$

Here, $z_{k,l}$ is the kth sample in the lth segment of the actual measurements and $\hat{z}_{k,l}$ is the reconstructed signal using the state estimate. The algorithm then selects the best (adaptive) parameter pair for the segment as the one that results in minimum MSE.

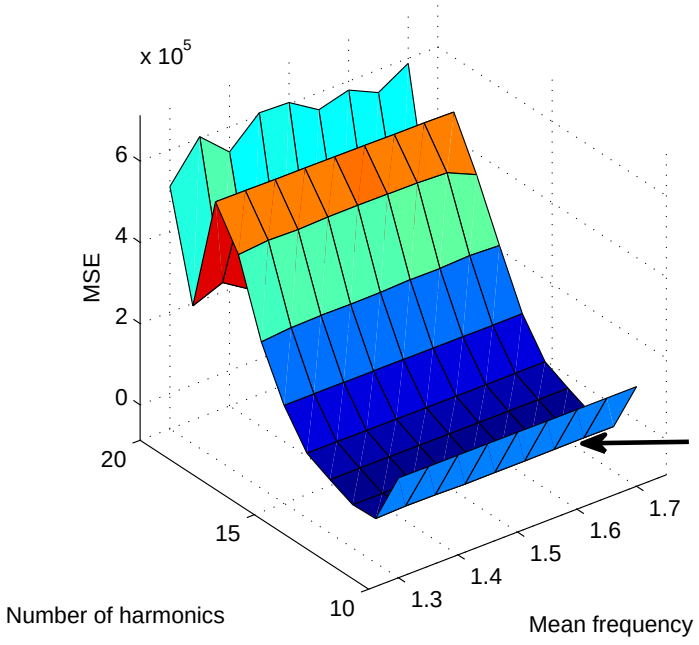


Figure 3.1: Surface plot of the MSE for the entire chosen range of the adaptive parameters for a segment from a normal sinus rhythm ECG signal. The arrow indicates the minimum MSE point for one adaptive parameter pair.

To summarize, real ECG signals (obtained from the online Physionet database [89]) are first divided into equal length segments; the cardiovascular parameters of the segments are then estimated using the EKF tracker. For each segment, this estimation is performed over the range of the number of harmonics $[N_h^{(\min)}, N_h^{(\max)}]$

and the range of the mean cardiac frequency $[\bar{f}^{(\min)}, \bar{f}^{(\max)}]$. The adaptive parameters for each segment are selected as the ones that minimize the MSE calculated using (3.3). The estimates provided by the adaptive parameters are subsequently used to categorize different types of ECG signals. The key steps in the adaptive multi-harmonic ECG modeling method are shown in Algorithm 1 below.

Algorithm 1 Adaptive Multi-harmonic ECG Modeling

1. Divide the ECG signal into L equal length segments, $l = 1, \ldots, L$.
2. In the lth segment, perform the following steps:
   (a) Set the number of harmonics N_h and mean cardiac frequency $\bar{f}$ in the ECG state-space model (3.1)-(3.2) to vary within the range $[N_h^{(\min)}, N_h^{(\max)}]$ and $[\bar{f}^{(\min)}, \bar{f}^{(\max)}]$, respectively.
   (b) For each N_h and $\bar{f}$ in Step 2 (a), estimate the cardiovascular parameters of the ECG data in the lth segment (time $k = 1, \ldots, N_l$) using the EKF tracker [21].
   (c) Using (3.3), compute the MSE between the measured ECG signal and the reconstructed ECG signal obtained from the estimated cardiovascular parameters.
   (d) Select the best (adaptive) number of harmonics N_h and mean cardiac frequency $\bar{f}$ for the lth segment as those which minimize the MSE.

3.3 Simulations and Discussion

In order to demonstrate the performance of the multi-harmonic ECG model performing adaptive parameter estimation, real ECG signals from the online Physionet database [89] were utilized.

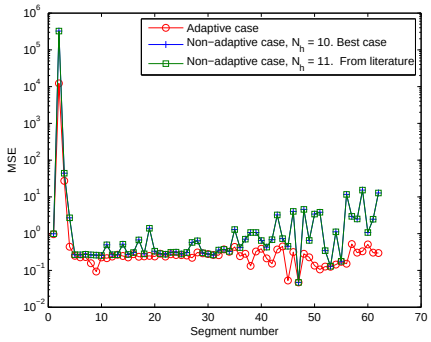
The ECG signals used to demonstrate the results were obtained from the MIT-BIH normal sinus rhythm, MIT-BIH supraventricular arrhythmia, MIT-BIH malignant ventricular ectopy, and MIT-BIH atrial fibrillation databases [89]. The ECG signals were sampled at 500 Hz. The EKF tracker was applied to estimate the model parameters. The estimation was performed with the number of har-

monics spanning the range $[N_h^{(\min)}, N_h^{(\max)}] = [10, 20]$ and mean cardiac frequency spanning the range $[\bar{f}^{(\min)}, \bar{f}^{(\max)}] = [1.3, 1.5]$ Hz. These ranges were chosen based on the values used for these parameters in the literature. As mentioned in Section 3.2, the ECG signals were divided into segments of length approximately equal to the ECG beat length. For the cases where the parameters are not varied adaptively, the number of harmonics is kept fixed at $N_h = 11$ (this is the value used in previous works [30]) and the mean cardiac frequency is set to $\bar{f} = 1.5$ Hz. For additional comparison purposes, an alternate choice is also considered for fixed number of harmonics, as the value that was chosen most frequently by the algorithm as the best parameter minimizing the MSE; this is referred to as the 'best case'. The AR parameter in the model was taken as $\alpha = 0.9987$.

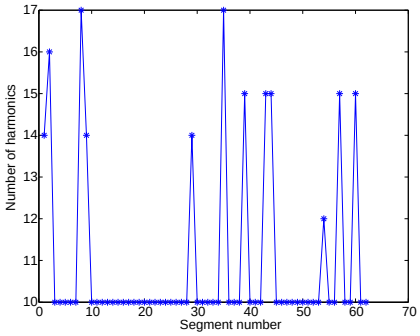
The performance of the algorithm working under adaptive conditions, wherein the best model parameter pair (number of harmonics and mean cardiac frequency) is selected to minimize the reconstruction MSE (3.3) at each time step, is compared to the performance of the algorithm using a fixed set of parameters. Figures 3.2(a), 3.2(c), 3.2(e), and 3.2(g), show plots of the MSE as a function of the segment number for ECG signals with normal sinus rhythm and supraventricular arrhythmia, malignant ventricular arrhythmia, and atrial fibrillation, respectively. For the adaptive parameter case, the minimum MSE values are shown as a function of the segment number. In the cases where the parameters are non-adaptive, the MSE is shown for the two fixed parameter cases as mentioned above. It can be seen that by adaptively optimizing the model parameters, the performance of the algorithm can be improved significantly. This is because the algorithm adaptively selects the model parameters that result in the lowest MSE, instead of performing estimation with fixed user-defined parameter values. Figures 3.2(b) and 3.2(d), 3.2(f), and 3.2(h), show the variation of the optimum number of harmonics as a function of the segment number for signals

with normal sinus rhythm, supraventricular arrhythmia, malignant ventricular arrhythmia, and atrial fibrillation, respectively. This elucidates the adaptive ability of the algorithm in selecting the best value of the number of harmonics that minimizes the MSE.

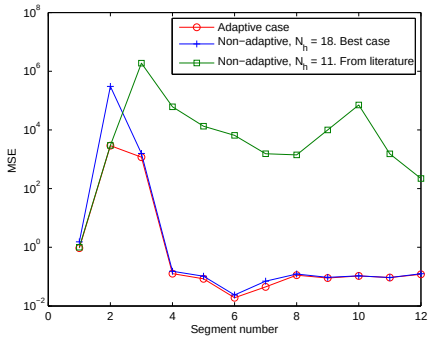
Figure 3.1 shows a surface plot of the MSE for the entire chosen range of the adaptive parameters for a segment from a normal sinus rhythm ECG signal. The plot illustrates the minimum MSE obtained for a specific parameter pair and shows how the algorithm selects the best adaptive parameters without requiring *a priori* information. Note that in the example shown, the variation of the MSE with the mean cardiac frequency is much smaller than that with the number of harmonics. Because of this, the variation of the optimum mean cardiac frequency for each segment is not shown.



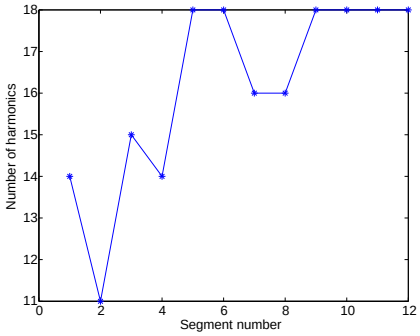
(a) Normal sinus rhythm



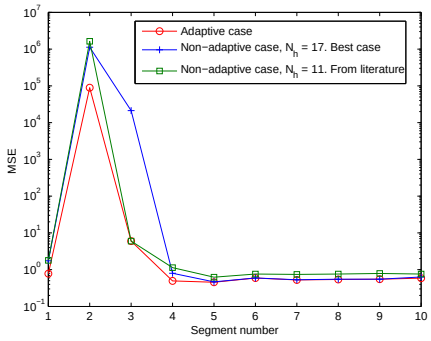
(b) Normal sinus rhythm



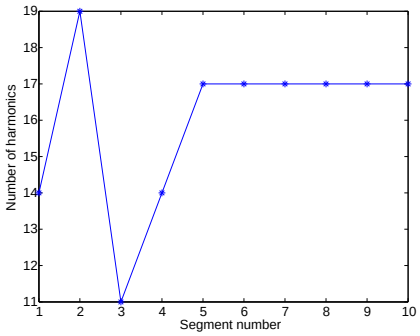
(c) Supraventricular arrhythmia



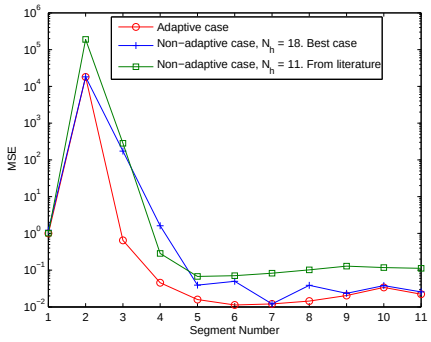
(d) Supraventricular arrhythmia



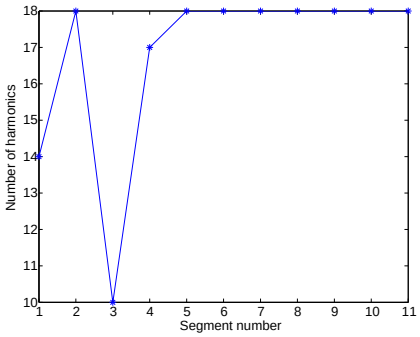
(e) Malignant ventricular arrhythmia



(f) Malignant ventricular arrhythmia



(g) Atrial fibrillation



(h) Atrial fibrillation

Figure 3.2: MSE of the multi-harmonic ECG modeling algorithm with fixed and adaptive model parameters, and variation in the optimum number of harmonics for different types of ECG signals.

Chapter 4

ECG MODELING USING INTERACTING MULTIPLE MODELS

The concept of multiple switching dynamic models is relevant in a number of applications since many dynamic systems can be better characterized by a set of possible modes of operation rather than a single mode [81]. The IMM algorithm [34] can be used to describe such hybrid systems and can estimate states and modes that transition according to a Markov process. In particular, ECG signals fit perfectly into this framework because of their dynamical nature and the presence of morphologies that vary between individuals and diseases. The multiple model framework can be used to describe the ECG signal by adaptively using different representations (models) depending on the nature of the signal, and the IMM algorithm is used estimate the parameters of each of these models.

In this chapter, a multiple model setup for describing ECG signals is proposed using the IMM framework. The IMM algorithm for dynamically changing systems is first described, followed by the proposed state-space framework for ECG signals. Next, the framework for modeling ECG signals using multiple models is described in detail. Finally, simulation results showing the effectiveness of the algorithm in tracking different types of ECG signal morphologies by adaptively choosing the model are demonstrated using real ECG signals.

4.1 IMM Algorithm

At each time k, in a multiple mode framework the system also depends on the system mode m_k, and thus the general state-space equations can be written as [21]

$$\mathbf{x}_k = \mathbf{f}_{m_k}(\mathbf{x}_{k-1}) + \mathbf{g}_{m_k}(\mathbf{u}_{k-1}), \tag{4.1a}$$

$$\mathbf{z}_k = \mathbf{h}_{m_k}(\mathbf{x}_k) + \mathbf{q}_{m_k}(\mathbf{v}_k), \tag{4.1b}$$

where $\mathbf{f}_{m_k}(\cdot)$ is the mode-dependent state transition function, $\mathbf{g}_{m_k}(\cdot)$ is the modeling error function, $\mathbf{h}_{m_k}(\cdot)$ is the measurement function, and $\mathbf{q}_{m_k}(\cdot)$ is the measure-

ment noise function, under mode m_k at time k. The system mode is modeled by an M-state first-order Markov chain with transition probabilities $\pi_{ij} = \Pr(m_k = j \mid m_{k-1} = i)$, for $i, j = 1, \ldots, M$. The initial mode probabilities are given by $\mu_0^{(i)} = \Pr(m_0 = i)$, $i = 1, \ldots, M$ [21].

The IMM algorithm comprises of three stages, namely interaction, filtering, and combination [34], performed at each time step k as follows. In the interaction stage, the mixing probabilities of the system defined as $\mu_{k-1}^{(i|j)} = \Pr(m_{k-1} = i \mid m_k = j, \mathbf{Z}_{k-1})$ and conditioned on the set of measurements $\mathbf{Z}_{k-1} = \{\mathbf{z}_1, \ldots, \mathbf{z}_{k-1}\}$ are first calculated using the mode probabilities $\mu_{k-1}^{(i)}$ from the previous time step as

$$\mu_{k-1}^{(i|j)} = \frac{\pi_{ij}\, \mu_{k-1}^{(i)}}{\sum_{i=1}^{M} \pi_{ij}\, \mu_{k-1}^{(i)}}, \quad i, j = 1, \ldots, M. \tag{4.2}$$

The mixing probabilities are next used to calculate the filter input parameters for the prediction step of the filtering stage as a weighted sum of the system parameter estimates from the previous time step. For example, the input state parameter to each mode-matched filter (the filter which computes the estimates for the corresponding state model) is computed as a mixture or interaction of previous state parameter estimates, given by

$$\hat{\mathbf{x}}_{k-1|k-1}^{(0j)} = \sum_{i=1}^{M} \mu_{k-1}^{(i|j)}\, \hat{\mathbf{x}}_{k-1|k-1}^{(i)}, \quad j = 1, \ldots, M, \tag{4.3}$$

where $\hat{\mathbf{x}}_{k-1|k-1}^{(0j)}$ is the input state parameter for mode j and $\hat{\mathbf{x}}_{k-1|k-1}^{(i)}$ is the state parameter estimate for mode i at the previous time $k-1$.

The filtering stage can be considered as a bank of mode-matched filters, each of which uses the current measurement $\mathbf{z}_k$ and performs a prediction and update step to provide an updated distribution of the system parameters at time k under the corresponding mode of operation. The updated state parameter

estimates are denoted by $\hat{\mathbf{x}}_{k|k}^{(j)}$, $j = 1, \ldots, M$. Additionally, the mode likelihoods

$$\Lambda_k^{(j)} = p\left(\mathbf{z}_k | \mathbf{Z}_{k-1}, m_k = j\right), \quad j = 1, \ldots, M, \tag{4.4}$$

are calculated, typically using a Gaussian approximation as

$$p\left(\mathbf{z}_k | \mathbf{Z}_{k-1}, m_k = j\right) \approx \mathcal{N}(\mathbf{z}_k; \mathbf{h}_j(\hat{\mathbf{x}}_{k|k-1}^{(j)}), \mathbf{S}_k^{(j)}) \,. \tag{4.5}$$

In the above equation, it can be seen that the likelihood for mode j can be approximated using a Gaussian density that is a function of the measurement $\mathbf{z}_k$ with mean $\mathbf{h}_j(\hat{\mathbf{x}}_{k|k-1}^{(j)})$ representing the estimated measurement using the corresponding state estimate from the prediction step and a covariance $\mathbf{S}_k^{(j)}$. The posterior mode probabilities are updated for current time k as

$$\mu_k^{(j)} = \frac{\Lambda_k^{(j)} \sum_{i=1}^{M} \pi_{ij} \, \mu_{k-1}^{(i)}}{\sum_{j=1}^{M} \Lambda_k^{(j)} \sum_{i=1}^{M} \pi_{ij} \, \mu_{k-1}^{(i)}} \,, \tag{4.6}$$

for $j = 1, \ldots, M$. In the final combination stage, the estimates from the different mode-matched filters are weighted by the corresponding mode probabilities and aggregated to calculate the overall estimate of the system parameters. The combined state parameter estimate is

$$\hat{\mathbf{x}}_{k|k} = \sum_{j=1}^{M} \mu_k^{(j)} \, \hat{\mathbf{x}}_{k|k}^{(j)} \,. \tag{4.7}$$

4.2 State-space Model using IMMs

To describe the ECG signals using a multiple mode setup, time-domain polynomial functions of different orders with coefficients that vary over time were used. In this framework each polynomial order is associated with a mode of the system's operation and the coefficients of the corresponding polynomial function form the state vector. The measurement model for the time-domain ECG signal can be written as

$$z_k = a_{k,0} + \sum_{m=1}^{M} a_{k,m} \, k^m + v_k, \tag{4.8}$$

where $a_{k,0}, \ldots, a_{k,M}$ are the polynomial coefficients, and v_k is zero-mean white Gaussian measurement noise with variance R. The ECG signal model in (4.8) can also be written in matrix-vector form as

$$z_k = \mathbf{H}_{m_k} \mathbf{x}_k + v_k, \tag{4.9}$$

where $\mathbf{H}_{m_k} = [1 \; k \; k^2 \; \ldots \; k^M]$, and the polynomial coefficients form the state vector $\mathbf{x}_k = [a_{k,0} \; \ldots \; a_{k,M}]^T$. In particular, using a polynomial function of order m_k for mode $m_k \in \{1, \ldots, M\}$, only the first m_k+1 coefficients in $\mathbf{x}_k$ are non-zero.

The state evolves in time according to a simple first-order Markov model, given by

$$\mathbf{x}_k = \mathbf{F}_{m_k} \mathbf{x}_{k-1} + \mathbf{G}_{m_k} \mathbf{u}_{k-1}. \tag{4.10}$$

The state transition and modeling error matrices, $\mathbf{F}_{m_k}$ and $\mathbf{G}_{m_k}$, are $(M+1) \times (M+1)$ diagonal matrices whose m'th diagonal element is given by $[\mathbf{F}_{m_k}]_{m'} = [\mathbf{G}_{m_k}]_{m'} = 1$ for $m' = 1, \ldots, m_k + 1$ and $[\mathbf{F}_{m_k}]_{m'} = [\mathbf{G}_{m_k}]_{m'} = 0$ for $m' = m_k + 2, \ldots, M + 1$. Thus, both $\mathbf{F}_{m_k}$ and $\mathbf{G}_{m_k}$ are mode-dependent matrices, and are used to ensure that when the system is under mode m_k, the polynomial coefficients $a_{k,m}$ are zero for $m > m_k$. The variance of the zero-mean white Gaussian modeling error $\mathbf{u}_{k-1}$ is small, so that the model describes polynomial coefficients varying slowly over time. The slowly varying linear, quadratic, and cubic polynomial modes are thus used to represent the morphology of the ECG signals over short time segments, such as the P wave, PR segment, the Q, R and S waves of the QRS complex, etc.

4.3 Framework for ECG Modeling with IMM

As discussed in Section 4.2, the ECG signals are modeled using polynomial functions of different orders. Since the IMM offers multi-mode flexibility, ECG signals of different morphologies, even those with abrupt changes, can be modeled using

transitioning polynomials of different orders depending on the nature of the data. This avoids the problems encountered in [24, 32], where due to phase-wrapping of each beat of the ECG signal, abrupt morphological changes, such as premature ventricular contractions (PVC), cannot be tracked. In addition, the proposed model is fairly straightforward and does not require any preprocessing steps to initialize the tracker. Also, different portions of the ECG signal, such as the P wave, QRS complex, and T wave, do not have to be delineated before modeling using mathematical representations as in [8, 9, 17, 19], since the system mode can automatically adapt at each time step if the need arises.

For the real ECG data used in this work, polynomial functions of three different orders, namely linear, quadratic, and cubic polynomials, were found to be sufficient. Hence the number of modes (highest polynomial order) was set to $M = 3$. Also, the state-space model in (4.9)and (4.10) is linear and Gaussian, and hence a KF can be used in the filtering stage of the IMM algorithm to estimate the unknown parameters (polynomial coefficients). This is known as the IMM-KF algorithm [82].

Based on the parameter estimates obtained from the IMM-KF algorithm, classification of ECG signals into different classes is proposed. Since this classification needs to be performed for each beat of the ECG data, the data is first divided into beats based on the peak location and phase as in [32]. The IMM-KF algorithm is re-initialized at the beginning of each new beat and parameters are estimated afresh. In addition, a simple low-pass filter is used to eliminate the effects of baseline wander and power-line interference on the parameter estimates [45].

To summarize, real ECG signals (obtained from the MIT-BIH arrhythmia database [89]) are first separated into beats based on the peak location (provided by the MIT-BIH arrhythmia database) and phase as in [32]. The ECG beats are then modeled using the IMM with $M = 3$ polynomial modes of operation, and

the IMM-KF is used to estimate the coefficients of the polynomial used by each mode. The estimated polynomial coefficients are subsequently used as features in a classifier to differentiate between different types of ECG signals. The main steps of the IMM ECG modeling method are shown in Algorithm 2 below.

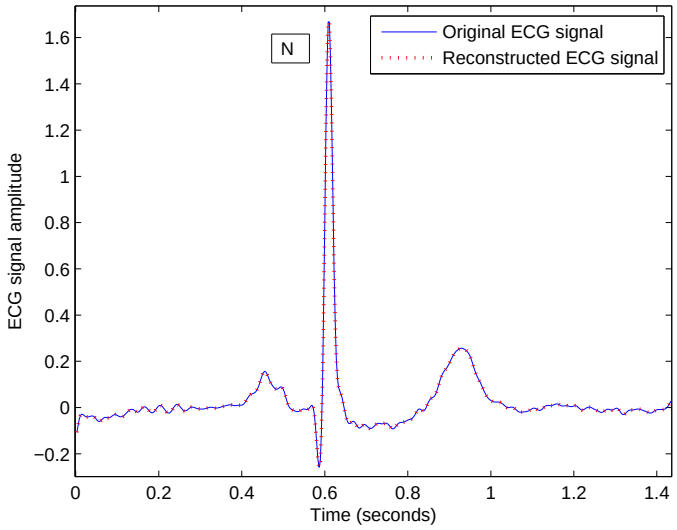
Algorithm 2 IMM ECG Modeling

1. Initialize the polynomial coefficients and the transition and mode probabilities for the $M = 3$ linear, quadratic, and cubic order polynomial ECG models.
2. For time $k = 1, 2, \ldots$, perform the following steps:
   (a) Compute the IMM mixing probabilities and input parameters for the $M = 3$ mode-matched filters using the transition and mode probabilities and polynomial coefficients from time $k - 1$ via (4.2)-(4.3).
   (b) Estimate the polynomial coefficients at time k for the $M = 3$ modes using the KF with the ECG state-space model (4.8)-(4.10), the input parameters from Step 2 (a), and the ECG measurement at time k. Also calculate the probabilities for the $M = 3$ polynomial modes at time k using (4.4)-(4.6).
   (c) Obtain the noise-free reconstructed ECG signal at time k using the combined polynomial coefficient estimate (4.7) with the measurement function in (4.9).

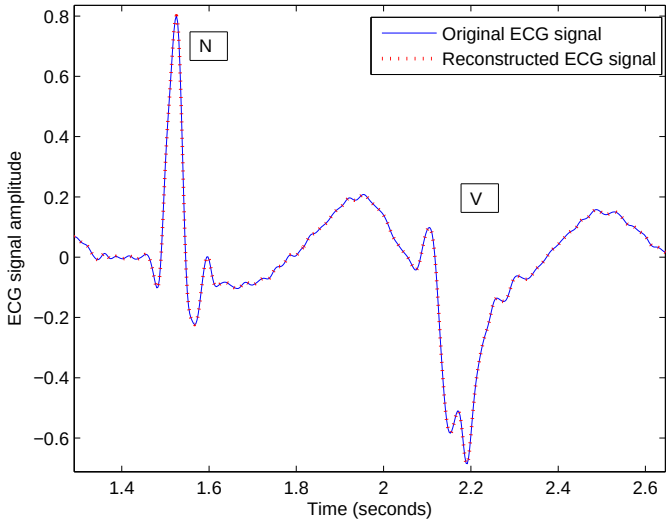
Modeling ECG data via the IMM framework provides an approach to adaptively utilize different order polynomial representations for the ECG signals depending on their morphology. The IMM based state space ECG model utilizes switching polynomial modes with slowly-varying coefficients to represent ECG morphologies over short time segments. As will be demonstrated in the following Section, the algorithm can track different types of ECG signals without the need for *a priori* or user-defined information.

4.4 Simulations and Discussion

The performance of the proposed IMM-KF ECG modeling method is demonstrated using real ECG signals obtained from the MIT-BIH arrhythmia database [89]. The sampling rate for the signals is 360 Hz. Preprocessing was carried out



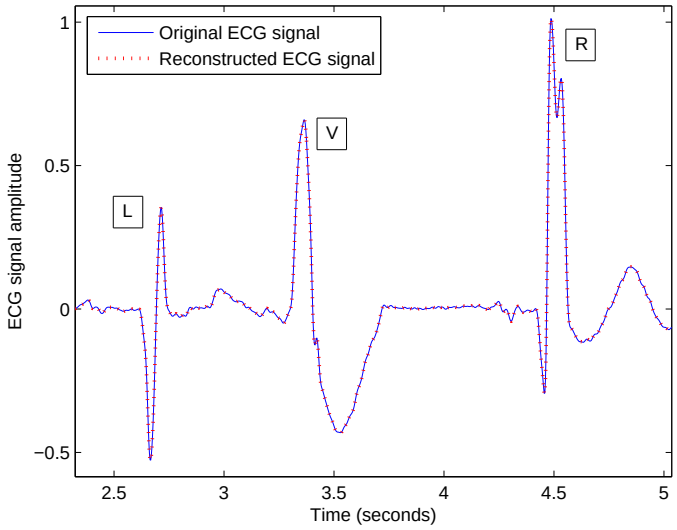
(a) N type ECG beat



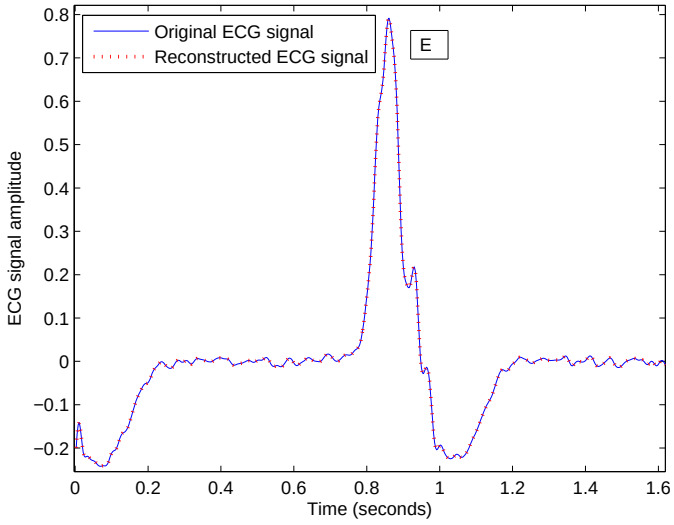
(b) N and V type ECG beats

Figure 4.1: Original and reconstructed ECG signals using the IMM-KF algorithm for different beat types. The letters in the square boxes are the beat labels.

to remove baseline wander and power-line interference with the help of a simple lowpass filter [45]. The IMM-KF algorithm is initialized with mode probabilities $\mu_0^{(i)} = 1/3$ $(i = 1, 2, 3)$, and transition probabilities $\pi_{ij} = 0.8$ for $i = j$ and $\pi_{ij} = 0.1$ for $i \neq j$ $(i, j = 1, 2, 3)$. The polynomial model coefficients were initialized using uniformly random values in $[-0.5, 1.5]$; it was observed that the



(a) L, V and R type ECG beats



(b) E type ECG beat

Figure 4.2: Original and reconstructed ECG signals using the IMM-KF algorithm for different beat types. The letters in the square boxes are the beat labels.

algorithm is not very sensitive to this choice.

Figures 4.1 and 4.2 shows plots of the real ECG signals along with the reconstruction using the parameter estimates from the proposed IMM-KF ECG model. The tracking performance of the algorithm is shown for different types of

ECG morphologies: Figure 4.1(a) for normal sinus rhythm (N) type ECG, Figure 4.1(b) for a premature ventricular contraction (PVC) ECG (MIT-BIH arrhythmia notation V), and Figures 4.2(a) and 4.2(b) for left bundle branch block (L), right bundle branch block (R), and ventricular escape (E) signals. For clarity, the plots show only a few seconds of the actual data duration. It can be seen that the tracker performs well and the reconstructed signals closely match the original ECG signals. In particular, from the tracking results for PVCs or V type beats, which occur intermittently within the ECG signal and are characterized by QRS complexes that do not have a fixed shape [72], it is seen that the proposed algorithm can track abrupt changes in morphology without having *a priori* information about the input data. The Gaussian model based algorithm presented in [24, 32] did not have this capability of tracking abrupt changes in morphology.

Chapter 5

ECG MODELING USING SEQUENTIAL MARKOV CHAIN MONTE CARLO METHOD WITH SIMULTANEOUS MODEL SELECTION

The IMM-KF method requires knowledge of the mode transition probabilities and its performance was found to be somewhat sensitive to those parameters. An alternate approach to this model, where the coefficients of the multi-mode polynomial representations vary slowly with every time step, is to model the polynomial coefficients as constant (static) over time segments of the ECG signals. Specifically, although ECG signals are dynamic in nature, they can be well represented using polynomial functions of different orders with coefficients that are constant over short time segments. So, instead of the transitioning modes utilized in the IMM framework, the model for the system (polynomials of different orders) can be selected for each segment. The system parameters can then be estimated using a particle algorithm framework provided by the SMCMC filter [36].

In this chapter, a multiple model setup for describing ECG signals is proposed using the SMCMC filter with simultaneous model selection using polynomial functions whose coefficients are assumed to be constant over certain segments of the signals, the lengths of which are determined adaptively. First, the concept of the SMCMC filter working with simultaneous model selection is described. The state-space model for ECG signals using the SMCMC filtering approach is given next, followed by the specific steps of the proposed algorithm. The effectiveness of the algorithm in tracking different types of ECG signal morphologies, the ability of the framework in adaptively selecting between multiple models and the superior performance of the multiple modeling framework are finally demonstrated with simulations using real ECG data.

5.1 SMCMC Filtering and Simultaneous Model Selection

In the particle filtering approach, the posterior pdf of the unknown state of the dynamical system is represented using particles and weights. However, particle algorithms were developed for systems where the parameters are time-varying and dynamic in nature, and can become unstable in the presence of static parameters [90]. To preserve the stability of the solution, ideas based on making a dynamic parameter assumption for static parameters have been suggested using improved mixing properties and kernel smoothing [90]. Other algorithms that do not depend upon a dynamic assumption for static parameters, by preserving the stability using complete rejuvenation of the particles, when required, have been discussed in [36, 91].

The algorithm of [36], referred to as the SMCMC filter, which combines sequential importance sampling (SIS) with Markov Chain Monte Carlo (MCMC) techniques, is utilized to estimate the parameters of the ECG signal model. The SIS is first used to process the measured ECG samples sequentially, after which a rejuvenation test is performed to check whether the SIS particles need rejuvenation. If rejuvenation is necessary, the particles are completely rejuvenated using MCMC methods. Thus, the three steps in the algorithm are: SIS, rejuvenation test, and MCMC.

The first step in the SMCMC filter is to propagate the weights of the particles representing the posterior distribution by incorporating the new incoming measurement using SIS. At time $k-1$, let the posterior distribution $p\left(\mathbf{x}|\mathbf{Z}_{k-1}\right)$ over the state vector $\mathbf{x}$ be represented by N_s particles $\mathbf{x}_1, \ldots, \mathbf{x}_{N_s}$ with weights $w_{k-1,1}, \ldots, w_{k-1,N_s}$. Since the unknown parameters are static and do not vary with time, we drop the time subscript for $\mathbf{x}$ during our discussion of the SMCMC filter, unless otherwise specified. Denoting the particle-weight pairs at time $k-1$

by $(\mathbf{x}_1, w_{k-1,1}), \ldots, (\mathbf{x}_{N_s}, w_{k-1,N_s})$, the updated posterior density at time k using the new measurement $\mathbf{z}_k$ is obtained by the weight update given as [36]

$$w_{k,j} \propto p(\mathbf{z}_k|\mathbf{x}_j, \mathbf{Z}_{k-1})\, w_{k-1,j}, \quad j = 1, \ldots, N_s. \tag{5.1}$$

Since the SIS uses a finite number of particles and the particle values are static, as new measurements get processed the weights lose diversity over time and only a very small percentage of weights remain non-zero. This causes the algorithm to become unstable. In order to counter this problem the SMCMC algorithm employs rejuvenation of particles whenever necessary. The rejuvenation test used here is based on the Kullback-Leibler (KL) divergence, which provides a measure of the distance between two probability distributions. Specifically, when each of the subsequent n measurements is obtained, the KL divergence between the posterior distributions $p(\mathbf{x}|\mathbf{Z}_k)$ and $p(\mathbf{x}|\mathbf{Z}_{k+n})$ is computed, and rejuvenation is performed if it exceeds a threshold. Since the particle-weight pairs representing these posterior pdfs are $(\mathbf{x}_1, w_{k,1}), \ldots, (\mathbf{x}_{N_s}, w_{k,N_s})$ and $(\mathbf{x}_1, w_{k+n,1}), \ldots, (\mathbf{x}_{N_s}, w_{k+n,N_s})$, respectively (the particles $\mathbf{x}_1, \ldots, \mathbf{x}_{N_s}$ do not change during the SIS step and the weights are updated according to (5.1)), the KL divergence is simply [36]

$$KL(w_{k+n}, w_k) = \sum_{j=1}^{N_s} w_{k+n,j} (\log w_{k+n,j} - \log w_{k,j}). \tag{5.2}$$

During rejuvenation, a new set of particles is generated using MCMC. The independent Metropolis-Hastings (IMH) is a popular MCMC method used for its simplicity. In the IMH method, if rejuvenation is to be performed at time $k+n$, a new set of particles representing the target density $p(\mathbf{x}|\mathbf{Z}_{k+n})$ is generated by sampling i.i.d. from a Gaussian proposal density $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\Sigma}_\mathbf{x})$ whose mean and covariance are computed using the most recent SIS particles and weights as [36]

$$\boldsymbol{\mu}_\mathbf{x} = \sum_{j=1}^{N_s} w_{k+n,j}\, \mathbf{x}_j, \tag{5.3}$$

$$\boldsymbol{\Sigma}_\mathbf{x} = \sum_{j=1}^{N_s} w_{k+n,j} (\mathbf{x}_j - \boldsymbol{\mu}_\mathbf{x})(\mathbf{x}_j - \boldsymbol{\mu}_\mathbf{x})^T. \tag{5.4}$$

The number of samples n processed between each rejuvenation is referred to as the batch size.

As for the IMM based ECG model of Chapter 4, linear, quadratic and cubic polynomials are used to model the ECG signals in the SMCMC algorithm. Model selection is carried out in conjunction with parameter estimation in the SMCMC filter framework. In general, assuming that there are M models $H_1, \ldots, H_M$, and denoting the parameter vector for model i as $\mathbf{x}^{(i)}$, the posterior density at time k is [36]

$$p(\mathbf{x}|\mathbf{Z}_k) = \sum_{i=1}^{M} P(H_i|\mathbf{Z}_k)\, p(\mathbf{x}^{(i)}|\mathbf{Z}_k, H_i). \tag{5.5}$$

The posterior pdf $p(\mathbf{x}^{(i)}|\mathbf{Z}_k, H_i)$ over the parameters can be obtained for each model H_i using SIS as described previously. The posterior model probability $P(H_i|\mathbf{Z}_k)$, which is used as a weight to determine the contribution of each model towards the final estimate, is calculated recursively as

$$P(H_i|\mathbf{Z}_k) \propto p(\mathbf{z}_k|\mathbf{Z}_{k-1}, H_i)\, P(H_i|\mathbf{Z}_{k-1}), \quad i = 1, \ldots, M. \tag{5.6}$$

The model likelihood $p(\mathbf{z}_k|\mathbf{Z}_{k-1}, H_i)$ is the expected likelihood with respect to $p(\mathbf{x}^{(i)}|\mathbf{Z}_{k-1}, H_i)$ and can be approximated as

$$p(\mathbf{z}_k|\mathbf{Z}_{k-1}, H_i) \approx \sum_{j=1}^{N_s^{(i)}} w_{k-1,j}^{(i)}\, p(\mathbf{z}_k|\mathbf{x}_j^{(i)}, \mathbf{Z}_{k-1}, H_i), \tag{5.7}$$

where $N_s^{(i)}$ is the number of particles used by model H_i, and $\mathbf{x}_j^{(i)}$ and $w_{k-1,j}^{(i)}$ are the particles and weights for model H_i at time $k-1$.

5.2 State-space Model for the SMCMC Filter with Simultaneous Model Selection

To formulate the ECG state-space model, the coefficients of the different order polynomial representations are assumed to be constant over time segments of the ECG signals. This is a viable assumption since certain segments of the ECG

signals, such as the P waves, PR segment, Q waves, R waves, S waves, ST segment and T waves, can be modeled using a single polynomial function, the order of which depends upon the signal morphology. Instead of delineating the ECG signal by using preprocessing algorithms, segments are used to define a time interval over which the ECG signal model parameters are assumed to be constant. In particular, the location of these segments is not specified *a priori* but rather determined adaptively based on the measured ECG data.

Based on the assumption that model parameters are constant within a given segment, the ECG signal is represented as

$$z_{k_l} = a_{l,0} + \sum_{m=1}^{M} a_{l,m}\, k_l^m + v_{k_l}. \tag{5.8}$$

Here, z_{k_l} is the k_lth ECG sample in the lth segment, with $k_l = 1, \ldots, N_l$, and N_l is the total number of samples in the lth segment. The measurement noise v_{k_l} is assumed to be white Gaussian with zero mean and variance R. The unknown static polynomial coefficients form the state vector $\mathbf{x}_l = [a_{l,0}\ a_{l,1}\ \ldots\ a_{l,M}]^T$ in the lth segment. In particular, under model H_i only the first $i+1$ elements of $\mathbf{x}_l$ are non-zero.

Since the model parameters are constant over a time segment, the state equation is given, for the lth segment, by

$$\mathbf{x}_{k_l} = \mathbf{x}_l. \tag{5.9}$$

To adaptively delineate the ECG signal into segments over which the model parameters can be assumed to be constant, the model likelihoods $p(H_i|\mathbf{Z}_k)$ are monitored. The model likelihoods give a measure of how well the models describe the given data, and small likelihood values are indicative of a poor fit. Thus, if at time k the likelihoods $p(H_i|\mathbf{Z}_k) = p(\mathbf{z}_k|\mathbf{Z}_{k-1}, H_i)\, p(H_i|\mathbf{Z}_{k-1})$ for all M models $H_1, \ldots, H_M$ fall below a threshold, a new segment is started at time $k+1$ and the

SMCMC particle filter algorithm is re-initialized. To obtain the noise-free reconstructed ECG signal for a segment, the combined polynomial coefficient estimate (5.5) is computed at final time in the segment and used with the measurement function in (5.8).

5.3 Framework for ECG Modeling with SMCMC Filtering and Simultaneous Model Selection

The suitable model for representing ECG data can change depending on the shape of the ECG signal in various time segments, such as the P wave, QRS complex, ST segment, etc. The proposed algorithm adaptively delineates the data into segments wherein the ECG signal parameters are assumed to be static under the SMCMC framework.

Due to the flexibility offered by the algorithm performing simultaneous model selection, ECG signals with different shapes can be modeled, including those with abrupt changes in morphologies. This avoids the issues faced in [24,32] due to the inability of the algorithm to track unexpectedly changing morphologies such as PVCs in the ECG. In addition, no preprocessing steps and *a priori* information is required for initialization of the algorithm. Delineation algorithms as used in [8, 9, 17, 19] are also not necessary.

Similar to the IMM-KF based algorithm proposed in Chapter 4, polynomial functions of three different orders, namely linear, quadratic, and cubic polynomials, are used to represent the ECG signals, setting the number of models to $M = 3$. We use the SMCMC algorithm with simultaneous model selection to obtain estimates of the polynomial coefficients. These estimates are used subsequently to perform classification of different types of ECG signals. As mentioned earlier, since classification is performed on a beat-to-beat basis, the ECG signals are divided into beats using the peak information of the beats provided by the MIT-BIH database.

To summarize, real ECG signals obtained from the MIT-BIH arrhythmia database [89] are first processed to remove recording artifacts such as baseline wander and power-line interference using lowpass filters as described in [45]. The measured signals are next separated into beats using the signal peak information and the phase of the signal as in [32]. Each ECG beat is then modeled using the three different order polynomials. The model parameters (polynomial coefficients) are estimated by the SMCMC algorithm that performs simultaneous model selection. The estimated parameters are subsequently used as features to classify between different types of ECG signals. The major steps in the SMCMC ECG modeling method are shown in Algorithm 3 below.

Algorithm 3 SMCMC ECG Modeling

1. Start a new segment l. Initialize the particles and weights for the posterior distribution over polynomial coefficients and the model probabilities for the $M = 3$ linear, quadratic, and cubic order polynomial ECG models.
2. For time $k_l = 1, 2, \ldots$, for the $M = 3$ polynomial models perform the following steps:
   (a) Using SIS ((5.1)), compute the weights at time k_l using the weights from time $k_l - 1$ and the ECG measurement at time k_l.
   (b) Calculate the KL divergence (5.2), and if it exceeds a threshold τ_1 perform rejuvenation of the particles.
   (c) Compute the posterior model probabilities at time k_l using the model probabilities from time $k_l - 1$ and the model likelihoods based on the ECG measurement at time k_l using (5.6)-(5.7).
   (d) Compare the model likelihoods to a threshold τ_2, and if they fall below the threshold, go to Step 3.
3. Obtain the noise-free reconstructed ECG signal for the lth segment using the combined polynomial coefficient estimate (5.5) at final time $k_l = N_l$ with the measurement function in (5.8). Go to Step 1.

5.4 Simulations and Discussion

The performance of the SMCMC ECG modeling method is demonstrated using real ECG signals from the MIT-BIH arrhythmia database [89]. All the signals are sampled at 360 Hz. Artifacts, such as baseline wander and power-line interference, are removed using a lowpass filter [45]. In the SMCMC filter, the particles representing the polynomial model coefficients in a segment are initialized using values from local linear, quadratic, and cubic polynomial approximation at the start of the segment. The threshold $\tau_1 = 0.1$ is used in the KL divergence based rejuvenation test, and the threshold $\tau_2 = 10^{-10}$ is used with the model likelihoods for defining the start of a new segment. The measurement noise variance R is chosen to be the same for all the models and is set equal to 10^{-4}. It must be noted that the value of R should be set such it always exceeds the amount of noise present in the signal, since otherwise it can leading to large tracking errors.

The prior model probabilities were set to $P(H_i|\mathbf{Z}_0) = \{0.9, 0.09, 0.01\}$ for $i = 1, 2, 3$ respectively. The higher order polynomial models can always track the signals corresponding to the lower order polynomial models, using small higher order coefficients. For example, a quadratic polynomial can be modeled using a cubic function provided the value of the highest order coefficient of the cubic function is close to zero. In such situations, it is natural to favor a lower order representation, and thus larger prior probabilities are assigned to the lower order models. This is demonstrated in Figure 5.1 using a simple example. 500 samples of a synthetic signal that corresponds to a quadratic polynomial with a sampling frequency of 1000 Hz are generated. The samples are generated in the time range 0.1 s to 0.6 s. Figure 5.1(a) shows the noisy data used as the algorithm input and the reconstructed signal that was estimated using the parameter estimates from the algorithm, which is seen to be very close to the noise-free data. The data is

not adaptively segmented for this case since a static parameter model is valid for the entire time range. The top plot in Figure 5.1(b) shows the model probabilities evaluated at each time step when the prior model probabilities are all assumed to be equal. It can be seen that the model probabilities for the quadratic and cubic models are almost equal, since both models can represent the data with sufficient accuracy. This means that an almost equal weighting is given to the estimates from both these models (even though the noisy data originated from a quadratic model) when calculating the final estimate using (5.5). This problem can be prevented by using unequal prior model probabilities, for example, using values $P\left(H_i|\mathbf{Z}_0\right) = \{0.9, 0.09, 0.01\}$ as described previously, which helps in giving higher preference to the lower order quadratic model over the cubic one. The model probabilities evaluated using this biasing scheme are shown in the bottom plot of Figure 5.1(b), from which it can be easily seen that the (simpler) quadratic model achieves a higher probability over the cubic model.

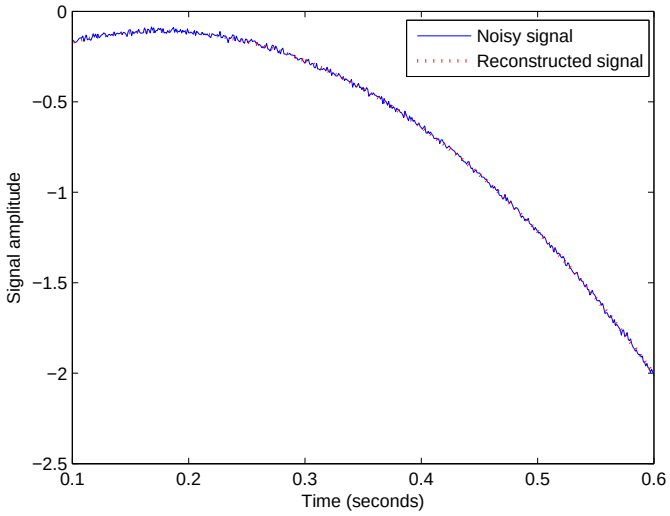
However, if the number of data samples is sufficiently large, the biasing scheme for the initial model probabilities is not necessary, since the algorithm will have enough number of samples to infer the correct model from the data. This is demonstrated in Figure 5.2. A synthetic quadratic polynomial is generated using parameters similar to those used to generate the signal in Figure 5.1(a), but with 2000 samples, so that the data now spans between 0.1 to 1.1 s. Figure 5.2(a) shows the noisy data and the reconstructed signal estimated using the algorithm parameter estimates. The model probabilities evaluated at each time step using equal prior model probabilities and unequal prior model probabilities with the biasing scheme are shown in the top and bottom plot, respectively in Figure 5.2(b). It is seen that in both cases the algorithm correctly converges to the quadratic model, since the quadratic model achieves the highest probability. It must be noted that in the case with equal prior model probabilities, the algorithm

converges to the correct model after a few samples have elapsed, but converges quickly when the prior model probabilities are unequally weighed.

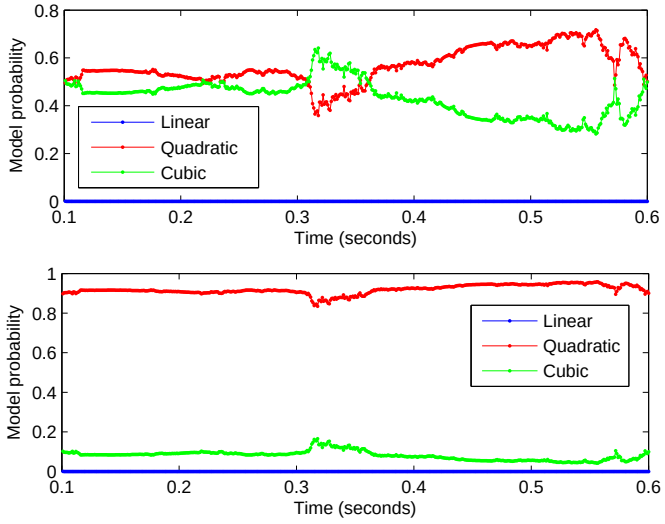
Since ECG signals are usually sampled at smaller sampling rates, the number of samples available in each adaptively selected segment is not sufficient for the algorithm to infer the correct model. Thus, for demonstrating the results of the SMCMC algorithm using real ECG signals, the biased weighing scheme for prior model probabilities with $P(H_i|\mathbf{Z}_0) = \{0.9, 0.09, 0.01\}$ for $i = 1, 2, 3$ is used.

Figures 5.3 and 5.4 demonstrates the tracking capability of the algorithm. Figures 5.3(a), 5.3(b), 5.4(a), and 5.4(b), show plots of the real ECG signals and the model reconstruction for N, L, V, R, E, and j type beats. It can be seen that the algorithm performs well in tracking ECG signals of different morphologies (the reconstructed signals are very close to the original ECG signals). In addition, the proposed algorithm can easily track (without requiring *a priori* information) ECG signals with abrupt changes in morphology, eg. PVCs which occur abruptly in between beats of different types and were not tracked using the Gaussian model for ECG signals presented in [24, 32].

Figure 5.5 illustrates the model selection capability of the algorithm using N and L type beats. In order to provide a clear illustration of the results only a single typical beat for each ECG signal type is shown. Figures 5.5(a) and 5.5(b) show the original and reconstructed N and L type ECG beats, respectively. The black asterisks indicate the end-times for the adaptive segments over which the model parameters were assumed static. Figures 5.5(c) and 5.5(d) show the selected models (linear, quadratic, or cubic polynomial models) for the segments of the N and L type ECG beats, respectively. Here, the model numbers 1, 2, and 3, represent linear, quadratic, and cubic order polynomials, as described in Section 5.2. In particular, the selected models for a segment are those models that have the highest model probability at the end of the segment. It can be



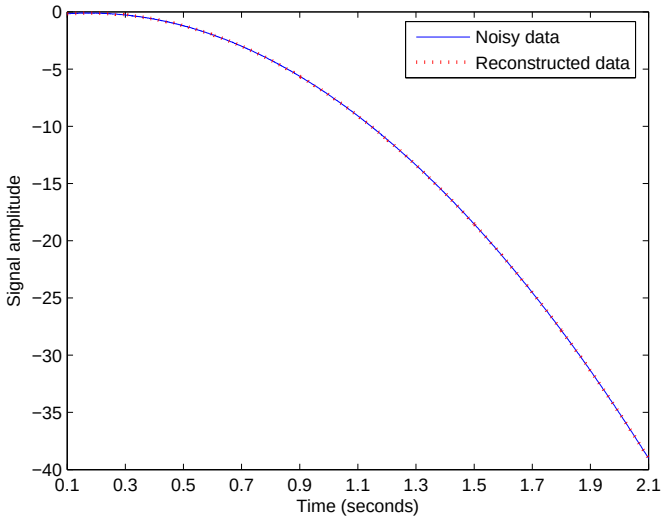
(a)



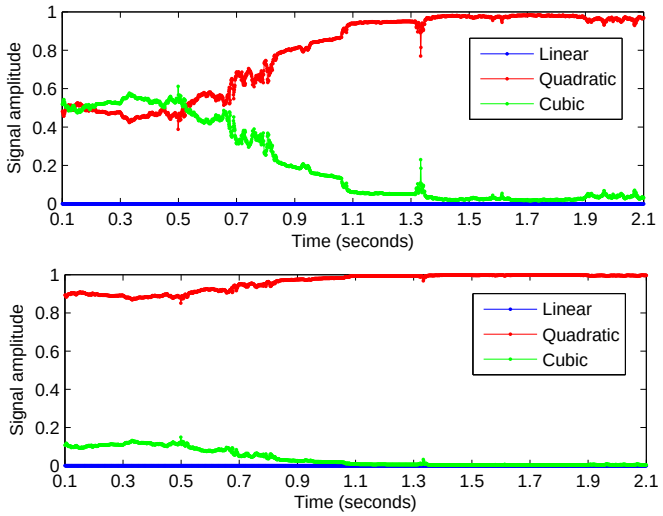
(b)

Figure 5.1: (a) Synthetically generated noisy signal from a quadratic model and the reconstructed signal with 500 data samples; (b) Model probabilities evaluated using equal prior model probabilities (top) and unequal prior model probabilities (bottom) using SMCMC algorithm with 500 data samples.

seen from the plots that the algorithm can adaptively select between the different polynomial models based on the morphology of the ECG signal at different times. Figures 5.5(e) and 5.5(f) show the probabilities of the linear, quadratic, and cubic



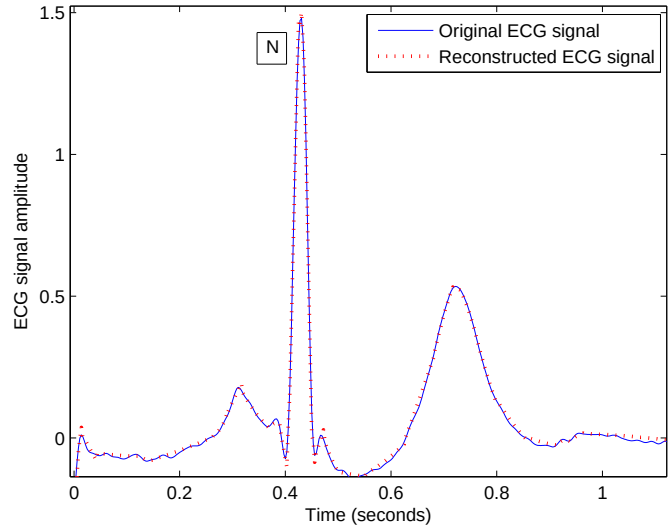
(a)



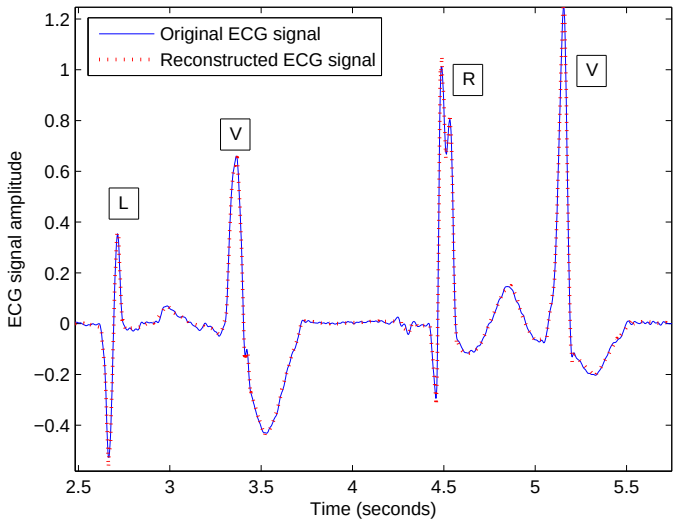
(b)

Figure 5.2: (a) Synthetically generated noisy signal from a quadratic model and the reconstructed signal with 2000 data samples; (b) Model probabilities evaluated using equal prior model probabilities (top) and unequal prior model probabilities (bottom) using SMCMC algorithm with 2000 data samples.

order polynomial models at the end of each segment. The plots show that in most cases the algorithm converges to a specific model by the end of a segment (the probability of one model is much higher than the others).



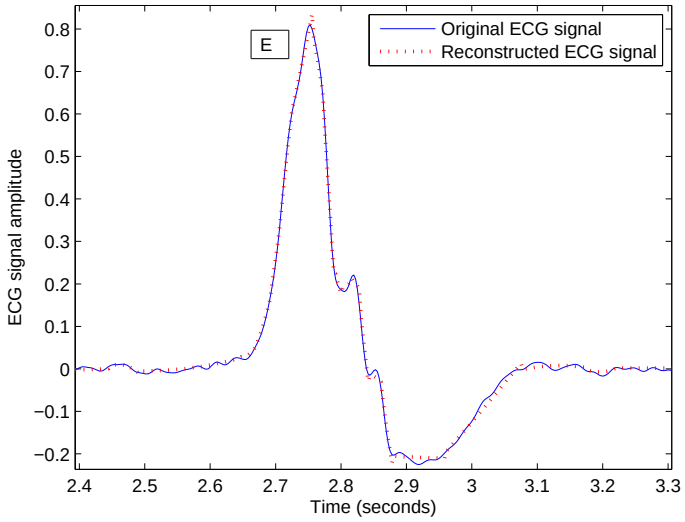
(a) N type beat



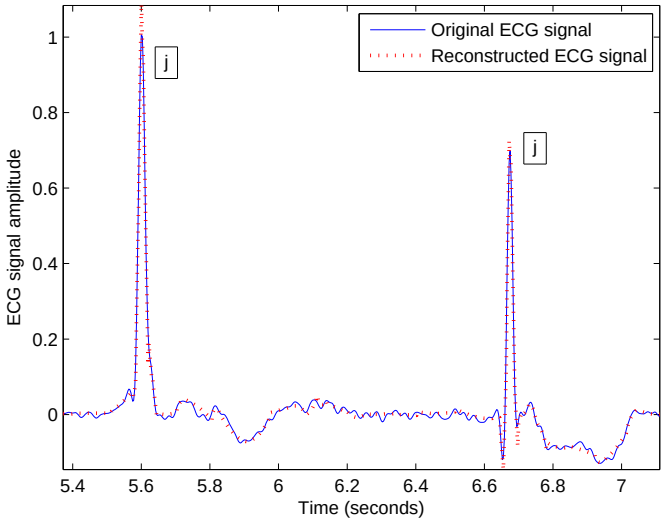
(b) L, V and R type beats

Figure 5.3: Original and reconstructed ECG signals from the SMCMC filtering algorithm with simultaneous model selection for different beat types. The letters in the square boxes are the beat labels.

A key advantage of our algorithm is its ability to track ECG signals of different and abruptly changing morphologies without requiring prior knowledge about the data. We demonstrate the superior performance of the proposed method by comparing its tracking capability and reconstruction root mean-squared error



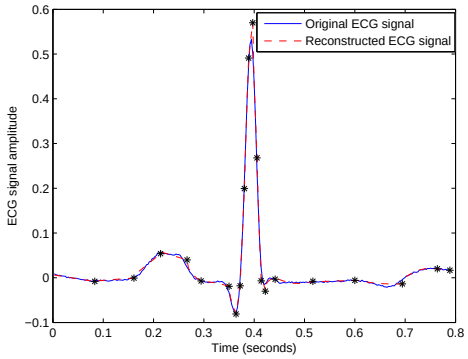
(a) E type beat



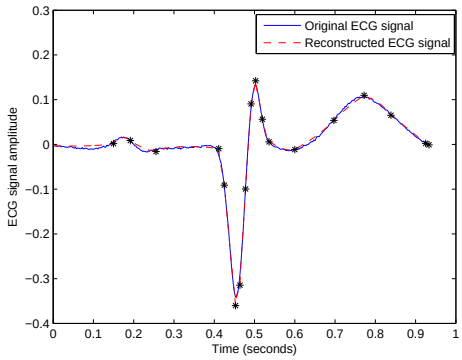
(b) j type beats

Figure 5.4: Original and reconstructed ECG signals from the SMCMC filtering algorithm with simultaneous model selection for different beat types. The letters in the square boxes are the beat labels.

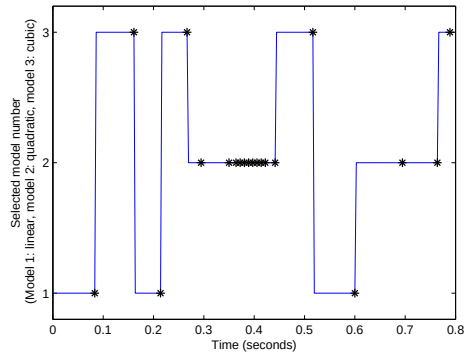
(RMSE) to that obtained from the nonlinear Bayesian framework for modeling ECG signals using Gaussian functions [24,32]. The Gaussian ECG model is chosen for comparison since it is based upon a similar statistical framework. For both methods, the reconstruction RMSE is calculated using Monte Carlo simulation,



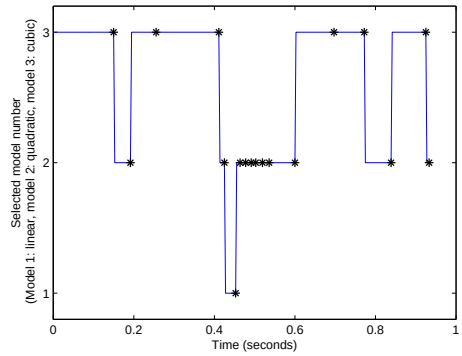
(a) Original and reconstructed ECG signal



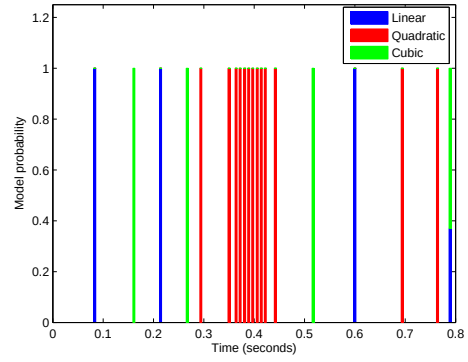
(b) Original and reconstructed ECG signal



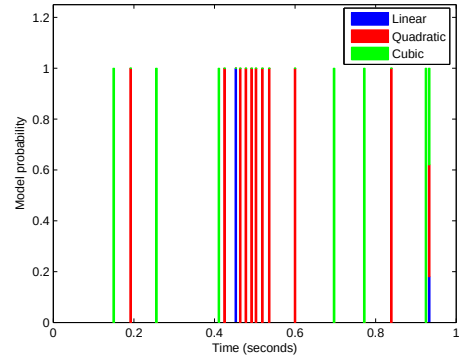
(c) Selected models for the segments



(d) Selected models for the segments



(e) Model probabilities for the segments



(f) Model probabilities for the segments

Figure 5.5: Illustration of model selection using typical ECG beats. The black asterisks indicate the end-times of the adaptive segments over which the model parameters were assumed static. (a), (c) and (e) N type ECG beat; (b), (d) and (f) L type ECG beat.

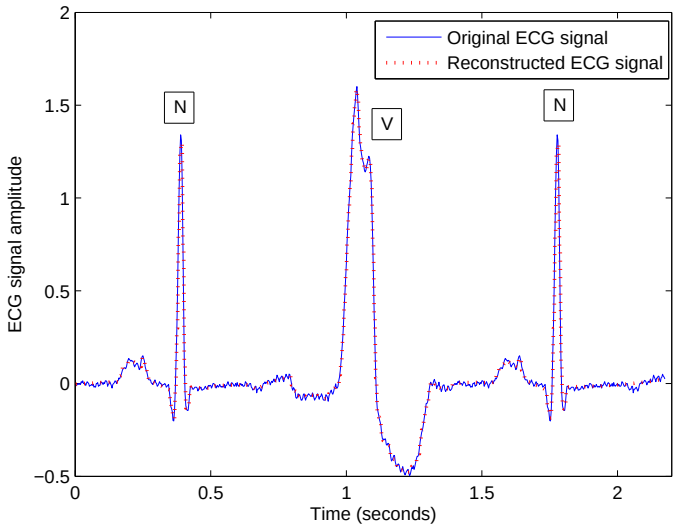
and is expressed as

$$\text{RMSE}_k = \sqrt{\frac{1}{N_r}\sum_{r=1}^{N_r}(\tilde{z}_k - \hat{z}_{k,r})^2}, \tag{5.10}$$

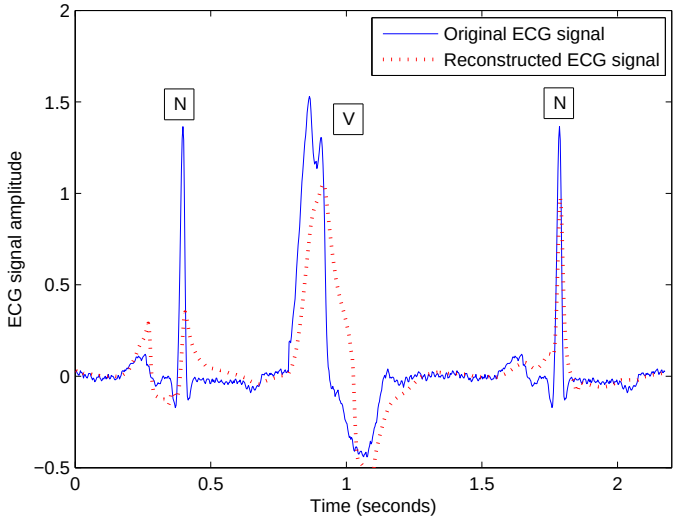
where $\tilde{z}_k$ denotes the noise-free reference ECG signal at time k, $\hat{z}_{k,r}$ is the reconstructed ECG signal at time k for the rth Monte Carlo run, and N_r is the number of Monte Carlo runs. The noise-free reference ECG signal is obtained by averaging noisy real ECG beats of the respective type from the MIT-BIH arrhythmia database [89] (assuming that the noise in the ECG signals is additive and independent, the SNR increases as more signals are averaged [92]). To demonstrate the ability of our algorithm to track abruptly changing morphologies, we created a typical example of such morphologies with the averaged noisy real ECG beats of N and V types. The N_r ECG signals used for modeling and reconstruction were also obtained from the MIT-BIH arrhythmia database. Also, the RMSE defined in (5.10) is the square root of a sample average and different from the MSE of (3.3) which uses a time average.

Figure 5.6 shows plots of the and reconstructed ECG signals using estimates from the SMCMC algorithm with simultaneous model selection and the Gaussian ECG model, respectively, for a typical run. As seen from Figure 5.6(a), the SMCMC filtering approach leverages its multiple model flexibility to track the different morphologies in the data without need for pre-processing. In contrast, from Figure 5.6(b) it can be observed that the Gaussian approach does not track the ECG data well and misses some of the fiducial points, such as the QRS complex of the first N type beat, among others. As mentioned previously, PVC beats occur randomly within ECG beats of different types and these abrupt changes in morphology were not tracked using the Gaussian model in [24, 32] which uses a phase-wrapping method to generate the initial filter estimates.

Table 7.1 shows the average RMSE between the noise-free reference and



(a) SMCMC ECG modeling method with simultaneous model selection



(b) Gaussian ECG model [24, 32]

Figure 5.6: Original and reconstructed ECG signal from a typical Monte Carlo run.

ECG signals reconstructed using our proposed SMCMC and IMM-KF multiple model frameworks, Gaussian ECG model [24, 32] and the SMCMC framework when only a single fixed polynomial model is used. The fixed-order polynomial cases are linear (M1), quadratic (M2) and cubic (M3) polynomials. The RMSE values are calculated using (5.10) with $N_r = 500$ Monte Carlo runs and averaged

Table 5.1: Comparison of average RMSE (x 10^{-2}).

Time (seconds)	SMCMC	IMM-KF	Gaussian model	SMCMC		
				M1	M2	M3
0.15-0.28	2.71	2.29	6.71	3.16	3.41	3.18
0.28-0.3	2.52	1.82	8.03	2.01	2.60	2.69
0.33-0.43	22.22	22.38	25.38	23.59	23.67	23.67
0.43-0.60	1.57	1.94	7.08	2.42	2.65	2.59
0.60-0.75	1.89	1.77	2.61	2.32	2.49	2.48
0.90-1.00	28.80	27.65	35.27	36.93	36.83	36.81
1.00-1.30	22.57	22.29	39.39	32.22	32.12	32.12
1.50-1.63	1.79	1.68	2.58	2.33	2.78	2.53
1.63-1.68	3.09	2.97	3.59	3.78	3.94	3.86
1.68-1.78	10.46	10.01	13.36	11.06	11.36	11.40
1.78-1.95	8.66	9.35	11.05	9.89	10.04	10.02
1.95-2.10	1.59	1.62	2.53	2.09	2.35	2.27

over certain segments of the signal corresponding to the times indicated in the first column of Table 7.1 (for example, the time 0.15-0.28 s corresponds to time range of the P wave in the first N type beat of the signal). It can be observed from Table 7.1 that both the SMCMC and IMM-KF algorithms outperform the Gaussian method by achieving a smaller reconstruction RMSE. This demonstrates that the use of multiple models to track ECG signals can be highly advantageous, as the resulting parameter estimation is adaptive to changes in morphology and consequently more accurate. This is important in clinical settings because the accuracy of the parameter estimation algorithm can greatly affect the outcome of cardiac disease diagnosis. The advantage of using multiple models is further substantiated due to the fact that the RMSE performance of the SMCMC and IMM-KF algorithms using multiple polynomial orders is also observed to be superior to that of fixed-order polynomial ECG modeling (M1, M2 and M3) with the SMCMC framework as seen in Table 7.1. Similar results were obtained showing a better performance for the multiple model algorithms when compared to the fixed-order polynomial ECG modeling with the IMM-KF framework.

Chapter 6

ECG ARRHYTHMIA CLASSIFICATION USING BAYESIAN MAXIMUM LIKELIHOOD CLASSIFIER

ECG signals represent recordings of the electrical activity of the human heart. Any alterations in the normal electrical pattern of the heart helps in diagnosing the presence of cardiac abnormalities. Thus, the recorded ECG signals are a useful diagnostic tool for identifying and analyzing the condition of an individual's heart. Automatic classification or screening of ECG signals is of great value to medical practitioners since it can alleviate the potentially painstaking process of performing this task manually. As a result, much research has been devoted to this area, within the sphere of signal processing techniques. A number of these techniques were discussed in Section 1.3

In this work, ECG signal classification is first performed for five different types of ECG signals using the estimated parameters from the proposed models as features in a simple Bayes ML classifier. This is to demonstrate the effectiveness of the model parameters in being able to differentiate between different ECG classes. In the next chapter, a patient-specific classifier based on Bayesian nonparametric methods is proposed. In this chapter, a short insight into the generation of ECG signals and the causes of arrhythmia is outlined first. Later, the Bayes ML classifier is described. Finally, the features extracted from each of the three proposed ECG model parameters are explained and the performance of the Bayes ML classifier using these features is demonstrated for different types of arrhythmia.

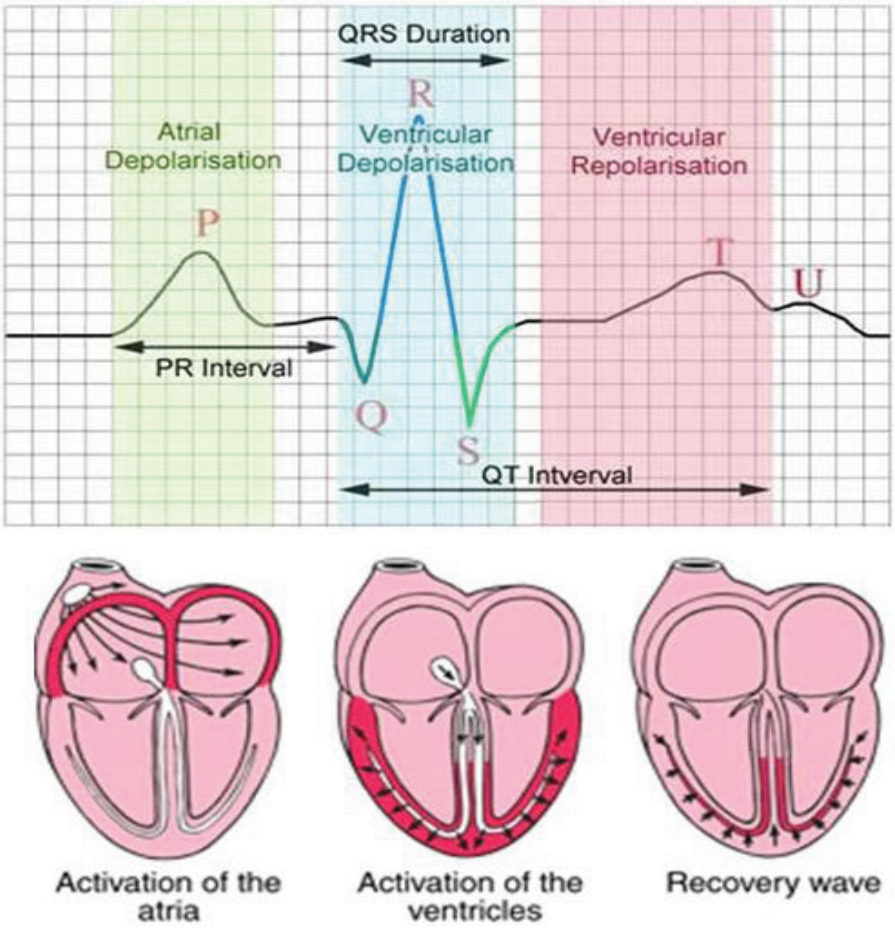


Figure 6.1: Depiction of electrical activity in the heart for a single cardiac cycle (ECG beat) [1].

The electrical activity of the heart during each ECG beat begins with the depolarization of the pacemaker cells in the sinus node located in the top right portion of the heart. Depolarization represents the loss of inherent negative polarity of the cardiac cells. The depolarization and subsequent electrical activity continues propagating from cell to cell producing a depolarization wave that proliferates through the atria, which when recorded using electrodes placed on the surface of the body displays a burst of electrical activity known as the P wave. The depolar-

ization wave passes through the atrio-ventricular node that causes the electrical activity to pause for a fraction of a second, thus giving rise the flat PR segment in the ECG recording. Finally, the wave of depolarization reaches the ventricles during which the QRS complex is recorded. As soon as the depolarization is complete in each portion of the heart, cardiac cells restore their polarity and the repolarization wave starts propagating in the same direction, repolarizing the cells. The repolarization of the atria is masked by ventricular depolarization and is thus not recorded on the ECG. However, the repolarization of the much larger ventricles is recorded as the T wave of the ECG signal. All this activity represents one cardiac cycle of the heart otherwise known as an ECG beat and is depicted in Figure 6.1.

Any disturbance in the rate, regularity, site of origin or conduction of the cardiac electrical activity is termed as an arrhythmia [72]. There are several types of arrhythmias ranging from a relatively benign atrial fibrillation to a serious life-threatening condition such as ventricular tachycardia which could result in sudden cardiac death. A lot of research has been performed to automate the detection of life-threatening conditions such as ventricular fibrillation and tachycardia [93–95]. In this work, the major focus is on arrhythmias that are not immediately life-threatening, but could turn so if not closely monitored and treated.

6.2 Bayes Maximum-Likelihood Classification Method

The Bayes ML classifier is a supervised learning technique that works as follows. Given a feature vector $\mathbf{y}$, of dimension $N_{\mathbf{y}}$, the Bayes ML classifier calculates and ranks the likelihood $p(\mathbf{y}|C_q)$ of the feature vector conditioned on each of the considered classes C_q, which are assumed here to follow multivariate Gaussian distributions. Considering N_q possible classes, the likelihood of feature vector $\mathbf{y}$

in class C_q is given by

$$p(\mathbf{y}|C_q) = \frac{1}{(2\pi)^{\frac{N_\mathbf{y}}{2}} |\mathbf{\Sigma}_q|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}_q)^T \mathbf{\Sigma}_q^{-1} (\mathbf{y}-\boldsymbol{\mu}_q)}, \tag{6.1}$$

for $q = 1, \ldots, N_q$, where $\boldsymbol{\mu}_q$ and $\mathbf{\Sigma}_q$ are the mean and covariance of the Gaussian model for class C_q. The means and covariances are determined using a set of training feature vectors from each class. For classification of a given test feature vector $\mathbf{y}$, the likelihood function in (6.1) is evaluated for each class C_q using the corresponding mean and covariance. Equal prior probabilities are assumed for all the classes, and the classifier output is the class C^* that maximizes the log-likelihood, i.e.,

$$C^* = \arg\max_q \ \log p(\mathbf{y}|C_q). \tag{6.2}$$

The N_q classes here correspond to various ECG signal types, and the feature vector $\mathbf{y}$ is formed using the appropriate parameter estimates that are unique to each class, based on the proposed ECG models.

6.3 Classification with the Multi-harmonic ECG Model

Using the parameters from the adaptive multi-harmonic ECG model, four types of ECG signals are classified. These were ECG signals with normal sinus rhythm (Class C_1) and signals with three different types of arrhythmia [72], namely supraventricular arrhythmia (arrhythmia arising in the upper chambers of the hearts above the ventricles, i.e., either in the atria or the atrio-ventricular node [72], denoted as Class C_2), malignant ventricular arrhythmia (arrhythmia that originates in the ventricles and can potentially lead to cardiac arrest and hemodynamic collapse [72], denoted as Class C_3), and atrial fibrillation (arrhythmia arising due to chaotic atrial activity leading to an irregular ventricular rate [72], denoted as Class C_4). Thus for this case the number of classes is $N_q = 4$.

In order to perform classification utilizing the estimated parameters from the ECG model with multiharmonic components,the adaptively selected best ECG

Table 6.1: Confusion matrix showing classification results using the multi-harmonic model with adaptive parameter estimation.

Class	C_1	C_2	C_3	C_4
C_1	0.89	0	0	0.11
C_2	0.06	0.83	0.06	0.06
C_3	0	0	0.89	0.11
C_4	0	0	0.11	0.89

parameter estimates are used. The best estimates correspond to the parameter pair representing the number of harmonics and mean cardiac frequency that minimize the reconstruction MSE. The instantaneous cardiac frequency estimates (given by f_k in (3.2)) are used to formulate the feature vector. To classify the four ECG signal classes, only six f_k estimates from around the QRS region are used. This limits the feature vector dimensionality to $N_{\mathbf{y}} = 6$.

6.3.1 Classification Results

Table 6.1 shows the classification results in the form of a confusion matrix. The (i, j)th entry of the confusion matrix shows the fraction of signals in class C_i that are classified to class C_j. Thus, the diagonal entries of the confusion matrix give the correct classification rates and the off-diagonal entries indicate the misclassification rates. It can be seen that the classifier performs fairly well, with a correct classification rate of nearly 0.9 (90%) for three of the four classes used.

6.4 Classification with the IMM-KF ECG Model

Five specific types of ECG signals for classification and comparison, are considered here thus providing $N_q = 5$ classes. These are the normal sinus rhythm beats (N), two types of conduction block arrhythmias (caused due to unexpected delays in propagation of the normal electrical activity originating in the sinus node [72]), namely the left bundle branch block (L) and right bundle branch block (R), and two types of escape rhythm arrhythmias (caused when electrical activity originates in locations other than the sinus node [72]), namely the ventricular escape beat

(E) and junctional or nodal escape beat (j). The letters used to denote each class are the standard arrhythmia notation used by the MIT-BIH arrhythmia database [89]. It should be noted that the conduction block arrhythmias of types L and R are characterized by inverted R waves and presence of double R waves (also known as RSR' waves or "rabbit ears") respectively, in lead V1 of the ECG lead configuration [72]. In addition, since escape beats originate in locations other than the sinus node, the P wave is usually absent in these signals [72]. The features were designed for our classifier based on these characteristics of the chosen types of ECG signals.

The estimated parameters of the IMM-KF ECG model are used to define the features used for performing ECG signal classification. In the proposed model, at each time k, a total of $M+1$ (combined) polynomial coefficient estimates are available. In order to limit the feature dimensionality and reduce complexity, for the purpose of classification we only use the estimated $a_{k,1}$ coefficient from the state vector, which was found to be most distinctive for classifying the chosen types of ECG signals. Additionally, in order to further reduce the feature vector size, only six features are computed ($N_{\mathbf{y}} = 6$) as follows. The first five features are obtained from average estimates of $a_{k,1}$ at five regions in the QRS complex (local averages are used for robustness to abrupt signal changes). The sixth feature is obtained from the mean of the $a_{k,1}$ coefficients in the P wave. Note that we do not perform QRS complex detection or wave delineation to find the location of the P wave. Instead, the signal peak information that was employed to perform beat separation is used. The position of the QRS complex and the P wave are roughly estimated based on peak information by noting that the QRS interval is about 70 ms to 120 ms long, with a PR interval of 0.12 s to 0.2 s occurring before the QRS interval [96].

Table 6.2: Confusion matrix showing classification results using the IMM-KF ECG model.

Beat type	N	L	R	E	j
N	0.98	0	0	0	0.02
L	0.01	0.99	0	0	0
R	0.02	0	0.98	0	0
E	0	0	0.02	0.98	0
j	0.01	0	0	0	0.98

6.4.1 Classification Results

Table 6.2 shows the classification results in the form of a confusion matrix. It can be seen that the classifier performs very well, achieving an average correct classification rate of 0.98 (98%).

6.5 Classification with the SMCMC ECG Model

For classification the same five types of ECG signals used by the classifier for the IMM-KF model, namely: normal sinus rhythm (N) signals, left bundle branch block (L), right bundle branch block (R), ventricular escape (E), and junctional or nodal escape (j) beats are used. This makes the number of classes $N_q = 5$.

The noise-free reconstructed ECG signal obtained using the estimated polynomial coefficients parameters from the SMCMC filter with simultaneous model selection is used to derive features for classification of the ECG signals. As before, the feature vector dimension is limited to $N_{\mathbf{y}} = 6$. The first five features are obtained from averages of the reconstructed measurements at five regions in the QRS complex. The sixth feature is obtained from the mean of the reconstruction for the P wave.

6.5.1 Classification Results

Table 6.3 shows the classification results in a confusion matrix. It can be seen that the classifier performs very well, with an average correct classification rate

Table 6.3: Confusion matrix showing classification results using the SMCMC ECG model.

Beat type	N	L	R	E	j
N	0.98	0	0	0	0.02
L	0	0.99	0	0.01	0
R	0	0	0.98	0.02	0
E	0.02	0	0	0.98	0
j	0.01	0	0	0	0.99

of 0.98 (98%).

Table 6.4: Comparison of classification results.

Beat type	Correct classification rate			
	[8]	[9]	IMM-KF	SMCMC
N	0.98	0.98	0.98	0.98
L	0.97	0.97	0.99	0.99
R	0.94	0.99	0.98	0.98
E	0.90	0.96	0.98	0.98
j	–	0.91	0.98	0.99

The classification results from the IMM-KF and SMCMC ECG models are compared with those presented in [8, 9] using fuzzy-hybrid neural networks and support vector machines, respectively. These works were chosen for comparison as they investigated the same types of arrhythmias for classification. Therein, ECG signal delineation was first used to detect the QRS complex, followed by Hermite polynomial fitting. The feature vector consisted of about 17 parameters (15 Hermite polynomial coefficients and some temporal features of the actual QRS data). Table 6.4 shows a comparison of the classification performance. It can be seen that our classification results using the proposed multiple model methods compare favorably with those results, despite the use of a fairly small feature set. In particular, our results for the correct classification rates of the nodal escape (j) type beats are considerably better because our features include information about the P wave, which is absent in these beats [72]. This feature was not considered in [8, 9].

Chapter 7

PATIENT-SPECIFIC ECG ARRHYTHMIA CLUSTERING USING BAYESIAN NON-PARAMETRIC METHODS

As described in Chapter 1, the challenges faced by ECG classification algorithms include having the ability to classify between ECG signals that exhibit large variations of morphology, while preserving the inter-patient variability, and the non-availability of large amounts of training and testing data to validate the algorithms in a clinical setting. Although a number of algorithms based on supervised learning methods were previously designed, these fail to account for variations between different individuals, and hence a lot of focus is being laid on patient-specific techniques that preserve the inter-patient differences in ECG morphology. But patient-specific ECG classifiers [46, 50] based on global and local classifiers and patient-adaptable neural networks [42, 57, 67] still require training data to train the global classifiers and the neural networks, respectively. Although unsupervised learning algorithms [43, 49, 52, 60] based on clustering different types of ECG beats provide a good means of retaining patient-specific information and do not require training and testing data, existing algorithms are based on use of finite number of clusters. Since ECG data is constantly evolving, this might lead to inadequate number of clusters.

In order to address the challenges faced by ECG classification algorithms in general, and the drawbacks of the aforementioned works, an adaptive learning ECG classification method that is based on a Bayesian nonparametric method [74, 97] is proposed in this work. The adaptability of the algorithm to several classes of diseases and changes in morphology is achieved because Bayesian nonparametric methods place no restrictions on the models and their parameters, as well as on the number of classes or clusters that the data might belong to. The adaptive learning framework based on the Dirichlet process (DP) [98] is used in

this work. To be specific, features of the ECG signal are modeled using a DP mixture model that does not set a limit on the number of mixture components or clusters in the ECG data. The clusters identified in the ECG data can be labeled by an expert, or alternatively, an automated method can be used if sufficient training data is available. The results of the proposed clustering algorithm based on the DP framework are first validated assuming the cluster labels are available and have been provided by an expert. In addition, in order to make cluster identification automatic, the Bayesian ML method is used to identify the cluster labels, assuming the availability of training data.

The DP adaptive learning framework is first described in this chapter along with the blocked Gibbs sampling technique that is used to estimate the mixture parameters. The features selected as the input data to the DP algorithm are described next. The results of the clustering algorithm using real ECG data are provided assuming access to cluster labels provided by an expert. Finally, the use of the Bayesian ML method in automatically providing the cluster labels is demonstrated.

7.1 Dirichlet Process Mixture Modeling

Nonparametric Bayesian methods offer flexibility in representing data using models that can have infinite number of parameters. The DP is one such method, which works by placing prior distributions on parameters [74, 97]. These prior distributions are combined with the data likelihood to obtain a mixture model with infinite number of mixture components (clusters), which is known as the DP mixture model [99].

Because of its highly flexible modeling properties, the DP has provided a natural framework for application to problems such as speaker diarization [100], music analysis [101], protein modeling [102] etc. The advantages offered by the DP GMM when applied to the problem of ECG signal classification include no

requirement of information about the clusters and no restrictions on their number, in addition to requiring no separate testing and training datasets, since the clusters are automatically learned over time from the data.

7.1.1 DP Model Formulation

In this work, a given feature vector $\mathbf{y}$, corresponding to features from a single ECG beat as described in Section 7.2 is modeled using the DP mixture model as

$$p\left(\mathbf{y}|\mathbf{w},\boldsymbol{\Theta}\right)=\sum_{m=1}^{M}w_m\,p\left(\mathbf{y}|\boldsymbol{\theta}_m\right). \tag{7.1}$$

In the above equation, $\mathbf{w}=\{w_1,\ldots,w_M\}$ are the mixture weights that sum to one, $\boldsymbol{\Theta}=\{\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_M\}$ is the set of parameters characterizing the clusters, $p\left(\mathbf{y}|\boldsymbol{\theta}_m\right)$ represents the pdf parametrized by $\boldsymbol{\theta}_m$, and M is the maximum number of mixture components. For any given dataset or set of feature vectors $\mathbf{Y}=\{\mathbf{y}_1,\ldots,\mathbf{y}_N\}$, where N is the number of feature vectors, the effective number of mixture components ($\leq M$), their mixture weights $\mathbf{w}$, and parameters $\boldsymbol{\Theta}$ have to be estimated.

The mixtures used for the DP in this work are Gaussian mixtures, wherein the parameters of each model are the mean and covariance of the corresponding Gaussian. This is known as the DP Gaussian mixture modeling method (GMM), and is referred to as the DP GMM. For a DP GMM, the pdf $p\left(\mathbf{y}|\boldsymbol{\theta}_m\right)$ of the mth mixture component is specified by a Gaussian distribution with the parameters representing the mean and covariance of the Gaussian, i.e., $\boldsymbol{\theta}_m=\{\boldsymbol{\mu}_m,\boldsymbol{\Sigma}_m\}$. So, for an $N_{\mathbf{y}}$-dimension feature vector $\mathbf{y}$, the pdf for the mth mixture component is, $p\left(\mathbf{y}|\boldsymbol{\theta}_m\right)\triangleq\mathcal{N}\left(\mathbf{y};\boldsymbol{\mu}_m,\boldsymbol{\Sigma}_m\right)$. Rewriting (7.1) using the precision $\boldsymbol{\Sigma}_m^{-1}$ for each mixture component instead of the covariance gives,

$$p\left(\mathbf{y}|\mathbf{w},\boldsymbol{\mu},\boldsymbol{\Sigma}^{-1}\right)=\sum_{m=1}^{M}w_m\,\mathcal{N}\left(\mathbf{y};\boldsymbol{\mu}_m,\boldsymbol{\Sigma}_m^{-1}\right). \tag{7.2}$$

If the number of mixture components is known, ML techniques such as the EM algorithm [103, 104] can be utilized to estimate the parameters by maximizing the likelihood. In particular, the EM algorithm was employed in [60, 73] to determine the parameters of the GMM used to represent ECG signal features. However, if the number of mixture components is unknown, Bayesian nonparametric methods using the DP framework provide a good approach to representing such models.

The DP represents a distribution over another base distribution. It is characterized by two parameters, namely the scalar concentration parameter α and the base distribution G_0. Any draw from a DP is an almost surely discrete distribution represented by [97],

$$G \sim DP\,(\alpha, G_0). \tag{7.3}$$

The concentration parameter α controls the closeness of the distribution G to G_0. Since G_0 is a continuous-valued distribution, separate random draws from the distribution always return distinct values. Also, the discrete nature of G implies that separate draws from it can correspond in value with a positive probability.

A dataset $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$ can then be characterized by using the set of parameters $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\}$ described using the DP as a prior distribution as,

$$\boldsymbol{\theta}_n | G \sim G,\ n = 1, \ldots, N. \tag{7.4}$$

Because of the discreteness of G, the parameters $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\}$ can coincide in value, thus inducing clustering of the corresponding data points $\{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$. The extent to which the parameters coincide in value is determined by α, with a larger α indicating that lesser number of parameters coincide in value, i.e., more clusters are formed. The assignment of a given parameter $\boldsymbol{\theta}_n$ to a cluster is mathematically characterized using the conditional density of $\boldsymbol{\theta}_n$ given the rest of

the parameters $\boldsymbol{\Theta}^{(-n)}$ (other than $\boldsymbol{\theta}_n$), and obtained by integrating out G in (7.4) using the Pólya-urn relation [98, 105],

$$p\left(\boldsymbol{\theta}_n|\boldsymbol{\Theta}^{(-n)}, \alpha, G_0\right) = \frac{1}{\alpha + N - 1} \sum_{m=1}^{M} \nu_m^{(-n)}\, \delta(\boldsymbol{\theta}_n, \boldsymbol{\theta}_m) + \frac{\alpha}{\alpha + N - 1}\, G_0(\boldsymbol{\theta}_n). \quad (7.5)$$

In the above equation, $\nu_m^{(-n)}$ is the number of variables in $\boldsymbol{\Theta}^{(-n)}$ equal to $\boldsymbol{\theta}_m$. Note that more than one $\boldsymbol{\theta}_n$ in $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\}$ can be associated with the same mixture parameter $\boldsymbol{\theta}_m$ in $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M\}$ from (7.1). The conditional probability given in (7.5) implies that each variable $\boldsymbol{\theta}_n$ is assigned to an existing cluster m (whose parameters are $\boldsymbol{\theta}_m$) with probability $\nu_m^{(-n)} \Big/ (\alpha + N - 1)$ or is assigned to a new cluster with probability $\alpha \Big/ (\alpha + N - 1)$. The tradeoff between using existing clusters or creating new ones is determined by the concentration parameter α. Under the DP framework, the joint distribution of the variables $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\}$ does not change even if the ordering of the variables is altered.

Since separate draws from the discrete distribution G can correspond in value with a positive probability, there exist an infinite set of probabilities corresponding to the frequency of each possible value that G can return, that are distributed according to a stick-breaking process [106], given by,

$$\begin{aligned}
\boldsymbol{\theta}_m &\sim G_0,\ m = 1, \ldots, \infty, && (7.6)\\
\beta_j &\sim \text{Beta}(1, \alpha),\ j = 1, \ldots, \infty, && (7.7)\\
w_m &= \beta_m \prod_{j=1}^{m-1} (1 - \beta_j),\ m = 1, \ldots, \infty, && (7.8)\\
G(\boldsymbol{\theta}) &= \sum_{m=1}^{\infty} w_m\, \delta(\boldsymbol{\theta}, \boldsymbol{\theta}_m) && (7.9)
\end{aligned}$$

The dataset $\mathbf{Y}$ which needs to be modeled in terms of an underlying set of clusters, can be described as a hierarchial Bayesian model using the DP as the prior distribution as [97],

$$G \sim DP(\alpha, G_0), \tag{7.10}$$

$$\boldsymbol{\theta}_n|G \sim G,\ n = 1, \ldots, N, \tag{7.11}$$

$$\mathbf{y}_n|\boldsymbol{\theta}_n \sim p(\mathbf{y}_n|\boldsymbol{\theta}_n),\ n = 1, \ldots, N. \tag{7.12}$$

The above characterization implies that each $\mathbf{y}_n$, which is an observed data point or feature vector belonging to the dataset $\mathbf{Y}$, is associated with a set of hidden or unknown variables $\boldsymbol{\theta}_n$. In other words, $\mathbf{y}_n$ is actually drawn from a pdf $p(\mathbf{y}_n|\boldsymbol{\theta}_n)$ whose parameters are given by $\boldsymbol{\theta}_n$. The stick-breaking procedure in (7.6)-(7.9) can be combined with (7.10)-(7.12) to give,

$$\boldsymbol{\theta}_m \sim G_0,\ m = 1, \ldots, \infty, \tag{7.13}$$

$$\beta_j \sim \text{Beta}(1, \alpha),\ j = 1, \ldots, \infty, \tag{7.14}$$

$$w_m = \beta_m \prod_{j=1}^{m-1} (1 - \beta_j),\ m = 1, \ldots, \infty, \tag{7.15}$$

$$c_n|\mathbf{w} \sim \text{Categorical}(\mathbf{w}),\ n = 1, \ldots, N, \tag{7.16}$$

$$\mathbf{y}_n|c_n \sim p(\mathbf{y}_n|\boldsymbol{\theta}_{c_n}),\ n = 1, \ldots, N. \tag{7.17}$$

In the above equations, c_n is an unknown variable that indicates the cluster membership of the corresponding data point $\mathbf{y}_n$. Using this characterization of $\mathbf{y}_n$, the general mixture model given in (7.1) can be written in terms of an infinite DP mixture model as,

$$p(\mathbf{y}|\mathbf{w}, \boldsymbol{\Theta}) = \sum_{m=1}^{\infty} w_m\, p(\mathbf{y}|\boldsymbol{\theta}_m). \tag{7.18}$$

Thus, in general $M = \infty$ in (7.1). However, for practical purposes M is set to a finite value, provided that the error ϵ, due to the truncation of the number of clusters is within tolerable limits [105]. The truncation limit is selected according to [105],

$$\epsilon \approx 4\, N\, e^{-(M-1)/\alpha}. \tag{7.19}$$

7.1.2 Estimation of DP Model Parameters using Blocked Gibbs Sampling

As seen in Section 7.1.1, given the dataset or the set of features $\mathbf{Y}$, the aim is to infer the properties of the mixtures that constitute this data set. These properties include the effective number of mixtures, and their corresponding weights and parameters. This can be cast as a Bayesian inference problem, where the goal is to obtain the posterior pdf over the unknown parameters given a set of observations, in this case, the data set $\mathbf{Y}$. To be specific, given $\mathbf{Y}$, the posterior pdf $p(\mathbf{\Theta}, \mathbf{c}, \mathbf{w}|\mathbf{Y})$, where $\mathbf{c} = \{c_1, \ldots, c_N\}$, has to be estimated.

Several MCMC methods [80], which generate samples from a desired posterior pdf based on a constructed Markov chain, can be used for this purpose. One such method is the Gibbs sampling method which relies on the availability of conditional densities of each parameter given the other parameters and the data. The Markov chain is obtained by sampling each parameter which is a random variable conditioned on the previously sampled values of the other parameters (random variables), and the data. In this work, the MCMC method of blocked Gibbs sampling is used to sample from the joint distributions of sets or blocks of the unknown random variables, given the previously sampled values of other variables, to form the pdf of interest. The blocked Gibbs sampling algorithm that was developed for inference in DP mixture models in [105] is employed here.

Specifically, for each iteration i in the Markov chain, i.e., each Gibbs iteration, samples are drawn iteratively from the conditional pdfs of each parameter that are conditioned on previously sampled values of the other parameters. This can be given by [105],

$$\boldsymbol{\theta}_m^{(i)} \sim p(\boldsymbol{\theta}_m|\mathbf{c}^{(i-1)}, \mathbf{Y}),\ m = 1, \ldots, M, \tag{7.20}$$

$$c_n^{(i)} \sim p(c_n|\mathbf{\Theta}^{(i)}, \mathbf{w}^{(i-1)}, \mathbf{Y}),\ n = 1, \ldots, N, \tag{7.21}$$

$$w_m^{(i)} \sim p(w_m|\mathbf{c}^{(i)}),\ m = 1, \ldots, M. \tag{7.22}$$

The conditional posterior pdfs in (7.21)-(7.22) can be explicitly given as [105],

$$p\left(\boldsymbol{\theta}_m|\mathbf{c},\mathbf{Y}\right) \propto G_0(\boldsymbol{\theta}_m)\prod_{n:c_n=m} p\left(\mathbf{y}_n|\boldsymbol{\theta}_m\right),\ m=1,\ldots,M, \tag{7.23}$$

$$p\left(c_n|\boldsymbol{\Theta},\mathbf{w},\mathbf{Y}\right) = \sum_{m=1}^{M}\Big(w_m\, p\left(\mathbf{y}_n|\boldsymbol{\theta}_m\right)\Big)\,\delta(c_n,m),\ n=1,\ldots,N, \tag{7.24}$$

$$p\left(w_m|\mathbf{c}\right) = \beta_m\prod_{j=1}^{m-1}(1-\beta_j),\ m=1,\ldots,M. \tag{7.25}$$

In the above equation, $n : c_n = m$ represent the indices in $\mathbf{c} = \{c_1,\ldots,c_N\}$ for which $c_n = m$ and

$$\beta_m \sim \text{Beta}\Big(1+\nu_m, \alpha+\sum_{m'=m+1}^{M}\nu_{m'}\Big), \tag{7.26}$$

where ν_q is the number of elements in $\mathbf{c}$ that are equal to m.

Bayesian inference provides a method to solve for the posterior pdf by multiplying the prior and the likelihood function and dividing this by a normalizing constant which is the integration of the product of the prior and the likelihood function over the the entire parameter space. Doing this can be computationally expensive and analytical solutions may not always be available. An efficient way of performing the update for the posterior pdf is by using the concept of conjugate priors [107]. This method relies on finding pairs of prior and posterior pdfs which are conjugate to each other, because of which an analytical solution to the integral becomes available. Specifically, given the likelihood $p\left(\mathbf{y}_n|\boldsymbol{\theta}_m\right)$ in (7.23), the base distribution or the prior G_0 can be chosen appropriately, such that the prior and posterior pdf $p\left(\boldsymbol{\theta}_m|\mathbf{c},\mathbf{Y}\right)$, belong to the same family of distributions. This enables the update step for the posterior pdf over the parameter $\boldsymbol{\theta}$ to be performed efficiently. Such priors are referred to as conjugate priors.

Since a DP GMM is being used in this work, the likelihood $p\left(\mathbf{y}_n|\boldsymbol{\theta}_m\right)$ in (7.21) is a Gaussian pdf with unknown mean and precision, as seen in (7.2). It is

assumed that the prior G_0 in (7.21) is a Normal-Wishart distribution [107], i.e.,

$$G_0(\boldsymbol{\theta}) \triangleq \mathcal{NW}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}; \boldsymbol{\mu}_{\mathcal{N}}, \tau_{\mathcal{N}}, \phi_{\mathcal{W}}, \tau_{\mathcal{W}}\right), \tag{7.27}$$

where, $\boldsymbol{\mu}_{\mathcal{N}}$, $\tau_{\mathcal{N}}$, $\phi_{\mathcal{W}}$ and $\tau_{\mathcal{W}}$ are hyperparameters. Specifically, $\phi_{\mathcal{W}}\,(\phi_{\mathcal{W}} > N_{\mathbf{y}} - 1)$ representing the number of degrees of freedom and $\tau_{\mathcal{W}}$ representing the symmetric positive definite precision matrix are the Wishart hyperparameters and, $\boldsymbol{\mu}_{\mathcal{N}}$, $\tau_{\mathcal{N}}\boldsymbol{\Sigma}^{-1}$ denoting the mean vector and the symmetric positive definite precision matrix, respectively, where $\tau_{\mathcal{N}} > 0$, are the hyperparameters corresponding to the multivariate normal (Gaussian) component of the Normal-Wishart prior.

The choice of a Normal-Wishart prior given the multivariate Gaussian likelihood with unknown mean and precision, results in a posterior pdf that is characterized by a Normal-Wishart distribution. Thus,

$$p\left(\boldsymbol{\theta}|\mathbf{c}, \mathbf{Y}\right) \triangleq \mathcal{NW}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}; \tilde{\boldsymbol{\mu}}_{\mathcal{N}}, \tilde{\tau}_{\mathcal{N}}, \tilde{\phi}_{\mathcal{W}}, \tilde{\tau}_{\mathcal{W}}\right), \tag{7.28}$$

where $\tilde{\boldsymbol{\mu}}_{\mathcal{N}}$, $\tilde{\tau}_{\mathcal{N}}$, $\tilde{\phi}_{\mathcal{W}}$, and $\tilde{\tau}_{\mathcal{W}}$ are the updated hyperparameters of the Normal-Wishart distribution representing the posterior pdf and are given by [107],

$$\tilde{\boldsymbol{\mu}}_{\mathcal{N}} = \frac{\tau_{\mathcal{N}}\,\boldsymbol{\mu}_{\mathcal{N}} + N\boldsymbol{\mu}_{\mathbf{Y}}}{\tau_{\mathcal{N}} + N}, \tag{7.29}$$

$$\tilde{\tau}_{\mathcal{N}} = \tau_{\mathcal{N}} + N, \tag{7.30}$$

$$\tilde{\tau}_{\mathcal{W}} = \tau_{\mathcal{W}} + \boldsymbol{\Sigma}_{\mathbf{Y}} + \frac{\tau_{\mathcal{N}}\,N}{\tau_{\mathcal{N}} + N}\left(\boldsymbol{\mu}_{\mathcal{N}} - \boldsymbol{\mu}_{\mathbf{Y}}\right)\left(\boldsymbol{\mu}_{\mathcal{N}} - \boldsymbol{\mu}_{\mathbf{Y}}\right)^{T}, \tag{7.31}$$

$$\tilde{\phi}_{\mathcal{W}} = \phi_{\mathcal{W}} + N. \tag{7.32}$$

In the above set of equations, $\boldsymbol{\mu}_{\mathbf{Y}}$ and $\boldsymbol{\Sigma}_{\mathbf{Y}}$ are the sample mean and covariance, respectively, of the dataset or set of features $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$ with N data points. The update step for the posterior pdf over the parameters $\boldsymbol{\theta}_m = \{\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m^{-1}\}$ using the concept of conjugate priors thus just amounts to an update of the Normal-Wishart hyperparameters given by (7.29)-(7.32).

The complete steps of the blocked Gibbs sampling algorithm for estimating the parameters of a DP GMM using an $N_{\mathbf{y}}$-dimensional dataset are given in Algorithm 4.

7.2 Framework for ECG Beat Clustering using the DP GMM

In order to cluster ECG beats using the DP GMM framework, real ECG data is obtained from the MIT-BIH arrhythmia database [89]. The MIT-BIH arrhythmia database consists of 48 ECG recordings that were obtained by the Beth Israel Hospital Arrhythmia Laboratory and are each 30 min. long. These signals represent normal ECG beats along with those with several different types of arrhythmia that were annotated by cardiologists. Four of these recordings contain paced ECG beats (generated by an implanted pacemaker, MIT-BIH arrhythmia notation: P) and beats that are a fusion of normal and paced beats (f), and are ignored in this work. In addition to the N, L, R, E, j beats that were used for demonstrating the classification performance of the Bayes ML classifier in Chapter 6, other arrhythmias which include three types of supraventrical arrhythmia (arrhythmias originating in the atria or near the atrio-ventricular junction [72]), namely, atrial premature beat (A), aberrated atrial premature beat (a) and supraventrical premature beat (S), and one type of ventricular arrhythmia (caused due to rhythm disturbances below the atrio-ventricular junction [72]), namely premature ventricular contraction (V), were also considered for clustering using the DP algorithm. Two other types of beats also present in the data include beats caused due to the fusion of normal and ventricular beats (F), and unclassifiable beats (which represent the beats that the cardiologists were not able to annotate, MIT-BIH arrhythmia notation: Q). There were approximately $1,500$-$3,500$ beats in each record.

Algorithm 4 Blocked Gibbs sampling for DP GMM using an $N_{\mathbf{y}}$-dimensional dataset $\mathbf{Y}$

Repeat for $i = 1, 2, \ldots,$ Gibbs iterations:

1. Update for $\boldsymbol{\theta}_m^{(i)} = \{\boldsymbol{\mu}_m^{(i)}, \boldsymbol{\Sigma}_m^{-1(i)}\} \sim p\left(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m^{-1} | \mathbf{c}^{(i-1)}, \mathbf{Y}\right)$, $m = 1, \ldots, M$.
   (a) Let $\mathbf{Y}_m = \{\mathbf{y}_n : c_n^{(i-1)} = m\}$ and $N_m = |\mathbf{Y}_m|$, for $m = 1, \ldots, M$.
   (b) For all clusters, $m = 1, \ldots, M$, compute,

$$\begin{aligned}
\boldsymbol{\mu}_{\mathbf{y}_m} &= \frac{1}{N_m} \sum_{n:c_n^{(i-1)}=m} \mathbf{y}_n \\
\boldsymbol{\Sigma}_{\mathbf{y}_m} &= \frac{1}{N_m} \sum_{n:c_n^{(i-1)}=m} (\mathbf{y}_m - \boldsymbol{\mu}_{\mathbf{y}_m})^2 \\
\tilde{\boldsymbol{\mu}}_{\mathcal{N},m} &= \frac{\tau_{\mathcal{N}}\, \tilde{\boldsymbol{\mu}}_{\mathcal{N}} + N_m\, \boldsymbol{\mu}_{\mathbf{y}_m}}{\tau_{\mathcal{N}} + N_m}\,, \\
\tilde{\tau}_{\mathcal{N},m} &= \tau_{\mathcal{N}} + N_m\,, \\
\tilde{\tau}_{\mathcal{W},m} &= \tau_{\mathcal{W}} + \boldsymbol{\Sigma}_{\mathbf{y}_m} + \frac{\tau_{\mathcal{N}}\, N_m}{\tau_{\mathcal{N}} + N_m} (\mathbf{m} - \boldsymbol{\mu}_{\mathbf{y}_m}) (\mathbf{m} - \boldsymbol{\mu}_{\mathbf{y}_m})^T\,, \\
\tilde{\phi}_{\mathcal{W},m} &= \phi_{\mathcal{W}} + N_m\,.
\end{aligned}$$

   (c) Draw samples for $\boldsymbol{\Sigma}_m^{-1(i)}$ from the Wishart distribution, $\mathcal{W}\left(\boldsymbol{\Sigma}_m^{-1}; \tilde{\tau}_{\mathcal{W},m}, \tilde{\phi}_{\mathcal{W},m}\right)$, for $m = 1, \ldots, M$.
   (d) Finally draw samples for $\boldsymbol{\mu}_m^{(i)}$ from the Normal distribution, $\mathcal{N}\left(\boldsymbol{\mu}_m; \tilde{\boldsymbol{\mu}}_{\mathcal{N},m}, \frac{\boldsymbol{\Sigma}_m^{(i)}}{\tilde{\tau}_{\mathcal{N},m}}\right)$, for $m = 1, \ldots, M$.

2. Update for $c_n^{(i)} \sim p\left(c_n | \boldsymbol{\mu}^{(i)}, \boldsymbol{\Sigma}^{-1(i)}, \mathbf{w}^{(i-1)}, \mathbf{Y}\right)$, $n = 1, \ldots, N$.
   (a) Let $q_{m,n} \triangleq w_m^{(i-1)} \mathcal{N}\left(\mathbf{y}_n; \boldsymbol{\mu}_m^{(i)}, \boldsymbol{\Sigma}_m^{(i)}\right)$, $m = 1, \ldots, M$ and $n = 1, \ldots, N$.
   (b) Normalize $q'_{m,n} = \frac{q_{m,n}}{\sum_{m=1}^{M} q_{m,n}}$, $m = 1, \ldots, M$ and $n = 1, \ldots, N$.
   (c) Draw samples for $c_n^{(i)} \sim \sum_{m=1}^{M} q'_{m,n} \delta(c_n, m)$, $n = 1, \ldots, N$.

3. Update for $w_m^{(i)} \sim p\left(w_m | \mathbf{c}^{(i)}\right)$, $m = 1, \ldots, M$.
   (a) Draw samples $\beta_j \sim \text{Beta}\left(1 + \nu_m, \alpha + \sum_{m'=m+1}^{M} \nu_{m'}\right)$, where $\nu_m \triangleq |\{n : c_n^{(i)} = m\}|$, $m = 1, \ldots, M$.
   (b) Finally evaluate $w_m^{(i)} = \beta_m \prod_{j=1}^{m-1} (1 - \beta_j)$, $m = 1, \ldots, M$.

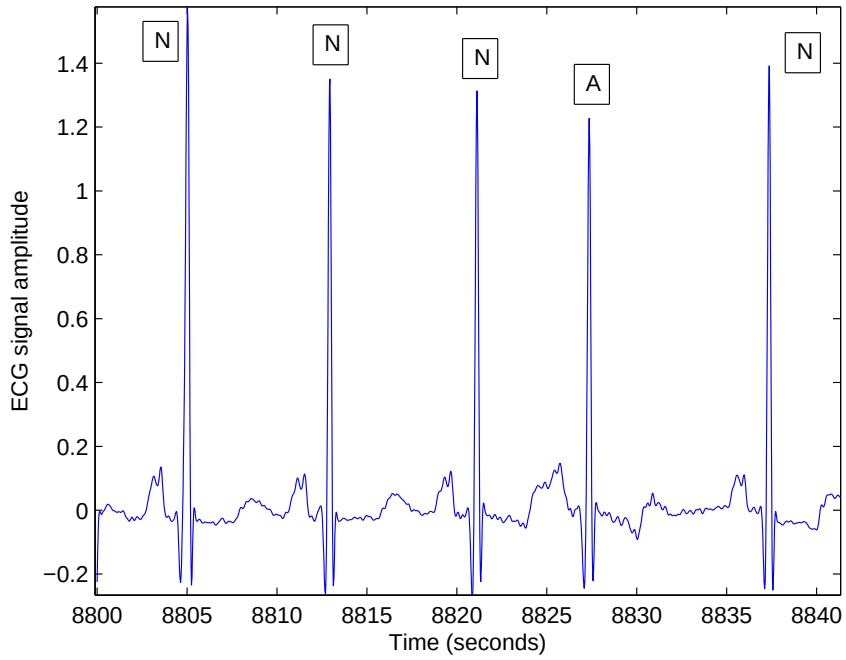


Figure 7.1: Illustration of beat types with morphological similarity and temporal differences using N and A type beats from Record 101 of MIT-BIH arrhythmia database.

7.2.1 Feature Design for ECG Beat Clustering

The ECG data is used as the input to the proposed SMCMC model with simultaneous model selection described in Chapter 5, and the estimated noise-free ECG signals are generated for each beat, from which the features describing the ECG morphology are obtained. The ECG signals are sampled at 360 Hz and each ECG beat has a duration of 1 s on an average. In order to limit the feature space size, only a fixed number of ECG morphology features are used as described in Section 6.5. These include the mean of the reconstructed ECG signal around the P wave and the local averages of the reconstructed ECG measurements at five regions in the QRS complex.

In addition, since the grouping of certain arrhythmias depends upon their timing information rather than their morphology [45, 56], temporal features are

also included in the feature set. An important temporal feature used to distinguish between ECG signals is the distance between successive R peaks of each ECG beat, and is known as the RR-interval. The pre-RR and post-RR intervals are defined as the distances between the R peak of a given beat and the R peak of the previous beat and the next beat, respectively. In this work, the average pre-RR and post-RR intervals, which are calculated for each ECG beat as being the local average of the pre-RR and post-RR intervals over ten neighboring beats, are used as the temporal features. An illustration of beats which are morphologically similar, but have differences in temporal features is shown in Figure 7.1 with N and A beats obtained from Record 101 of the MIT-BIH arrhythmia database. It can be observed that both the N and A type beats have similar morphologies, but since the A type beats occur due to premature depolarization of the atria [72], their pre-RR and post-RR intervals are respectively, smaller and larger, when compared to those of N type beats. Also, since certain arrhythmias such as premature ventricular contraction give rise to QRS complexes that are wider than the QRS complexes of other beat types [72], the QRS width (time between the QRS onset and QRS offset) is also included in the feature set. An example of the differences between QRS widths of V type and other beats can be seen in Figure 7.4 using data from Record 208 of the MIT-BIH arrhythmia database. It can be seen that the V type beat has a wider QRS complex when compared to both N and F type beats. Thus, the feature set $\mathbf{y}$ of each ECG beat consists of:

(a) Mean of the noise-free reconstructed ECG samples from around the P wave.

(b) Local averages of noise-free reconstructed ECG samples at 5 points in the QRS complex.

(c) Average pre-RR and post-RR intervals.

(d) QRS width.

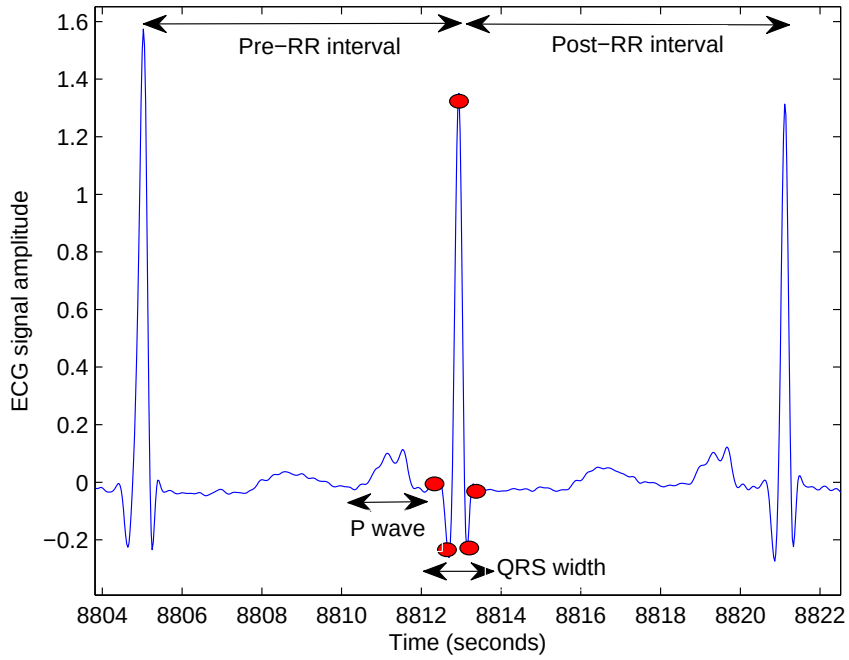


Figure 7.2: Illustration of the features used to form the feature set. The dots denote the selected 5 points in the QRS complex for a typical case.

Also, in order to find the location of the P wave and the QRS complex, the method described in Section 6.4 based on the R peak location and the QRS, PR intervals can be used. However, since the aim is to differentiate between additional ECG beat classes (in addition to those classified using the Bayes ML classifier in Sections 6.4 and 6.5), some of which have intermittent changes in morphology (such as the V type beats), this method is not employed. Instead, the ECG delineation routine *"ecgpuwave"* based on the algorithm in [108] and available on the online Physionet database [89] is utilized to find the onset and offset positions of the P wave and the QRS complex. An illustration of the selected feature set is shown Figure 7.2, wherein, the delineation of the P wave, typical selection of 5 points in the QRS complex, the temporal features, and the QRS width are illustrated.

For each ECG recording in the MIT-BIH arrhythmia, the dataset consists

of the feature vectors obtained from each ECG beat as described above. Each feature vector consists of 9 samples, and hence the dimension of each data point in the dataset is $N_{\mathbf{y}} = 9$. This dataset is used as the input to the DP algorithm which then clusters the data. The number of clusters in the DP algorithm is limited by the truncation limit in (7.19). However, the number of clusters found in the actual dataset is lesser than total number of clusters used by the algorithm. Hence, only a few clusters have significant mixture weights associated with them. In other words, only a few weights out of the estimated mixture weights in $\mathbf{w} = \{w_1, \ldots, w_M\}$ are significant. Also, similar types of beats may be assigned to more than one cluster if a single Gaussian is not adequate the represent all the data points corresponding to a particular type of beat.

7.3 Cluster Labeling using Bayes ML Method

As mentioned previously, the DP only clusters the data, but does not assign labels to the clusters obtained in the data. Once clustering is performed, the data can be labeled using expert knowledge, if it is available. An ECG beat can then said to be correctly classified if it falls in a cluster whose dominant beat (which represents the beat type to which majority of the data points in the cluster belong) is of the same type. Incorrect classification takes place when an ECG beat falls in a cluster where it is the non-dominant beat.

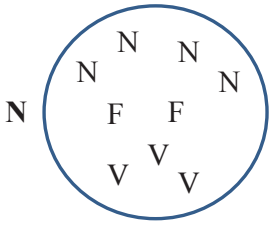
In scenarios wherein external expert knowledge is not available, a supervised learning algorithm can be used to label the generated ECG clusters, if training data is available. So, once each individual ECG recording is clustered using the DP algorithm, each of these clusters is labeled using a supervised learning algorithm. Thus, this helps in preserving the inter-patient variability between the ECG recordings of different patients and makes the algorithm patient-adaptable by first differentiating between the beats of each individual, and then using the supervised approach only to label these clusters.

In this work, the Bayes ML method described in Section 6.2 is used for this purpose. In order to make the Bayes ML method work in conjunction with the DP algorithm, the same feature vector structure as described in Section 7.2 is used for training. The training is performed using a portion of the data from each recording that had not been previously clustered, and which is representative of the beat types that were sought to be clustered. In cases where sufficient training data is not available for a certain beat type, the beat is only labeled assuming that expert knowledge is available (in the form of annotations already provided by the database), and not automatically labeled using the Bayes ML method. For each ECG recording, the testing data consists of the means of all the clusters found by the DP algorithm. For labeling a given test feature vector (one of the cluster means), the likelihood function in (6.1) is firstly evaluated using the corresponding mean and covariance of each beat type used for training. Finally, the label of the beat type whose mean and covariance maximized the log-likelihood function in (6.2) is assigned to the cluster. This means that a single label is assigned to all data points in the cluster, and thus if a certain data point corresponds to a beat type whose true label is not the same as the assigned label, then it is said to be misclassified.

7.4 Simulations and Discussion

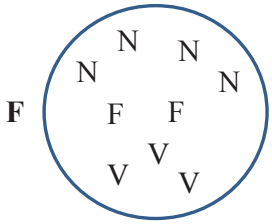
For implementing the DP algorithm, the concentration parameter was set to $\alpha = 3$. The number of mixtures in the DP is truncated to $M = 41$ terms by setting the error in (7.19) to $\epsilon = 10^{-2}$. The Gibbs sampler was initialized by assuming that all the data points in the given dataset fall into the same cluster. Also, the number of iterations performed for burn-in and sample collection from the Gibbs sampler were set to 5000 and 6500 respectively.

Firstly, results demonstrating the algorithm performance are shown assuming that expert knowledge is available to label the clusters. In this method,



Beat type	Classification rate
N	4/4
V	0/3
F	0/3

(a) Using labels provided by an expert



Beat type	Classification rate
N	0/4
V	0/3
F	2/2

(b) Using labels from Bayes ML classifier

Figure 7.3: Example showing evaluation of performance of the DP algorithm using the different labeling schemes discussed. The labels inside the circle which denotes the cluster, are the true labels and the labels outside the circle in bold are the assigned labels.

any beat that is not the dominant beat in a cluster will be considered to be misclassified. In other words, the expert labels the dominant beat of the cluster and all the beats of the cluster are assigned the same label. It is assumed that expert knowledge is available in terms of the annotations provided by the MIT-BIH ar-

Table 7.1: Confusion matrix showing clustering results using the DP algorithm.

Beat type	N	L	R	j	A	a	S	V	E	F	Q
N	1	0	0	0	0	0	0	0	0	0	0
L	0	0.97	0	0	0	0	0	0.03	0	0	0
R	0	0	0.99	0	0	0	0	0.01	0	0	0
j	0	0	0.02	0.98	0	0	0	0	0	0	0
A	0.01	0	0	0	0.99	0	0	0	0	0	0
a	0	0	0	0	0.07	0.93	0	0	0	0	0
S	1	0	0	0	0	0	0	0	0	0	0
V	0.02	0	0	0	0	0	0	0.98	0	0	0
E	0	0	0	0	0	0	0	0	1	0	0
F	0.03	0	0	0	0	0	0	0.46	0	0.51	0
Q	0.13	0	0	0	0	0	0	0.27	0	0	0.6

rhythmia database. This is shown in Figure 7.3(a) using a simple example with one cluster consisting of 4 N, 3 V and 2 F (true labels) type beats. The dominant beat of the cluster is the N type beat and hence the assigned label is N, using which the correct classification rate of all the beats is indicated in the table at the bottom of the figure. The results of the performance of the DP clustering algorithm are shown in Table 7.1.

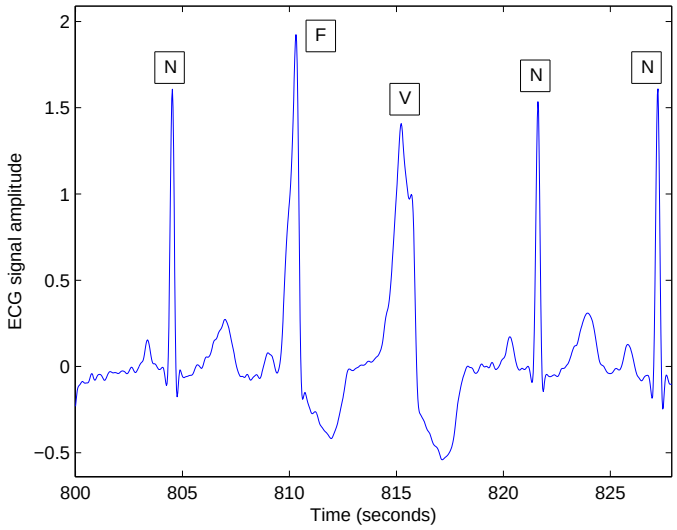
The results using the Bayes ML method for labeling the clusters (in case the expert knowledge is not available) are outlined next. In this case, misclassification occurs when the label assigned by the classifier to the cluster does not correspond to the true label of any beat type present in the cluster. This is demonstrated using an example in Figure 7.3(b). The cluster given by the DP algorithm consists of 4 N, 3 V and 2 F (true labels) type beats. The label assigned to the cluster using the Bayes ML classifier is F. This is different from the label assigned by an expert as seen in Figure 7.3(a), and shows that the output of the Bayes ML method does not necessarily depend on the dominant beat but is based on the cluster means found by the DP. However, it must be noted that this is not always the case, and that the label assigned by the Bayes ML method can correspond to that assigned by the expert in other instances. Using this assigned label, the

classification results are shown in the table given at the bottom of the figure. The results of the performance of the Bayes ML method for labeling the clusters are shown in Table 7.2.

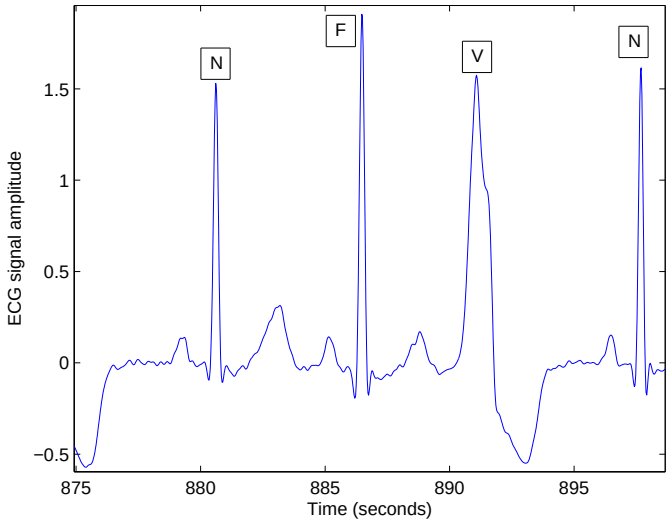
Table 7.2: Confusion matrix showing cluster labeling results using the Bayes ML method.

Beat type	N	L	R	j	A	V	E	F
N	0.99	0	0	0	0.002	0.001	0	0.007
L	0	0.97	0	0	0	0.03	0	0
R	0	0	0.99	0	0	0.01	0	0
j	0	0	0.02	0.98	0	0	0	0
A	0.02	0	0.03	0	0.97	0	0	0
V	0	0	0	0	0	0.98	0	0.02
E	0	0	0	0	0	0	1	0
F	0	0	0	0	0	0.01	0	0.99

From Tables 7.1 and 7.2, it can be seen that both the DP algorithm for clustering ECG data and the Bayes ML method for labeling the generated cluster perform fairly well for most of the beat types. From Table 7.1 it can be seen that the DP correctly clusters 98% of all the considered beats, excluding F, S and Q type beats. The performance of the DP algorithm for clustering the F type data is less accurate when compared to that of the other beats because the F beats are assigned to the clusters where either N or V beats are dominant, and are thus misclassified. This is because the F type beats represent the fusion of both N and V type beats and are morphologically very similar to both these beats, and there exists a considerable amount of uncertainty even among doctors to annotate them [60]. An illustration of the similarity of F type beats to both N and V types is shown using data obtained from Record 208 of the MIT-BIH arrhythmia



(a) F type ECG beat similar to a V type beat



(b) F type ECG beat similar to a N type beat

Figure 7.4: Illustration of morphological similarity of F type beats to both N and V beat types.

database in Figure 7.4. However, from Table 7.2, it can be seen that most of the F type beats have been labeled correctly by the Bayes ML method because the means of the clusters found by the DP classifier are closer to the original mean of the data with F type beats. Also, it is seen from Table 7.1 that the S type beats are incorrectly clustered. This is because of the number of representative S

type beats available in the entire database was very small when compared to the number of beats corresponding to the other beat types (Only a total of 2 S type beats were present in all the 44 records combined, in the MIT-BIH arrhythmia database). For the same reason, the S type beats were not labeled using the Bayes ML method. Similarly, since only a total of 15 Q type beats were present in all the analyzed records put together, the accuracy of correctly clustering these beats was lesser when compared to that of most of the other beat types. Thus, the Q type beats and also the a type beats are not labeled using the Bayes ML method, since enough training data was not available. For the considered beats, the Bayes ML method correctly labeled 98.3% of the total number of beats.

An example of the clustering and labeling performance is shown in Figure 7.5 for Record 207 of the MIT-BIH arrhythmia database. Figure 7.5(a) shows the different beat types of the record, with each beat type being associated with a different color. It must be noted that for illustrative purposes, the beats are grouped together according to their type, but the beats occur without any particular order in the actual ECG recording. Also, in this figure the y-axis (height of stem plot) is not indicative of any quantity and the difference in the heights of the stem plots for each group is only to show the differences between the memberships of different groups. Figures 7.5(b) and 7.5(c) show the clusters that the beats have been assigned to, and the cluster labels given by the Bayes ML method, respectively. In both these figures, the color of the line indicates the true label (beat type) for the corresponding beat, whereas the marker color indicates the cluster label to which the corresponding beat has been assigned to. In Figure 7.5(b), it is assumed that the dominant beat of the cluster has been labeled by an expert. A number of observations can be made from this figure. Firstly, it can be observed that the beats can get assigned to more than one cluster. For example, beats of the R type were assigned to clusters 14 and 16. Secondly, errors in clustering can

Table 7.3: Grouping of MIT-BIH arrhythmia database beat types into AAMI recommended beat types.

AAMI beat type	MIT-BIH arrhytmia database beat type
N^A	N, L, R, j
S^A	A, a, S
V^A	V, E
F^A	F
Q^A	P, f, Q

be noticed due to the fact that 39 L type and 1 R type beat were assigned to cluster 16 whose dominant beat was the V type. The labels given to the clusters using the Bayes ML method are shown in Figure 7.5(c). It can be seen that the clusters to which most of the beats of a particular beat type are assigned to are labeled correctly, whereas the smaller clusters are labeled erroneously. This is due to the fact that the small amount of data in these clusters causes the cluster mean to not be a good representation of the true value.

The Association for the Advancement of Medical Instrumentation (AAMI) recommends the reporting of performance of algorithms designed for cardiac signal processing using a set of pre-defined standards [109, 110]. According to these standards, the different heart beat types from the MIT-BIH arrhythmia database are grouped into 5 different classes. Each class can include beats of more than one type from the MIT-BIH arrhythmia database. The grouping of the different classes according to the AAMI practice is shown in Table 7.3, using which it can be seen that, for example, all the beats belonging to N, L, R and j types are grouped into a single class denoted by N^A. Also, since the P and f type beats are ignored in this work, the Q^A beat type only consists of Q type beats.

In order to report the performance of the proposed DP clustering algorithm according to the AAMI recommended practice, it was assumed that a beat belonging to N^A, S^A, V^A, F^A and Q^A types is correctly clustered if it was assigned to a cluster where the dominant beat belonged to the corresponding type. For

Table 7.4: Confusion matrix showing clustering results for the DP algorithm using AAMI recommended practice.

AAMI beat type	N^A	S^A	V^A	F^A	Q^A
N^A	1	0	0	0	0
S^A	0.02	0.98	0	0	0
V^A	0.02	0	0.98	0	0
F^A	0.03	0	0.46	0.51	0
Q^A	0.13	0	0.27	0	0.60

Table 7.5: Confusion matrix showing cluster labeling results with the Bayes ML method using AAMI recommended practice.

AAMI beat type	N^A	S^A	V^A	F^A
N^A	0.99	0	0.004	0.006
S^A	0.01	0.99	0	0
V^A	0	0	0.98	0.02
F^A	0	0	0.01	0.99

example, if any of the MIT-BIH arrhythmia database beat types, N, L, R and j, are assigned to a cluster where the dominant beat is one of these types, then correct classification takes place. A similar method is adopted to report the performance of the Bayes ML method that is used for automated labeling of the DP clusters. It must be noted that since an insufficient number of S, a and Q type beats are available to train the classifier, the S and a type beats are excluded from the AAMI beat type S^A, and the AAMI beat type Q^A is altogether ignored for reporting the labeling results using the Bayes ML method. Using the AAMI recommended practice, the performance of the DP clustering algorithm and the Bayes ML method for labeling the DP clusters are given in Tables 7.4 and 7.5.

In order to compare the results in this work to existing approaches, three approaches are chosen. These include the mixture-of-experts (MOE) approach in [50], clustering approach based on self-organizing maps (SOM) [52] and the linear discriminant approach in [46]. Among these works, the classifiers proposed in [50] by Hu *et al.* and in [46] by de Chazal *et al.* were based on the global-local classifier approach, whereas Lagerholm *et al.* proposed an unsupervised clustering

method in [52]. In [50], a global classifier was first used to classify an individual's ECG signal based on a large existing database of ECG signals. These results were combined with those from a local classifier that was trained specifically using data from the an individual's ECG record, using the MOE approach. However, the aim in this work was to classify the ECG signals as belonging to only two classes, namely V (premature ventricular contraction) and non-V. The performance or error measure that was defined in this work was given by,

$$R^{MOE} = \frac{\text{Number of true negative and true positive beats}}{\text{Total number of beats}} \tag{7.33}$$

SOMs were used in [52] to cluster the beats from each ECG record to a pre-defined number of clusters. It was assumed that expert knowledge was available to perform labeling of these clusters. The dominant beat of the cluster was assigned a label by the expert and this label was used to denote all the beats of that cluster. In order to facilitate the comparison of results in this work to those given by the MOE approach of [50], the performance measure used was given by,

$$R^{SOM} = \frac{\text{Number of correctly clustered V and E type beats}}{\text{Total number of beats}} \tag{7.34}$$

A global classifier based on linear discriminants (LD) to classify ECG beats into 5 classes based on the AAMI recommended practice was used in [46]. The results of the global classifier were validated by an expert and used to train a local classifier. The global and local classifier outputs were combined to produce a final classification result. The results in this work were compared to those in [50] and [52] using a performance measure based on determining the clustering performance of the method for each of the possible 5 AAMI classes. In order to do so, the classifier output for each heart beat was considered to be a cluster label and the performance measure was calculated using the number of correctly clustered V and E type beats ($CC_{V,E}$) and the number of correctly clustered non-V and

non-E type beats ($\mathrm{CC}_{\mathrm{non-V,E}}$) as,

$$R^{LD} = \frac{\mathrm{CC}_{\mathrm{V,E}} + \mathrm{CC}_{\mathrm{non-V,E}}}{\text{Total number of beats}} \tag{7.35}$$

In this work, a performance measure similar to the one used in [52] and defined in (7.34) is calculated in order to facilitate comparison. The calculations are performed using the DP clustering results in which it was assumed that an expert was available to label each cluster generated by the DP. The performance measure for the proposed DP algorithm is given by,

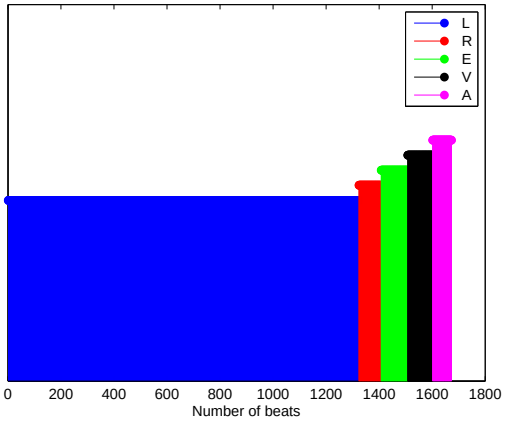
$$R^{DP} = \frac{\text{Number of correctly clustered V and E type beats}}{\text{Total number of beats}} \tag{7.36}$$

In order to facilitate a fair comparison, R^{DP} is evaluated for only the specific ECG records from the MIT-BIH arrhythmia database which were used in common by the three previous works for performance evaluation and comparison.

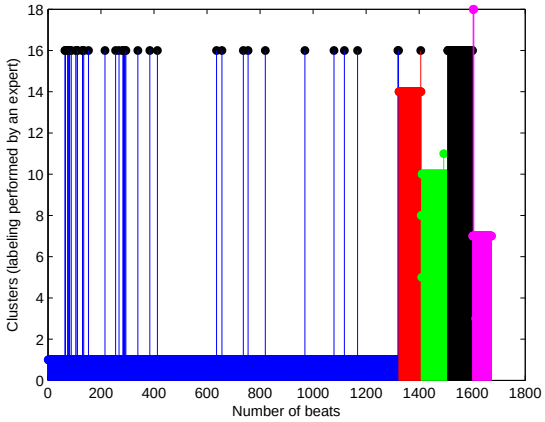
The values of all the aforementioned performance measures are enumerated in Table 7.5. It can be seen that the results using the proposed DP clustering algorithm compare favorably with the previous works. However, the DP algorithm offers a significant advantage over these other methods because it does not require separate training and testing datasets as in [46, 50], and can adaptively learn the number of clusters from ECG data which can evolve over time without requiring *a priori* information about the number of diseases as in [52].

Table 7.6: Comparison of clustering results.

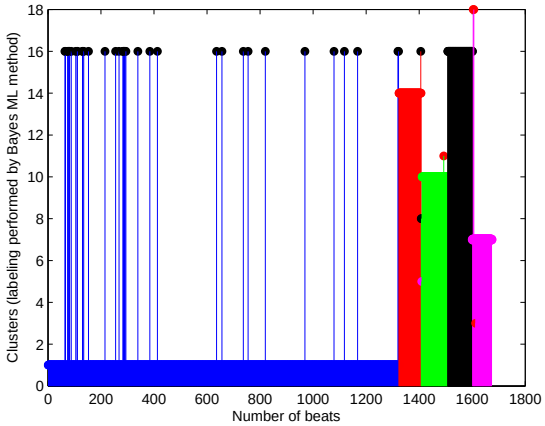
Record	R^{MOE}	R^{SOM}	R^{LD}	R^{DP}
200	0.81	0.98	0.97	1
201	0.95	0.99	–	0.99
202	0.72	0.99	0.99	1
203	0.87	0.97	–	1
205	0.97	0.99	–	0.99
207	0.88	0.97	–	1
208	0.91	0.99	–	0.97
210	0.93	0.98	0.97	0.99
213	0.92	0.98	0.99	0.96
214	0.98	0.99	1	0.99
215	0.98	0.99	–	1
219	0.97	0.99	0.99	0.95
221	0.99	1	0.99	1
223	0.94	0.99	–	0.97
228	0.99	1	0.99	0.98
231	0.99	0.99	1	1
233	0.98	0.99	0.99	1
234	0.99	1	1	1



(a) Different beat types in Record 207



(b) Clusters assigned to each beat, assuming the availability of cluster labels by an expert



(c) Clusters assigned to each beat, using cluster labels provided by the Bayes ML method

Figure 7.5: Example of clustering and labeling performance using Record 207 of the MIT-BIH arrhythmia database.

Chapter 8

CONCLUSIONS AND FUTURE WORK

In this work, the focus was on first developing statistical models for estimating ECG signal parameters that did not require preprocessing steps for filter initialization or ECG fiducial point delineation. Constructing such models, whose estimated parameters can later be used for automatic classification of several cardiac diseases is very helpful, and is of great importance as it helps in avoiding manual annotation and provides speedy diagnosis. To work towards this goal, novel methods for modeling ECG signals and adaptive cardiac parameter estimation using sequential Bayesian methods were proposed. To perform ECG signal classification, the proposed model parameters were first used as the feature set for a simple Bayesian ML classifier to classify between different disease types. In addition, an adaptive learning framework based on the DP was employed to cluster the ECG data and provided the basis for a patient-specific algorithm for classifying between different cardiac arrhythmias.

8.1 Conclusions

(a) **Adaptive parameter estimation:** Current ECG statistical models exhibit dependencies on requirement of *a priori* information about the ECG signals and presence of a number of user-defined parameters. In order to avoid these issues, a method for adaptive parameter selection using the existing multi-harmonic ECG model [29, 30] was presented. The proposed algorithm can adaptively select parameters such as number of harmonics and mean cardiac frequency by minimizing the estimation MSE. Thus, the selection of the best parameter pair which leads to an improvement in the estimation accuracy of the cardiac signal parameters represented by the model is enabled. Results using real ECG data from the online Physionet database [89] demonstrate

that the performance of the adaptive algorithm, in terms of the estimation MSE, is superior to the algorithm performance when the values of the parameter pair are fixed. It should be noted that the size of the state vector in this model is determined by the number of harmonics N_h, and is in the order of $2N_h + 3$. The use of a large number of harmonics should be avoided, as it can potentially increase model complexity and may lead to a scenario in which a large number of state variables have to be estimated from a given small number of data samples in each segment.

(b) **ECG classification using adaptively estimated parameters:** The instantaneous cardiac frequency estimates were used to classify between four types ECG signals, including, signals with normal sinus rhythm, supraventricular arrhythmia, malignant ventricular arrhythmia and atrial fibrillation. It was seen that the classifier performed fairly well, giving nearly a 90% correct classification rate.

(c) **ECG signal modeling using multiple models with the IMM-KF algorithm:** Another issue faced with existing ECG models is the use of a single representation to model different types of ECG morphologies. Since ECG morphologies vary across different diseases this might not be a feasible option for representing all types of ECG signals. In addition, some existing models delineate ECG signals such that each ECG segment is modeled using a separate representation. To preclude these possibilities, two novel ECG models based on utilizing multiple signal models were presented.

In the first proposed approach, the IMM technique was employed to model the ECG signal using three different polynomials, namely, linear, quadratic and cubic. The polynomial coefficients represent the model parameters and are estimated using a KF for each model. The final IMM-KF estimate for the polynomial coefficients is constructed as a weighted sum of estimates from

each model weighed by the corresponding model probabilities. Using the IMM-KF and without requiring pre-processing the ECG data, either to determine filter initialization parameters or to delineate the various segments, we were able to closely track real ECG signals of several morphologies, including those with abrupt changes such as PVCs. Such signals were not tracked using the statistical framework presented in [24, 32] using Gaussian functions due to phase-wrapping of each ECG beat.

(d) **ECG signal modeling using multiple models with the SMCMC algorithm:** The IMM-KF method however requires knowledge of the mode transition probabilities and its performance was found to be somewhat sensitive to those parameters. A second new approach to modeling ECG signals using SMCMC with simultaneous model selection was also presented. Similar to the IMM-KF model, the ECG signal is represented using linear, quadratic and cubic polynomial models. The model parameters represented by the polynomial coefficients are assumed to be constant over a given number of ECG samples (designated as a segment), and the parameter and model estimates which are obtained by the end of a segment (assumed to be the best estimates) are used to represent the samples in the segment. In addition, the ECG data is adaptively delineated into segments based on the morphology, by monitoring the value of the model likelihood function. Using the SMCMC approach the tracking of several different types of morphologies, including any abruptly occurring beats was demonstrated. The model selection ability of the algorithm using different types of ECG signals was also shown.

In addition, the proposed algorithms outperformed the estimation RMSE performance of the Gaussian ECG model in [24, 32], due to their ability to track intermittently occurring beats such as PVCs. The use of multiple models was also further substantiated by showing that the estimation RMSE of the al-

gorithms working with simultaneous model selection is much better than the estimation RMSE of the algorithms using a single fixed model for ECG signal representation.

(e) **ECG classification using parameter estimates from IMM-KF and SMCMC algorithms:** Classification was performed using features obtained from both the IMM-KF and SMCMC filters with a Bayesian ML classifier. While the $a_{k,1}$ polynomial coefficient representing the slope was used to extract the feature set from the IMM-KF model estimates, in the case of the SMCMC filter the reconstructed or estimated measurements were used. Results for classifying five ECG signal types, including, normal sinus rhythm signals and left bundle branch block, right bundle block, ventricular escape and junctional escape arrhythmias, were highly promising with an average correct classification rate of 98.5%, for both models. Comparison of these classification results with those presented in [8] and [9] using Hermite polynomials in conjunction with fuzzy-hybrid neural networks and support vector machines, revealed a comparable classification rate for all classes, with a fairly high improvement for the classification of junctional escape (j) beats. This was due to the fact that although the proposed classifier used a smaller dimension feature set, the information about the P wave, which is absent in such type of arrhythmias, was considered.

(f) **ECG clustering using adaptive learning DP framework and cluster labeling using Bayes ML method:** In order to preclude the dependence of ECG algorithms on huge amounts of training data and to preserve the inter-patient variability between ECG recordings from different patients, which is lost due to the use supervised learning algorithms for ECG signal classification, an adaptive learning method based on the DP framework is presented. In this method, firstly, the ECG data from each patient is individually clustered

using the DP algorithm. Since the DP algorithm does not provide labels for the clusters, the performance of the algorithm is first reported assuming the availability of expert knowledge, wherein the label assigned to a cluster corresponds to the label given to the dominant beat of the cluster by the expert. It is seen that the DP clustering algorithm performs fairly well correctly clustering 98% of all the considered beats, excluding the F, S and Q type beats. The performance of the algorithm was less accurate in the case of F type beats because their morphology closely resembled those of N and V type beats. Also, in the case of S and Q type beats, the lesser accuracy was due to the fact that these beats represented a very less percentage of the total number of available beats. Later, the application of a supervised Bayes ML method for labeling the clusters is demonstrated for situations wherein expert knowledge might be unavailable. It was seen that the classifier correctly labeled 98.3% of the beats (S and Q type beats were not labeled because of non-availability of sufficient training data). Using the clustering approach thus prevents the loss of inter-patient variability. The algorithm can be made patient-specific by performing the labeling using the clusters specific to each patient. The performance of the proposed DP clustering algorithm is also shown to compare favorably with the performance of the algorithms in [46, 50, 52] using the performance measures defined in (7.33) - (7.36). However, the DP algorithm offers the flexibility to adaptively learn the number of clusters from the ECG data without relying on an *a priori* knowledge about the number of diseases present in the given ECG data.

8.2 Future work

(a) It was assumed that the process noise in the proposed ECG models was white Gaussian noise. This was a sufficient assumption because the real ECG signals obtained from the MIT-BIH arrhythmia database did not contain manifesta-

tions of noise due to electrode movement and muscle artifact [89]. If these manifestations arise in the ECG data, the process noise can no longer be white Gaussian. For example, in [32], muscle artifact noise was modeled as colored noise by altering the slope parameter of the noise spectral density function. These scenarios present possible extensions to our proposed models.

(b) An extension to the multiple model framework proposed in this work can be made by including models that work with different measurement noise variances. This can ensure the choice of the best model even when the signal-to-noise ratio is reduced. For example, if the amount of noise in the ECG signal increases suddenly for a short duration of time, the selected measurement noise variance of the model might turn out to be lesser than the actual amount of noise present in the data, and this might lead to undesirable tracking results. This can be possibly avoided by using adding an extra degree of freedom in the model selection process, which enables the selection of the model with the appropriate amount of model noise.

(c) The features obtained from the SMCMC filter with simultaneous model selection, both for classification of ECG signals using the Bayes ML classifier and clustering of ECG signals using the DP algorithm are based on the noise-free reconstructed ECG signals. This is because the algorithm adaptively creates segments of data to work with the assumption of static parameters within a window. Thus, at the beginning of each segment all the model parameters including the point of reference for data samples are reinitialized. So, in order to use the polynomial coefficient estimates from different segments as features, estimates have to be adjusted to a common point of time reference, such as the beginning of the beat. However, by performing this adjustment to the reference time, and using the polynomial coefficient estimates as the features, additional distinguishing features may be obtained, and this might even lead

to a smaller and more efficient feature set.

(d) The DP algorithm used in this work employs a GMM and thus fits a Gaussian distribution to each of the clusters. However, the use of other distributions such as exponential or log-normal distributions can also be investigated.

(e) The clusters in the ECG data generated by the DP algorithm are labeled automatically using the Bayes ML method when expert knowledge is assumed to be unavailable. The Bayes ML method uses the estimated cluster means from the DP algorithm to identify the different clusters. However, the DP algorithm also estimates the cluster covariances. This information can also used in cluster identification by employing a method such as Kullback-Leibler divergence to label the clusters. The use of a more powerful supervised algorithm such as hidden Markov models etc. can also be explored. A combination of expert knowledge and supervised learning can also be used to identify or label the clusters. This would be especially helpful when enough training data is not available to train the supervised learning algorithm, for certain beat types. For such beat types, expert knowledge can be used to label the corresponding clusters, whereas the supervised learning algorithm can be employed for other beat types for which sufficient training data is available.

REFERENCES

[1] "IVLine," http://www.ivline.info/2010/05/quick-guide-to-ecg.html, 2010.

[2] U. Acharya, J. Suri, J. Spaan, and S. Krishnan, *Advances in Cardiac Signal Processing*. Springer-Verlang, 2007.

[3] T. Y. Young and W. H. Huggins, "On the representation of Electrocardiograms," *IEEE Transactions on Biomedical Engineering*, vol. 10, no. 3, pp. 86–95, July 1963.

[4] S. Karlsson, "Representation of ECG records by Karhunen-Loeve expansion," in *International Conference on Medical and Biological Engineering*, 1967, p. 105.

[5] N. Ahmed, P. J. Milne, and S. G. Harris, "Electrocardiographic data compression via orthogonal transforms," *IEEE Transactions on Biomedical Engineering*, vol. 22, no. 6, pp. 484–487, November 1975.

[6] L. G. Herrera-Bendezu and B. G. Denys, "Feature identification of ECG waveforms via orthonormal functions," in *Computers in Cardiology*, September 1990, pp. 641–644.

[7] L. Sornmo, P. O. Borjesson, M.-E. Nygards, and O. Pahlm, "A method for evaluation of QRS shape features using a mathematical model for the ECG," *IEEE Transactions on Biomedical Engineering*, vol. 28, pp. 713–717, October 1981.

[8] S. Osowski and T. H. Linh, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Transactions on Biomedical Engineering*, vol. 48, pp. 1265–1271, November 2001.

[9] S. Osowski, L. T. Hoai, and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 582–589, April 2004.

[10] A. Ahmadian, S. Karimifard, H. Sadoughi, and M. Abdoli, "An efficient piecewise modeling of ECG signals based on Hermitian basis functions," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, August 2007, pp. 3180–3183.

[11] W. Philips, "ECG data compression with time-warped polynomials," *IEEE Transactions on Biomedical Engineering*, vol. 40, no. 11, pp. 1095–1101, 1993.

[12] J. P. Marques de Sá and C. Abreu-Lima, "Optimal autoregressive modeling of ECG signals," in *Computers in Cardiology*, 1987, pp. 147–149.

[13] L. Kang-Ping and W. Chang, "QRS feature extraction using linear prediction," *IEEE Transactions on Biomedical Engineering*, vol. 36, pp. 1050–1055, October 1989.

[14] Z. Li and M. Ma, "ECG modeling with DFG," in *Proc. Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society*, Jan. 2005, pp. 2691–2694.

[15] M. P. S. Chawla, H. K. Verma, and V. Kumar, "ECG modeling and QRS detection using principal component analysis," in *International Conference on Advances in Medical, Signal and Information Processing*, July 2006, pp. 1–4.

[16] R. Dubois, P. Roussel, M. Vaglio, F. Extramiana, F. Badilini, P. Maison-Blanche, and G. Dreyfus, "Efficient modeling of ECG waves for morphology tracking," in *Computers in Cardiology*, September 2009, pp. 313–316.

[17] R. Borsali, A. Nait-Ali, and J. Lemoine, "ECG compression using ensemble polynomial modeling: Comparison with the wavelet-based technique," *Biomedical Engineering*, vol. 39, pp. 138–142, May 2005.

[18] S. Jokić, S. Krčo, V. Delič, D. Sakač, I. Jokič, and Z. Lukič, "An efficient ECG modeling for heartbeat classification," in *Symposium on Neural Network Applications in Electrical Engineering*, September 2010, pp. 73–76.

[19] S. Jokić, V. Delič, Z. Peric, S. Krčo, and D. Sakač, "Efficient ECG modeling using polynomial functions," *Elektron Elektrotech*, vol. 110, 2011.

[20] R. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME*, vol. 82, pp. 35–45, 1960.

[21] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.

[22] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, February 2002.

[23] R. Sameni, M. Shamsollahi, C. Jutten, and M. Babaie-Zadeh, "Filtering noisy ECG signals using the extended Kalman filter based on a modified dynamic ECG model," in *Computers in Cardiology*, 2005, pp. 1017–1020.

[24] O. Sayadi and M. Shamsollahi, "ECG denoising and compression using a modified extended Kalman filter structure," *IEEE Transactions on Biomedical Engineering*, vol. 55, pp. 2240–2248, 2008.

[25] M. Mneimneh, E. Yaz, M. Johnson, and R. Povinelli, "An adaptive Kalman filter for removing baseline wandering in ECG signals," in *Computers in Cardiology*, 2006, pp. 253–256.

[26] M. Ebrahim, J. Feldman, and I. Bar-Kana, "A robust sensor fusion method for heart rate estimation," *Journal of Clinical Monitoring and Computing*, vol. 13, pp. 385–393, November 1997.

[27] Q. Li, R. G. Mark, and G. D. Clifford, "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter," *Physiological Measurement*, vol. 29, no. 2, pp. 15–32, January 2008.

[28] M. Tarvainen, S. Georgiadis, P. Ranta-Aho, and P. Karjalainen, "Time-varying analysis of heart rate variability signals with a Kalman smoother alorithm," *Physiological Measurement*, vol. 27, no. 3, pp. 225–239, March 2006.

[29] J. McNames and M. Aboy, "Statistical modeling of cardiovascular signals and parameter estimation based on the extended Kalman filter," *IEEE Transactions on Biomedical Engineering*, vol. 55, pp. 119–129, January 2008.

[30] S. Kim, L. Holmstrom, and J. McNames, "Multiharmonic tracking using marginalized particle filters," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, August 2008, pp. 29–33.

[31] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Transactions on Biomedical Engineering*, vol. 50, pp. 289–294, March 2003.

[32] R. Sameni, M. B. Shamsollahi, C. Jutten, and G. D. Clifford, "A nonlinear Bayesian filtering framework for ECG denoising," *IEEE Transactions on Biomedical Engineering*, vol. 54, pp. 2172–2185, December 2007.

[33] S. Edla, J. Zhang, J. Spanias, N. Kovvali, A. Papandreou-Suppappola, and C. Chakrabarti, "Adaptive parameter estimation of cardiovascular signals using sequential Bayesian techniques," in *Asilomar Conference on Signals, Systems and Computers*, November 2010, pp. 374–378.

[34] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting multiple model methods in target tracking: A survey," *IEEE Transactions on Aerospace Electronic Systems*, vol. 34, pp. 103–123, January 1998.

[35] S. Edla, N. Kovvali, and A. Papandreou-Suppappola, "Electrocardiogram signal modeling using interacting multiple models," in *Asilomar Conference on Signals, Systems and Computers*, November 2011, pp. 471–475.

[36] D. S. Lee and N. K. K. Chia, "A particle algorithm for sequential Bayesian parameter estimation and model selection," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 326–336, 2002.

[37] S. Edla, N. Kovvali, and A. Papandreou-Suppappola, "Sequential Markov chain Monte Carlo filter with simultaneous model selection for Electrocardiogram signal modeling," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, August-September 2012, pp. 4291–4294.

[38] World Health Organization, "World Health Organization. Cardiovascular Disease," http://www.who.int/cardiovascular_diseases/en/, 2012.

[39] P. Bozzola, G. Bortolan, C. Combi, F. Pinciroli, and C. BroHet, "A hybrid neuro-fuzzy system for ECG classification of myocardial infarction," in *Computers in Cardiology*, September 1996, pp. 241–244.

[40] C. De Capua, A. Meduri, and R. Morello, "A remote doctor for homecare and medical diagnoses on cardiac patients by an adaptive ECG analysis," in *IEEE International Workshop on Medical Measurements and Applications*, May 2009, pp. 31–36.

[41] M. Morabito, A. Macerata, A. Taddei, and C. Marchesi, "QRS morphological classification using artificial neural networks," in *Computers in Cardiology*, September 1991, pp. 181–184.

[42] R. Watrous and G. Towell, "A patient-adaptive neural network ECG patient monitoring algorithm," in *Proc. Computers in Cardiology*, Sep. 1995, pp. 229–232.

[43] M. Shahram and K. Nayebi, "ECG beat classification based on a cross-distance analysis," in *International Symposium on Signal Processing and its Applications*, vol. 1, 2001, pp. 234–237.

[44] H.-S. Chow, G. Moody, and R. Mark, "Detection of ventricular ectopic beats using neural networks," in *Computers in Cardiology*, October 1992, pp. 659–662.

[45] P. de Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 1196–1206, July 2004.

[46] P. de Chazal and R. Reilly, "A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2535–2543, December 2006.

[47] P. Tadejko and W. Rakowski, "Mathematical morphology based ECG feature extraction for the purpose of heartbeat classification," in *International Conference on Computer Information Systems and Industrial Management Applications*, June 2007, pp. 322–327.

[48] E. Pasolli and F. Melgani, "Active learning methods for Electrocardiographic signal classification," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 6, pp. 1405–1416, November 2010.

[49] M. Nait-Hamoud and A. Moussaoui, "Two novel methods for multiclass ECG arrhythmias classification based on PCA, fuzzy support vector machine andunbalanced clustering," in *International Conference Machine and Web Intelligence*, October 2010, pp. 140–145.

[50] Yu Hen Hu, S. Palreddy, and W. Tompkins, "A patient-adaptable ECG beat classifier using a mixture of experts approach," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 9, pp. 891–900, September 1997.

[51] L. Senhadji, G. Carrault, J. J. Bellanger, and G. Passariello, "Comparing wavelet transforms for recognizing cardiac patterns," *IEEE Engineering in Medicine and Biology Magazine*, vol. 14, no. 2, pp. 167–173, March–April 1995.

[52] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sornmo, "Clustering ECG complexes using Hermite functions and self-organizing maps," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 838–848, July 2000.

[53] M. Owis, A. Abou-Zied, A.-B. M. Youssef, and Y. M. Kadah, "Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 7, pp. 733–736, July 2002.

[54] D. Ge, N. Srinivasan, and S. M. Krishnan, "Cardiac arrhythmia classification using autoregressive modeling," *Biomedical Engineering Online*, vol. 1, 2002.

[55] L.-Y. Shyu, Y.-H. Wu, and W. Hu, "Using wavelet transform and fuzzy neural network for VPC detection from the holter ECG," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1269–1273, July 2004.

[56] O. T. Inan, L. Giovangrandi, and G. T. A. Kovacs, "Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2507–2515, December 2006.

[57] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ECG signals," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 5, pp. 1415–1426, May 2009.

[58] J. Wiens and J. Guttag, "Patient-adaptive ectopic beat classification using active learning," in *Computing in Cardiology*, September 2010, pp. 109–112.

[59] ——, "Patient-specific ventricular beat classification without patient-specific expert knowledge: A transfer learning approach," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, August–September 2011, pp. 5876–5879.

[60] M. Llamedo and J. Martinez, "An automatic patient-adapted ECG heartbeat classifier allowing expert assistance," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2312–2320, August 2012.

[61] M. Faezipour, A. Saeed, S. Bulusu, M. Nourani, H. Minn, and L. Tamil, "A patient-adaptive profiling scheme for ECG beat classification," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1153–1165, September 2010.

[62] O. Sayadi, M. Shamsollahi, and G. Clifford, "Robust detection of premature ventricular contractions using a wave-based Bayesian framework," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 2, pp. 353–362, February 2010.

[63] O. Sayadi and M. Shamsollahi, "Life-threatening arrhythmia verification in ICU patients using the joint cardiovascular dynamical model and a Bayesian filter," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 10, pp. 2748–2757, October 2011.

[64] M. Llamedo and J. Martinez, "Heartbeat classification using feature selection driven by database generalization criteria," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 616–625, March 2011.

[65] M. R. Risk, J. F. Sobh, and J. P. Saul, "Beat detection and classification of ECG using self organizing maps," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, October–November 1997, pp. 89–91.

[66] P. Tadejko and W. Rakowski, "Hybrid wavelet-mathematical morphology feature extraction for heartbeat classification," in *International Conference on "Computer as a Tool"*, September 2007, pp. 127–132.

[67] W. Jiang and S. G. Kong, "Block-based neural networks for personalized ECG signal classification," *IEEE Transactions on Neural Networks*, vol. 18, no. 6, pp. 1750–1761, November 2007.

[68] I. Christov, I. Jekova, and G. Bortolan, "Premature ventricular contraction classification by the kth nearest-neighbor rule," *Physiological Measurement*, vol. 29, no. 2, pp. 15–32, January 2008.

[69] W. Gersch, D. M. Eddy, and E. Dong, "Cardiac arrhythmia classification: A heart-beat interval-Markov chain approach," *Computational Biomedical Research*, vol. 3, no. 4, pp. 385–392, August 1970.

[70] L. Szilagyi, "Application of Kalman filter in cardiac arrhythmia detection," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 20, 1998, pp. 98–100.

[71] D. Coast, R. Stern, G. Cano, and S. Briller, "An approach to cardiac arrhythmia analysis using hidden Markov models," *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 9, pp. 826–836, September 1990.

[72] M. Thaler, *The Only* EKG *Book You'll Ever Need*. Lippincott Williams & Wilkins, 2010.

[73] D. Novák, D. Cueta-Frau, P. Micó Tormos, and L. Lhotská, "Number of arrhythmia beats determination in Holter electrocardiogram: How many clusters?" in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 3, September 2003, pp. 2845–2848.

[74] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics*, vol. 1, pp. 209–230, 1973.

[75] V. E. Beneš, "Exact finite-dimensional filters with certain diffusion non linear drift," *Stochastics*, vol. 5, pp. 65–92, 1981.

[76] F. E. Daum, "Beyond Kalman filters: practical design of nonlinear filters," pp. 252–262, 1995.

[77] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall, 1979.

[78] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, pp. 465–479, 1971.

[79] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Adaptive Systems for Signal Processing, Communications, and Control Symposium*, 2000, pp. 153–158.

[80] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC, 1996.

[81] H. Blom, "An efficient filter for abruptly changing systems," in *IEEE Conference on Decision and Control*, vol. 23, December 1984, pp. 656–658.

[82] H. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Transactions on Automatic Control*, vol. 33, no. 8, pp. 780–783, August 1988.

[83] Y. Boers and J. Driessen, "Interacting multiple model particle filter," *IEE Proceedings on Radar, Sonar & Navigation*, vol. 150, pp. 344–349, October 2003.

[84] G. Ackerson and K. Fu, "On state estimation in switching environments," *IEEE Transactions on Automatic Control*, vol. 15, no. 1, pp. 10–17, February 1970.

[85] C. B. Chang and M. Athans, "State estimation for discrete systems with switching parameters," *IEEE Transactions on Aerospace Electronic Systems*, vol. 14, no. 3, pp. 418–425, May 1978.

[86] J. K. Tugnait, "Detection and estimation for abruptly changing systems," in *IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*, vol. 20, December 1981, pp. 1357–1362.

[87] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–732, 1995.

[88] F. Gustafsson, *Adaptive filtering and change detection*. John Wiley, 2000.

[89] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation Electronic Pages*, vol. 101, pp. 215–220, June 2000.

[90] C. Andrieu, A. Doucet, and E. Punskaya, "Sequential Monte Carlo methods for optimal filtering," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. deFreitas, and N. Gordon, Eds. Springer, 2001, pp. 79–95.

[91] N. Chopin, "A sequential particle filter method for static models," *Biometrika*, pp. 539–552, August 2002.

[92] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*. Prentice Hall PTR, 1998.

[93] R. Clayton, A. Murray, and R. Campbell, "Recognition of ventricular fibrillation using neural networks," *Medical and Biological Engineering and Computing*, vol. 32, pp. 217–220, 1994.

[94] S. J. L. Evans, H. Hastings, and M. M. Bodenheimer, "Differentiation of beats of ventricular and sinus origin using a self-training neural network," *Pacing and Clinical Electrophysiology*, vol. 17, no. 4, pp. 611–626, 1994.

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

[96] Association for the Advancement of Medical Instrumentation, *American National Standard for Cardiac Monitors, Heart Rate Meters, and Alarms (EC13-1983)*, ANSI/AAMI, Arlington, Va., 1984.

[97] M. I. Jordan, "Bayesian nonparametric learning: Expressive priors for intelligent systems," in *Heuristics, Probability and Causality: A Tribute to Judea Pearl*, R. Dechter, H. Geffner, and J. Halpern, Eds. College Publications, 2010.

[98] C. E. Antoniak, "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems," *Annals of Statistics*, vol. 2, pp. 1152–1174, 1974.

[99] M. D. Escobar and M. West, "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 577–588, June 1995.

[100] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "A sticky HDP-HMM with application to speaker diarization," *Annals of Applied Statistics*, June 2011.

[101] Y. Qi, J. Paisley, and L. Carin, "Music analysis using hidden Markov mixture models," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5209–5224, November 2007.

[102] D. Ting, G. Wang, M. Shapovalov, R. Mitra, M. I. Jordan, and R. L. Dunbrack, "Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical dirichlet process model," *PLoS Computational Biology*, vol. 6, no. 4, April 2010.

[103] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[104] D. G. S. Richard, O. Duda, and P. E. Hart, *Pattern Classification*. Wiley, 2001.

[105] H. Ishwaran and L. F. James, "Gibbs sampling methods for stick-breaking priors," *Journal of the American Statistical Association*, vol. 96, pp. 161–173, 2001.

[106] J. Sethuraman, "A constructive definition of Dirichlet priors," *Statistica Sinica*, vol. 4, pp. 639–650, 1994.

[107] D. Fink, "A compendium of conjugate priors," 1997. [Online]. Available: http://www.johndcook.com/CompendiumOfConjugatePriors.pdf

[108] P. Laguna, R. Jané, and P. Caminal, "Automatic detection of wave boundaries in multilead ECG signals: validation with the CSE database," *Computational Biomedical Research*, vol. 27, no. 1, pp. 45–60, February 1994.

[109] Association for the Advancement of Medical Instrumentation, *Recommended practice for testing and reporting performance results of ventricular arrhythmia detection algorithms*, ANSI/AAMI, Arlington, Va., 1987.

[110] ——, *Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms*, ANSI/AAMI, Arlington, Va., 1998.