Early detection of cardiac arrhythmia based on Bayesian methods from ECG data

Nasim Montazeri Ghahjaverestan

N° d'ordre : ANNÉE 2015

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

En Cotutelle Internationale avec
L'Université Technologique Sharif, Iran

pour le grade de
DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Traitement du Signal et Télécommunications

Ecole doctorale MATISSE
Ecole Doctorale Sciences De La Matière

Nasim MONTAZERI GHAHJAVERESTAN

Préparée à l'unité de recherche LTSI INSERM UMR 1099
Laboratoire Traitement du Signal et de l'Image
et au BiSIPL
Biomedical Signal and Image Processing Laboratory

Early detection of cardiac arrhythmia based on Bayesian methods from ECG data

Thèse soutenue à Rennes
le 10 juillet 2015

devant le jury composé de :

Raimon JANE
Dir. de Recherche. Univ. Polytechnique de Catalogne, Espagne/ rapporteur

Reza SAMENI
Assistant Professor, Univ. de Shiraz, Iran/ rapporteur

Anne HUMEAU
Prof. des Universités. Univ. d'Angers/examinateur

Emadeddin FATEMIZADEH
Assistant Professor, Univ. Technologique de Sharif, Iran/ examinateur

Alfredo HERNANDEZ
Directeur de Recherche INSERM/ directeur

Mohammad B. SHAMSOLLAHI
Prof. Des Universités. Univ. Technologique de Sharif, Iran/ directeur

Guy CARRAULT
Prof. des Universités. Univ. de Rennes 1/ co-directeur

Abstract

Apnea-bradycardia episodes (breathing pauses associated with a significant fall in heart rate) are the most common disease in preterm infants. Consequences associated with apnea-bradycardia episodes involve a compromise in oxygenation and tissue perfusion, a poor neuromotor prognosis at childhood and a predisposing factor to sudden-death syndrome in preterm newborns. It is therefore important that these episodes are recognized (early detected or predicted if possible), to start an appropriate treatment and to prevent the associated risks. In this thesis, we propose two Bayesian Network (BN) approaches (Markovian and Switching Kalman Filter) for the early detection of apnea bradycardia events on preterm infants, using different features extracted from electrocardiographic (ECG) recordings.

Concerning the Markovian approach, we propose new frameworks for two generalizations of the classical Hidden Markov Model (HMM). The first framework, Coupled Hidden Markov Model (CHMM), is accomplished by assigning a Markov chain (channel) to each dimension of observation and establishing a coupling among channels. The second framework, Coupled Hidden semi Markov Model (CHMM), combines the characteristics of Hidden semi Markov Model (HSMM) with the above-mentioned coupling concept. For each framework, we present appropriate recursions in order to use modified Forward-Backward (FB) algorithms to solve the learning and inference problems. The proposed learning algorithm is based on Maximum Likelihood (ML) criteria. Moreover, we propose two new switching Kalman Filter (SKF) based algorithms, called wave-based and R-based, to present an index for bradycardia detection from ECG. The wave-based algorithm is established based on McSarry's dynamical model for ECG beat generation which is used in an Extended Kalman filter algorithm in order to detect subtle changes in ECG sample by sample. We also propose a new SKF algorithm to model normal beats and those with bradycardia by two different AR processes.

We evaluate the performance of the proposed Markovian methods to detect event of interest using both simulated and real databases. In the case of simulated data, the performance of the proposed algorithms is evaluated in classification and detection procedures, in terms of confusion tables, sensitivity, specificity and time delay for detection task. The real signal database contains three feature time series extracted from raw ECG signals, acquired from preterm infants suffering from bradycardia. The proposed algorithms are evaluated in terms of the same metrics as in detection task of simulated data, which illustrate their ability in early detection of bradycardia episodes. The real ECG database is also applied in order to establish and assess the two Switching algorithms. The best results of AB detection precision are achieved by CHSMM (94.87% sensitivity and 96.52% specificity) and the lowest time delay is obtained by

using CHMM (0.73 s). Among methods in Switching approach, wave-based shows superior performance by 94.74% sensitivity, 94.17% specificity and 0.35 s time delay.

Résumé long en Français

L'apnée est une complication fréquente chez les nouveaux-nés prématurés, pouvant conduire à une diminution de la fréquence cardiaque (apnée bradycardie (AB)). La détection précoce de ces événements d'apnée-bardycardie peut réduire la durée d'hospitalisation, ainsi que les principaux effets secondaires liés aux thérapies complexes comme la réanimation ou l'intubation.

Position du problème clinique

La naissance prématurée de l'enfant arrive avant 37 semaines de gestation. Cette problématique est en augmentation constante dans la plupart des pays [Martin et al., 2010, Zeitlin, 2009]. Elle peut entraîner de nombreuses complications aux enfants en raison des dysfonctionnements des organes, qui ne sont pas complètement développés et apres pour la vie extra-utérine. L'un des problèmes le plus fréquent est l'épisode d'apnée-bradycardie, dont la répétition influence de manière négative le développement de l'enfant. Par conséquent, les enfants prématurés sont surveillés en continu par un système de monitoring installé dans les unités des soins intensifs néonatals (USIN). Ce système permet de déterminer l'évolution de l'état de santé de l'enfant et, depuis sa mise en marche, la qualité, l'espérance de vie et le pronostic de vie des prématurés ont été considérablement améliorés et la mortalité a été réduite.

En effet, les avancées technologiques en électronique, informatique et télécommunications ont conduit à l'élaboration de systèmes multivoies de monitoring néonatal de plus en plus performants. La plupart sont constitués de modules d'acquisition des données et d'une station centrale pour visualiser les signaux acquis en temps réel. Des alarmes sont produites lorsqu'une situation à risque est détectée, comme par exemple, une apnée-bradycardie. L'un des principaux signaux exploités dans ces systèmes est l'électrocardiogramme (ECG). L'ECG est un outil diagnostique qui mesure et enregistre l'activité électrique du cœur. Il reste aujourd'hui comme la technique la plus largement utilisée pour l'exploration de l'activité électrique cardiaque, puisqu'il s'agit d'une technique non invasive, simple, économique et sûre.

Même si l'analyse de l'ECG a évolué au fil des années et que le développement des méthodes de traitement des signaux électrocardiographiques a amélioré la caractérisation et le diagnostic des maladies cardiovasculaires, l'ensemble des informations fournies par l'ECG ne sont pas encore totalement exploitées dans les processus de décision, notamment en monitoring en USIN. Deux points particuliers sont à nous yeux sous exploités dans les stations de monitoring néonatal : i) l'exploration multi-variée des ondes et intervalles de l'ECG, afin de détecter précocement les épisodes d'apnée-

bradycardie et ii) les méthodes d'analyse restent des méthodes de détection et sont très peu anticipatives (prédiction de l'évènement).

Objectif principal de la thèse

Dans des travaux antérieurs [Dumont, 2008, Altuve, 2011], des méthodes pour répondre aux deux points mentionnées précédemment ont été proposées. Elles étaient fondées sur l'analyse de la *dynamique* des séries temporelles multivariées, extraites du signal ECG. Cette dynamique est analysée par l'utilisation de modèles de Markov cachés (MMC) [Rabiner, 1989a] et des modèles de semi-Markov cachés (MSMC) [Yu, 2010]. Le MMC est un modèle stochastique supposé avoir nombre fini d'états M avec la probabilité de générer l'observation (b_m). Le MSMC est similaire à un MMC classique, mais la différence principale est que le processus est semi-Markovien, dans le sens où un changement dans un état futur dépend à la fois l'état actuel, mais aussi du temps passé à cet état. Par conséquent, un paramètre aléatoire supplémentaire est défini dans les MSMC pour représenter la durée de stage dans l'état m, appelé le temps de séjour et désigné par d.

Les modèles MMC et MSMC ont été appliqués dans le domaine du traitement du signal pendant plus de deux décennies, en particulier dans le contexte de la reconnaissance automatique de la parole. Toutefois, ils constituent également des outils souples, flexibles et robustes pour le traitement des séries temporelles univariées et multivariées, y compris dans les cas d'observations discrètes ou continues. L'intérêt pour la théorie et les applications de ces modèles est en pleine expansion à d'autres domaines, par exemple: différents types de reconnaissance (visages, parole, geste, écriture, signature), bioinformatique (analyse des séquences biologiques, analyse de l'ECG, classification de l'électroencéphalogramme), environnement (direction du vent, précipitations, tremblements de terre) et finances (rentabilité journalière). Même si l'application de ces approches à la détection précoce des apnées-bradycardies a produit des résultats supérieurs aux autres méthodes de la littérature [Altuve et al., 2011b, Altuve et al., 2015], plusieurs limitations persistent dans la représentation des dynamiques multidimensionnelles. L'objectif principal de ce travail de thèse est donc d'améliorer la prise en compte des dynamiques multi-dimensionnelles en proposant des nouvelles approches basées sur un formalisme Bayesien, pour la détection précoce des apnées-bradycardies chez le nouveau-né prématuré.

Méthodes proposées

Dans cette thèse, nous proposons deux approches bayésiennes, basées sur les caractéristiques des signaux biologiques en vue de la détection précoce de l'apnée bradycardie des nouveaux-nés prématurés. D'abord avec l'approche de Markov caché, nous proposons deux extensions du MMC classique. La première, qui s'appelle Modèle de Markov caché couplé (MMCC), crée une chaîne de Markov à chaque dimension de l'observation et établit un couplage entre les chaînes. La seconde, qui s'appelle modèle semi-Markov caché couplé (MSMCC), combine les caractéristiques du modèle de MSMC avec le mécanisme de couplage entre canaux. Pour les deux nouveaux modèles (MMCC et

MSMCC), les algorithmes récursifs basés sur la version classique de Forward-Backward sont introduits pour résoudre les problèmes d'apprentissage et d'inférence dans le cas couplé.

En plus des modèles de Markov, nous proposons dans ce travail de thèse deux approches passées sur les filtres de Kalman (Switching Kalman Filter, SKF) pour la détection de l'apnée. La première utilise les modifications de la morphologie du complexe QRS et est inspirée du modèle générateur de McSharry, déjà utilisé en couplant avec un filtre de Kalman étendu dans le but de détecter des changements subtils de l'ECG, échantillon par échantillon. La deuxième utilise deux modèles AR (l'un pour le processus normal et l'autre pour processus de bradycardie). Les modèles AR sont appliqués sur la série RR, alors que le filtre de Kalman suit l'évolution des paramètres du modèle AR et fournit une mesure de probabilité des deux processus concurrents.

Evaluation et résultats expérimentaux

Les performances de l'ensemble de méthodes proposées est évalué sur des bases de données à la fois simulées et réelles. Concernant les bases simulées, des modèles dynamiques comme le modèle de FitzHugh-Nagumo [FitzHugh, 1961] sont utilisés. La performance des algorithmes proposés est évaluée pour la détection de perturbations et les critères clés sont les tableaux de contingence, la sensibilité, la spécificité et le délai de détection. Concernant les données réelles, elle est construite des signaux ECG des nouveau-nés prématurés souffrant d'apnée-bradycardie et suivis dans les unités de soins intensifs néonatals du CHU de Rennes. Pour chaque enregistrement de la base, trois indicateurs extraits de signaux ECG sont disponibles et représentés comme des séries temporelles, battement à battement (intervalle RR, durée du QRS et amplitude du QRS). Les annotations du début de chaque événement d'apnée-bradycardie sont également disponibles dans la base. Les algorithmes proposés sont appliqués sur ces séries temporelles et évalués en termes de sensibilité, spécificité, ainsi que le délai à la détection, ce qui permet de démontrer leur capacité dans la prévision des épisodes d'apnée-bradycardie. La même base de données ECG est appliquée afin d'établir et d'évaluer les deux algorithmes de SKF.

Concernant les approches Markoviennes, que se soit sur les données simulées ou réelles, les meilleurs résultats sont obtenus de l'application des méthodes MSMCC et MSMC, même à un nombre d'état relativement faible. En fait, une conclusion importante est que l'approche par modèles couplés facilite de manière notable la phase d'apprentissage des dynamiques caractéristiques de chaque événement et ce, d'autant plus que la dimension des observables est grande. En particulier, sur les données réelles, le MSMCC donne de meilleurs résultats que le MSMC, notamment en termes de délai à la détection, de spécificité et de la distance au point de détection parfaite. Ils est important de noter que le MSMCC est la seule approche à fournir de façon consistante les meilleures performances de détection (en termes de distance au point de détection parfaite) en fonction des caractéristiques utilisées en entrée et que les meilleures performances sont obtenues quand on utilise l'ensemble de séries disponibles (dimension 3). De plus, les simplifications et méthodes de calcul proposées ont des complexités similaires ou inférieures à celles décrites auparavant dans la littérature, ce qui ouvre la voie

vers une implémentation en ligne de nos méthodes. Ces contributions aux approches de Markov couplées, ont été publiées dans un article de conférence internationale et un article de journal international indexé ([2 et 3] dans liste ci-dessous).

En ce qui concerne les approches SKF, les principaux gains observés sont liés au délai à la détection, qui sont réduits par rapport aux modèles Markoviens, principalement dû à la simplicité de ces approches. Cependant, ce gain en temps de réponse vient au prix d'une plus faible performance. Globalement, les meilleurs résultats sont donc obtenus avec l'approche MSMCC proposée. Ces approches à base de SKF ont été récemment publiées dans un journal international indexé ([4] dans liste ci-dessous).

Perspectives

Comme travail futur, nous pouvons étudier l'utilisation clinique des méthodes proposées sur un plus grand nombre de cas et en utilisant d'autres signaux sources que l'ECG (par exemple les signaux respiratoires ou de saturation d'oxygène), pour la détection de l'AB. Nous pouvons également étudier d'autres événements pathologiques en plus des épisodes d'apnée-bradycardie. De plus, sachant que l'ensemble de méthodes proposées est générique, nous voulons appliquer ces méthodes sur d'autres applications cliniques, comme la détection précoce des apnées des adultes ou encore le suivi de données observées des dispositifs implacables de nouvelle génération.

Les méthodes proposées peuvent être encore étendues dans plusieurs aspects, comme par exemple l'amélioration des propriétés des paramètres de stage dans les MSMCC en utilisant des fonctions de densité de probabilités connues. De plus, nous pouvons améliorer la représentation des états en employant plus d'un noyeau Gaussien ou tiré d'une autre fonction de distribution.

Les travaux futurs concernant les approches SKF comprennent l'incorporation des méthodes d'estimation des paramètres AR, plutôt que l'approche ML, pour la procédure d'apprentissage dans les deux approches proposées. Dans l'approche basé sur les morphologies des complexes QRS, nous pouvons utiliser d'autres modèles dynamiques pour les tracés ECG [Ayatollahi et al., 2005,Das and Maharatna, 2013]. Enfin, nous pouvons aussi utiliser des observations multidimensionnelles, incluant la série RR et d'autres caractéristiques, comme entrée pour les approches SKF. Des modèles AR d'ordre plus elevé peuvent aboutir à une meilleure performance de détection, même s'ils requièrent une algorithmie plus complexe.

Publications

Ce travail de thèse a conduit aux publications suivantes :

1. N. Montazeri, M. B. Shamsollahi, G. Carrault, and A. I. HernÃ¡ndez, "Paroxysmal atrial fibrillation prediction using Kalman Filter," in Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies, 2011, p. 89.

2. S. Masoudi, N. Montazeri, M. Shamsollahi, D. Ge, A. Beuchee, P. Pladys, and A. Hernandez, "Early detection of apnea-bradycardia episodes in preterm infants

based on coupled hidden Markov model," in Signal Processing and Information Technology (ISSPIT), 2013 IEEE International Symposium on, 2013, pp. 000243-000248.

3. N. Montazeri, S. Masoudi, M. B. Shamsollahi, D. Ge, and A. I. Hernandez, "Coupled hidden Markov model based me[thod for apnea bradycardia detection", IEEE Journal of Biomedical and Health Informatics (J-BHI), - accepted and published in early print DOI : 10.1109/JBHI.2015.2405075.

4. N. Montazeri, M. B. Shamsollahi, D. Ge, and A. I. Hernandez, "Switching Kalman filter based methods for apnea bradycardia detection from ECG signals", Physiological Measurement, accepted, in press 2015.

Contents

Acronyms and notations

Acronyms

Models and methods

- **AC**: Accuracy
- **ApEn**: Approximated Entropy
- **AR**: Autoregressive
- **BN**: Bayesian Network
- **CHMM**: Coupled Hidden Markov Model
- **CHSMM**: Coupled Hidden Semi Markov Model
- **EKF**: Extended Kalman Filter
- **EM**: Expectation-Maximization
- **FB**: Forward-Backward
- **FN**: False-Negative
- **FP**: False-Positive
- **HAC**: Hierarchical Ascending Classification
- **HF**: High Frequency
- **HMM**: Hidden Markov Model
- **HSMM**: Hidden Semi Markov Model
- **KF**: Kalman Filter
- **KNN**: K-Nearest Neighbor
- **LF**: Low Frequency
- **MA**: Moving Average
- **MAP**: Maximum Aposteriory

- **ML**: Maximum Likelihood
- **MOB**: Mode on Beat
- **MOW**: Mode on Window
- **PCA**: Principal Components Analysis
- **PD**: Perfect Detection
- **PDF**: probability density function
- **PSD**: Power Spectral Density
- **ROC**: Receiver Operating Characteristic
- **SE**: Spectral Entropy
- **SEKF**: Switching Extended Kalman Filter
- **SEN**: Sensitivity
- **SKF**: Switching Kalman Filter
- **SPC**: Specificity
- **SSM**: State Space Model
- **SNR**: Signal to Noise Ratio
- **TN**: True-Negative
- **TP**: True-Positive
- **WLAR**: Maximum Likelihood Weighted Least Absolutes Residual

Medical Terms

- **AB**: Apnea Bradycardia
- **ANS**: Autonomic Nervous System
- **AP**: Action Potential
- **ARP**: Absolute Refractory Period
- **AV**: Atrioventricular
- **bpm**: beats per minute
- **ECG**: Electrocardiogram
- **HR**: Heart Rate

- **NICU**: Neonatal Intensive Care Unit
- **PF**: Pulse Frequency
- **PPG**: Photoplethysmography
- **RIP**: Respiratory Inductive Plethysmography
- **RRI**: RR Interval
- **SA**: Sinoatrial

Groups

- **BISIPL**: Biomedical Signal and Image Processing Laboratory
- **LTSI**: Laboratory of Signal and Image Processing

Chapter 1

Introduction

This thesis was developed under a joint supervision between two laboratories:

- The Laboratory of Biomedical Signal and Image Processing (BISIPL) of Sharif University of Technology.

- The Laboratory of Signal Processing and Image (LTSI) of the University of Rennes 1 where the recent projects conducted on analyzing the dynamics of time series by data mining [Dumont, 2008] and predicting the occurrence of apnea bradycardia (AB) in the preterm infants using methodologies seeking to exploit the dynamics of multivariate time series extracted from the electrocardiogram (ECG) [Altuve, 2011].

1.1 Problem definition

The premature birth of an infant happens before 37 weeks of gestation. This problem is increasing in most countries [Martin et al., 2010]; [Zeitlin, 2009] and can lead to many complications in children because of dysfunctions of organs, which are not fully developed and adapted for extra-uterine life. One of the most common problems is the episode of AB, whose repetition negatively influences the growth of the child [Pichler et al., 2003]; [Urlesberger et al., 1999]; [Janvier et al., 2004]. Therefore, preterm infants are continuously monitored by a monitoring system installed in the neonatal intensive care unit (NICU). This system allows to determine the evolution of health of an infant and consequently, the quality, life expectancy and prognosis of preterm life are significantly improved and mortality is reduced. Indeed, technological advances in electronics, information technology and telecommunications led to the development of neonatal monitoring multichannel systems. Most modules consist of data acquisition and a central station for inspecting the signals acquired in real time. Alarms are triggered when a risk situation such as AB is detected. One of the main monitored signals in these systems is ECG which is a diagnostic tool that measures and records the electrical activity of the heart and it is considered as the most widely used technique for exploring the performance of of heart, since it is non-invasive, simple, economic and safe.

Although the ECG analysis has evolved over the years and the treatment methods based on cardiac signals improved characterization and the diagnosis of cardiovascular diseases, all information provided by ECG are not yet fully exploited in the decision-

making process, including monitoring in the NICU. Two particular points can be considered in monitoring stations of neonates: i) multivariate exploration of waves and intervals of ECG to detect early episodes of AB and ii) improvement of the detection methods to predict an event.

In previous works [Dumont, 2008]; [Altuve, 2011], methods to address the two mentioned issues are proposed which were based on the use of the dynamic temporal multivariate series extracted from ECG. This dynamic is analyzed by the use of Hidden Markov Model (HMM) [Rabiner, 1989a] and Hidden semi Markov models (HSMM) [Yu, 2010]. HMM is a stochastic model assumed to have finite number of states M with the probability of generating the observation (b_m). HSMM is similar to a classic HMM, but the main difference is that the process is semi-Markov, in the sense that a change in a future state depends both the current hidden state and the time spent on this state. Hence, another random parameter for HSMM is defined for the remaining time duration in state m called sojourn time and notified as d. HMM and HSMM were used in the field of signal processing for more than two decades, particularly in the context of automatic recognition speech. Moreover, they are soft, flexible and robust tools for the processing of univariate (1-dimensional) and multivariate (having more than 1 dimension) time series, including the case of discrete or continuous observations. Interest in the theory and applications of these models are expanding to other areas, e.g. different types of recognition (faces, gestures, handwriting, signature), biomedical engineering (analysis of biological sequences, ECG analysis, classification of the electroencephalogram), environment (wind direction, rainfall, earthquakes) and finance (the daily profitability).

The main advantage of using HMM in the analysis of biomedical signals is based on the fact that: i) it can represent the time evolution of a variable through the standard parameters of HMM , ii) except initial values for model parameters, they do not require any prior knowledge about the data to be processed and iii) the number of states M is the only parameter which has to be set.

1.2 Subject of the thesis

In this work, the main purpose is the early detection of AB using various Bayesian Network (BN) [Ghahramani, 1998] methods including HMM and its generalizations. Expanding the pervious works in our group, we concentrate our study on specific framework for Coupled Hidden Semi Markov Model (CHSMM) [Natarajan and Nevatia, 2007a], which is a generalization of HMM integrating HSMM and Coupled Hidden Markov Model (CHMM) characteristics [Brand et al., 1997a]. However, it is essential firstly to establish new framework for CHMM. A CHMM represents a system with multivariate observations assigning a specific HMM to each of its dimension. The states of HMMs mutually interact in a way that a transition to a state in one channel at time t depends on states of all channels at time $t-1$. HSMMs can be replaced to HMMs in coupled structure in order to form a CHSMM framework. We primarily propose a framework for CHMM which is extended to CHSMM by considering parameter representing the sojourn time (d) of the states. The performance of the proposed frameworks

are compared to each other and to HMM and HSMM based methods in AB detection task.

Another common method in BN is State Space Model (SSM) and in special case of linearity, Kalman Filter (KF). In case of nonlinear dynamical model, Extended Kalman Filter (EKF) is applied for local linearization. Furthermore, the performance of KF is effectively dependent on the predefined knowledge of the state space coming along with its dynamic. Hence, their performance for modeling the systems with changeable dynamic is not reliable. On the other hand, as a solution, it is possible to model changeable dynamics by integrating a switch on as many KFs as the number of possible dynamics observed in the system, so that each one tracks one of the dynamics. This switch is a random variable whose status indicates to a KF and its corresponding dynamic. Allocating a HMM to the sequence of states of the switch, creates a combination of HMM and KF which is called Switching Kalman Filter (SKF) in literature [Murphy, 1998]. We also propose a (SKF) and a Switching Extended Kalman Filter (SEKF) based models for early detection of AB.

1.3 Overview of the thesis

This thesis is organized as follows:

- The chapter 2 summarizes anatomy and physiology of heart and the cardiac characteristics of AB. The physiological concepts of preterm infants and complications, including AB are also included. The basis of electrocardiography and ECG in preterm infants are detailed and the works that have already been proposed in this field are reviewed.

- The chapter 3 provides a brief literature review on strategies for classification and detection of Markov models including standard HMM, HSMM and coupling concepts in HMM. Then, our proposed CHMM and CHSMM are introduced.

- In chapter 4, the algorithms based on dynamical model of ECG, developed in our laboratory, are subsequently reviewed. Our proposed algorithms, SKF (R-based) and SEKF (wave-based), for modelling the changeable dynamic of cardiac system are presented.

- In chapter 5, the simulated and real database are introduced properly followed by description of evaluation methods and corresponding metrics.

- The chapter 6 details the results. First, a comparison on the results of various Markovian models are presented. Furthermore, the results of switching approach shows their effective contributions in AB detection.

- Finally, we summarize our findings and contributions in this thesis and suggests several fields of research for future works.

A list of publications related to this thesis is presented in Appendix C.

1.4 Main contributions

The main contributions of this thesis can be summarized as:

1. (a) the introduction of a novel framework for CHMM as a generalization of HMM to address the analysis of N-dimensional coupled signal features,

   (b) the derivation of the FB variables and solutions to the standard problems of the Markovian models according to the proposed framework,

   (c) the application of the proposed framework to AB detection, based on the log-likelihoods of competing models,

2. (a) Combination of HSMM characteristics with previously proposed CHMM framework to proposed CHSMM, a novel generalization of HMM, to address the analysis of N-dimensional coupled signal features,

   (b) the derivation of the related FB variables and solutions to the standard problems of the CHSMM according to the proposed framework,

   (c) the application of the proposed CHSMM to AB detection, and the comparison of the results with other existing methods,

3. (a) the introduction of a wave-based state space formulation according to SEKF for detecting bradycardia,

   (b) the derivation of a linear R-based model according to previously proposed SKF for detecting the bradycardia from RR signal, and

   (c) comparing the results achieved by wave-based and R-based methods with other identical previously presented methods.

Chapter 2

Apnea Bradycardia

2.1 Introduction

This chapter aims to give the physiological basis needed to understand our issue addressed as AB detection in premature newborns. Generalities of the cardiovascular system are presented in the next section including anatomical structure of heart, its electrical activity and system of excitation-conduction. Then, the definitions related to prematurity and neonatal apnea focusing on the causes, consequences and implications of premature birth and AB of prematurity are described. Finally, a review on methods for apnea detection especially the researches using ECG of preterm infants is presented to conclude the chapter.

2.2 Overview of the cardiovascular system

The cardiovascular system is a compound of the heart and blood vessels. This system ensures a continuous flow of blood to the organs and tissues of the body cell to supply them with oxygen and nutrients removal of metabolic products generated during their activity and transporting hormones produced by the endocrine glands to receptors. Heart as a vital organ of the circulatory system of warm-blooded living creatures is composed of highly complex structure. Naturally, in human being, according to its innate and vital requirements, its heart has the highest degree of evolution and cooperation between the organs of all living organisms. In order to know about the performance of this vital organ, it is essential to be familiar with its anatomy and physiological characteristics. Furthermore, the functionality of the heart in circulatory system to maintain bloodstream within the vessels can be reviewed.

2.2.1 Clinical anatomy of heart

Anatomically speaking, the human heart is a muscular organ in the chest cavity that functions as an independent pump, ensuring the progress of the blood within the vessels. It consists of four contractile cavities, two atria and two ventricles at the bottom of the atria. Ventricles are separated from each other by muscular walls. The left and right atria receive venous blood and are separated by the interatrial septum. The left and

right ventricles are divided by the interventricular septum, which ensures the expulsion of the blood in the pulmonary circulation, respectively and in the systemic circulation. However atrium and its below ventricle are related through atrioventricular valves, formed by fine connective tissue. The tricuspid valve between the atrium and the right ventricle and mitral valve separates the left atrium left ventricle, ensuring the blood to flow one way. The function of the atrioventricular valves is to prevent reentry of blood to the atria once it streams to the ventricles. There are also the sigmoid valves within pulmonary and right ventricle and also aortic artery and left ventricle, preventing the return of blood to the ventricles, once pumped to the pulmonary artery and to the aorta, respectively. The location of each elements of heart is illustrated in Fig. 2.1. The wall of the heart muscle is composed of three layers:

- The pericardium is the outer shell and the core consists of a layer of epithelial cells and connective tissue.
- The myocardium and the intermediate layer consists of predominantly of fibers contractile.
- The endocardium is the inside and consists of an additional layer epithelial cells and connective tissue.

Major part of contraction operation is performed by the myocardial muscles which consist principally of two different types of tissues. First one is the fabric nodal conduction. This tissue consists of cells having properties of automatic excitability and conductivity histologically. These characteristics allow regular and spontaneous generation of electrical impulses and the transmission of the pulses through the myocardium in an organized manner, to ensure adequate and efficient pumping contraction. The second type is contractile myocardial tissue which is forms the majority of myocardium and has properties of cellular excitability and conductivity to be capable of contracting. Unlike the nodal tissues, they are not able of being excited automatically. Hence these tissues requires to be triggered for contracture. The activity of the heart is quasi-periodic, called the cardiac cycle, pushing blood through the body continuously. The cardiac cycle consists of a period of ventricular relaxation (diastole), while the heart is filled with blood, followed by a period of ventricular contraction (systole), when the blood is directed out of the heart. By contracture of left and right ventricles, the most of the blood storage is discharged into aorta and pulmonary artery respectively. In the pulmonary circulation, blood uptakes the oxygen and excrete the carbon dioxide via the pulmonary veins and then returns to the left atrium. In general circulation, blood nourishes and refines the body cells. Hence it contains low degrees of oxygen and high degrees of carbon dioxide then, so it enters the right atrium through superior and inferior vena cava veins. During the relaxation of ventricles and following the loosening of atrioventricular valves due to the gradient pressure between their both sides, 3/4 of the amount of blood in atriums rushes to the corresponding ventricles. Hence, the appropriate performance of the these valves plays a crucial role in maintaining the proper functioning of heart in sync with its normal cycle [Fauci et al., 2008].

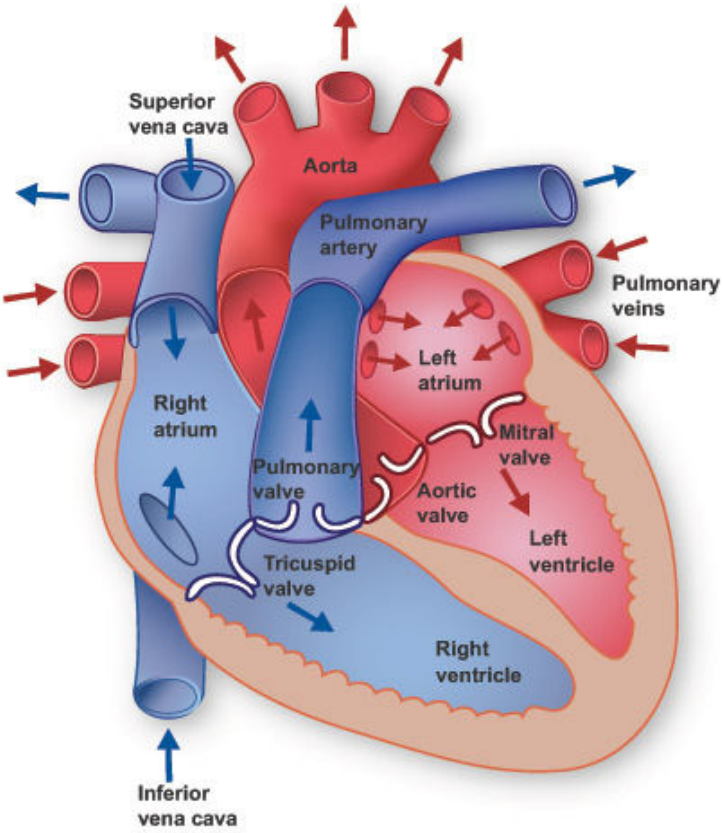

Figure 2.1: Heart anatomy. [THI, 2014]

2.2.2 Physiology of heart

Heart is a muscular organ capable of autonomous action which is specially influenced by higher nerve centers. Cardiac muscle, skeletal muscle, smooth muscle and neuronal tissue, histologically, possess tissues which are potentially capable of demonstrating an electrical potential difference between the two sides their cells membrane. The cardiac tissues are fine, capable of aerobic and long-term metabolism and tightly interconnected to autonomic efferent nerves (sympathetic and parasympathetic). Each cardiac cell (nodal or myocardial) has a two-layer membrane made of fat in between the layers, and several ionic valves. They are also surrounded and filled with a solution which contains ions. The three most important are: sodium (Na^+), potassium (K^+) and calcium (Ca^{2+}). Active and passive movements of ions through ionic channels, associated with crossing the cell membrane and their propagation form the basis of the cellular electrical activity. The ionic valves are able to selectively transmit the ions. Consequently, this feature leads to a difference in electrical potential on both sides at resting time of the cardiac cells. At rest, the interior of the cell membrane is electrically negative with respect to the extracellular environment, which is taken as reference. The potential is approximately -80 to -90 mv. Ions involved in the creation of the resting potential, mainly are sodium and potassium ions. When an electric pulse of sufficient magnitude, as a consequence of a nervous, mechanical or cell membrane stimulation, arrives at an excitable cell, the resting potential within the cell exhibits rapid polarity

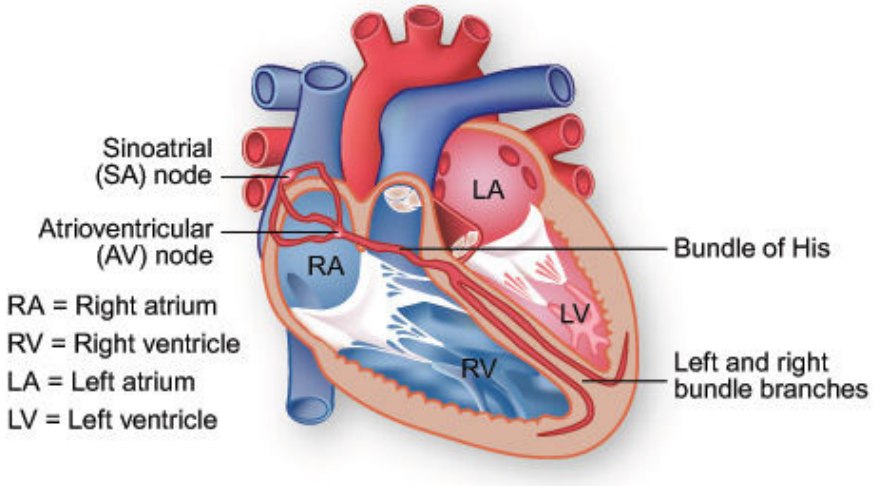

Figure 2.2: SA node: the natural pacemaker of the heart. AV node: transmitter of electrical stimulation to ventricles. [THI, 2014]

reversal and becomes positive relative to the outside. It increases to the point where it is called the potential (activation and opening) threshold. This process is known as cellular depolarization and it is generated by transmembrane opening which allows the passive entry of Na^+ in the intracellular space. At this level, the cell membrane potential reaches even to the 20+ to 30+ mv. This electric potential is called as action potential (AP). The return of the stimulated cardiac cell to its resting state is called cell repolarization which is consists of rapid and slow steps. Its first step is characterized by a rapid and short-time repolarization due to the inactivation of Na^+ channels and the outspreading of K^+. Slow voltage-sensitive calcium channels in intracellular open, allowing the passive diffusion of slow inward of Ca^{2+} ions into the cytoplasm cell of heart. Entered calcium ions sit on special positions, on Myosin fibers and cause the energy-consuming process of contraction in the muscle cell. Reconstructing the resting potential is begun by closing of specific ion channels and activating the potassium channels, which allows the expulsion of the active ions, facilitating the return to the original negative transmembrane potential. The cells that an external stimulation was incapable of causing a new AP in them till this stage, becomes excitable again. The time interval of non-excitability is called the absolute refractory period (ARP). The duration of ARP depends on the frequency at which the cell is stimulated. AP has the ability to move along the heart muscles. As mentioned, the cardiac cells are associated with the ability of ion transmission through intercellular channels. This implies that the action potential can spread rapidly to all connected muscle cells (Fig. 2.2).

The clinical representation of AP in heart muscle is the creation of contraction in the muscle. The start of the stimulation process and the contraction of cardiac muscles needs to have a special regulations, in order to generate a regular performance of heart for blood pumping throughout the body. In human heart, in the posterior region of the right atrium, the sinoatrial node (SA node) can be found among the specific muscles with special characteristics, that are constantly self-stimulating due to their low resting potential, which is equivalent to -60 mv. Hence, the fiber tissues of

this region, insistently generates the AP with velocity of nearly 80 beats per seconds and are tightly connected to the fibers of atrial tissues. Among these particular cells, SA node is called the dominant pacemaker of heart due to its rapid self-stimulation. SA node initiates the electrical excitation in atrial tissues through the links between ion channels of the atrial muscle. The muscle of atrium are electrically isolated from ventricles by a fibrous-connective tissue that acts as insulation. The only connection between the atria and ventricles in normal mode is the atrioventricular (AV) node which is a collection of heart muscle fibers called Purkinje that is able to conduct electrical signals. This activation is mediated at ventricle using specialized internodal pathways that connect the SA node to the AV node. AV node is located at the bottom of the right atrium and consisting of cells which exhibit a slow electrical conduction. The velocity of activation is physiologically slowed (approximately 100 ms) by AV node, before reaching to the His bundle. This property of AV node protects the ventricles from excessive number of activations of the AV node and activations diffused suddenly which optimizes ventricular contraction. Also, it reserves the regulation of pumping that the atria are activated before ventricles. The His bundle is located in the upper part of the interventricular septum and pass through the connective tissue fibers (not excitable) which electrically separate atria from the ventricles. In normal cases, the AV node is the only way of propagation of cardiac electrical activity between atria and ventricles. This is often called the atrioventricular junction. The His bundle includes a main trunk that divides into two branches, right to left. Purkinje fibers are the branches of His bundle which is ended in a network of fibers which reaches the ventricular walls. The Purkinje fibers terminate in anastomoses with myocardial muscle fibers, facilitating their excitement. The heart rate of a man lying is around 100-120 beats per minute (bpm).

2.2.3 ECG

The flow and the amplitude of the electric currents generated by depolarization and repolarization of the myocardial cells can all be detected by electrodes placed on the surface of the thorax. The electrical signal obtained is an ECG signal. The analysis of this electrical activity is proved to be an essential technique for the diagnosis of cardiovascular diseases and is a fundamental tool in cardiac monitoring. As the waves of depolarization and repolarization have direction and magnitude, they can be represented by a vector. The vectorial analysis of electrical activity of heart shows the main concepts of ECG. This means that the ECG is a complex spatial and temporal mapping of electrical potential of multiple myocardial fibers that are directed to the surface of the body and is recorded by surface electrodes. This implies that the activity of some regions can be weakened or discarded from the recorded signal. In order to have complete outlook of heart performance, it is common to record ECG from various places on the chest called leads.

The electrical activity generates special waveform patterns in ECG that are summarized as follows. ECG patterns starts with P waves which represents atrial depolarization. This wave characterizes by a low frequency and energy component which often its observation is limited to a few ECG leads, especially in noise conditions. Then, atrial repolarization is represented by Ta wave in opposite direction of P wave. Generally Ta

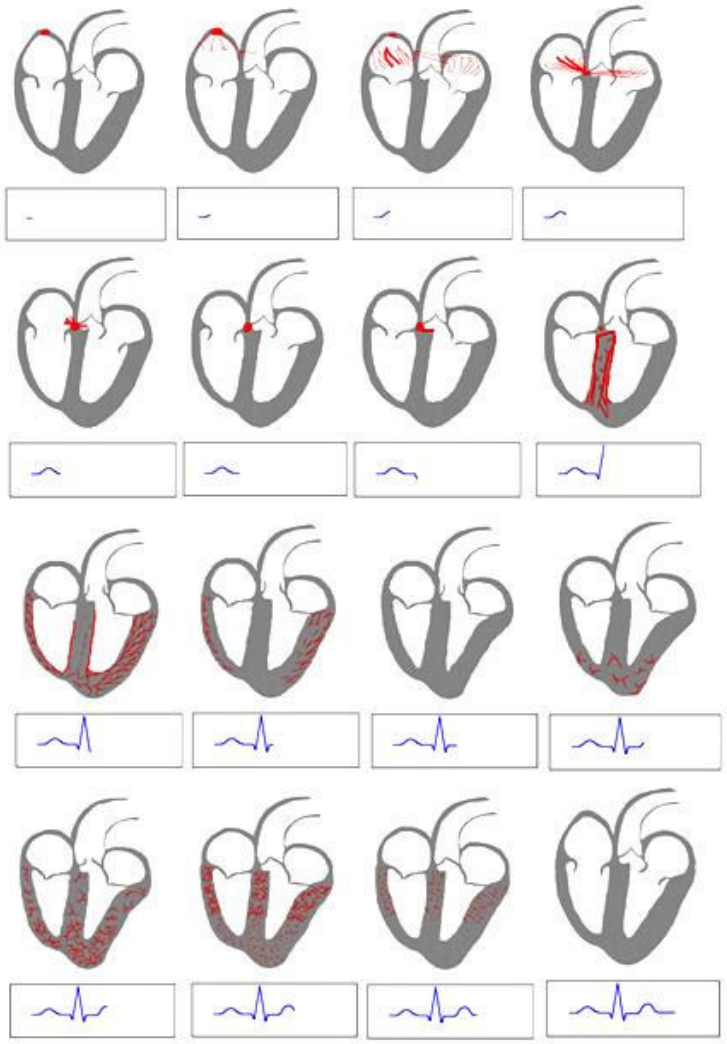

Figure 2.3: Conduction of electrical stimulation as the origin of ECG waveform generation. From left to right and up to bottom. [Wikipedia, 2011]

wave is not visible in the ECG as it coincides with the largest amplitude QRS complex. Ventricular depolarization is then represented by a deflection of the amplitude of ECG called as QRS complex. This complex consists of three consecutive waves: Q, R and S, which are respectively associated with the means of activation of different parts of ventricles. Ventricular repolarization is reflected by T wave Fig. 2.3.

In addition to the morphology of these waveforms, the distance between their locations in an ECG beat is another characterization. The RR interval: it is the elapsing time between the peaks of two consecutive R waves. The reverse of RR interval is equal to the spontaneous HR. PR interval: this is measured from the beginning of the P wave and the beginning of the QRS complex. This interval represents the time of conduction of the electrical activity of atria to the ventricles and the time of transmission of the front depolarization by the AV node. ST segment is defined as the time between the end of the QRS complex (or J point) and beginning of the rising phase of the wave T.

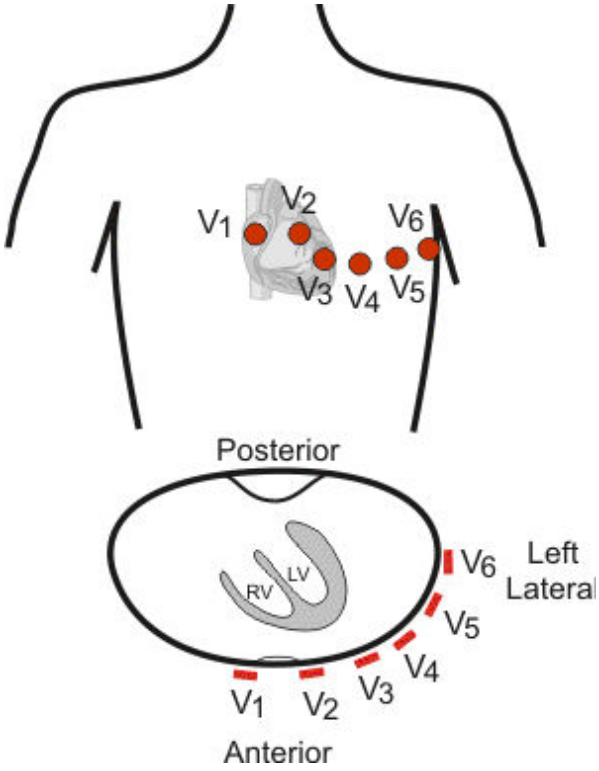

Figure 2.4: Conduction of electrical stimulation as the origin of ECG waveform generation. From left to right and up to bottom. [Klabunde, 2008]

This corresponds to the time segment wherein all the myocardial cells are depolarized. The QT interval is the time between the beginning of the QRS complex and the end of T wave. This wave is an indication of the length of the phases of the depolarization and ventricular repolarization. The QT interval varies with heart rate.

12 common ECG leads records the difference potential between the electrodes placed on the surface of the body. This leads are divided into two groups: 6 organ leads and 6 thoracic leads. Organ leads are divided in 3 bipolar (I, II, III) and three unipolar leads (aVR, aVL, aVF). Each of the dipole leads measures the differential potential between two electrodes placed on the organs that can be identified in table 2.1. 6 Chest leads are also recorded unipolar which are shown in Fig. 2.4.

ECG leads are placed in a way that if the depolarization spreads to their positive pole (Fig. 2.5), a positive deviation (upward) will be recorded and vice versa. If the axis of a depolarization vector is perpendicular to plane of the electrode, it creates a two-phase deviation (equal amounts of positive and negative). Normal atrial depolarization vector is towards the bottom left and reflects the spread of the SA node stimulation to the myocardium of the right atrium and then left atrium. Since the direction of this vector is toward the positive pole lead II and negative polarity of of aVR lead, P wave would be positive and negative in these leads respectively. In the case of an ectopic pacemaker in the lower part of the atrium or in the ventricle and atrium connection, a negative P waves in lead II or positive P waves in aVR lead can be observed. The cardiovascular system which appears in early embryological development and continues

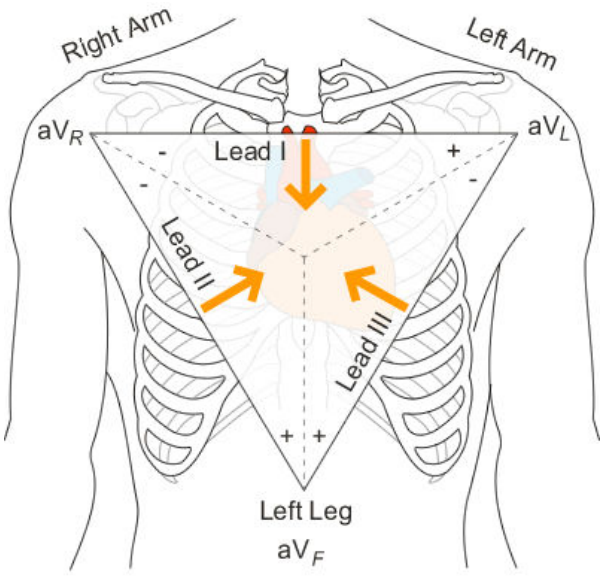

Figure 2.5: Conduction of electrical stimulation as the origin of ECG waveform generation. From left to right and up to bottom. [Nottingham, 2014]

Table 2.1: Organ Leads

Unipolar		Bipolar	
right hand	aVR	right/left hands	I
left hand	aVL	right hand/left foot	II
left foot	aVF	left hand/foot	III

to develop through the remaining fetal life, remains immature and is intended to continues the development for several weeks after term birth. In addition to maturation of the cardiovascular system after birth, a critical changes happen in the structure of cardiovascular system whereby circulatory shunts, including the ductus arteriosus, ductus venosus and foramen ovale close and transform the system from having a placental oxygen source to a pulmonary source [Hall, 2010]. In the premature infant the closures of shunts may not be complete, remaining a transition channel between the right and left ventricles of the heart. In this case, the blood flow in the aorta can return to the pulmonary artery and cause an increase in flow heart [Dageville, 2011]. However, the premature myocardium contains less contractile elements and is more compliant than the adult. Higher levels of HR, thin rib cage and the undeveloped cardio-respiratory system in preterm neonates produce very different ECG compared to those of normal adults. The most differences are notable as: 1- Intervals (RR, PR and QT) and durations of waveforms are shorter; 2- The size of the right ventricle is thicker than the left ventricle in newborns while the left ventricle is much thicker in adults; 3- R-waves has greater amplitude in the front aVR lead and V4, V1 and V2 leads; 4- S waves are deeper in the frontal branch of I lead and in precordial V5 and V6 leads; 5- The QT interval is prolonged in preterm infants; 6- The P-wave has greater amplitude in leads II and V1 [Schwartz et al., 2002].

2.3 Apnea bradycardia in preterm infants

Reviewing the functionality of heart and ECG generation phenomenon in previous section, we can approach to the concept of prematurity in newborn infants and study about apnea as one of its complication. This section also includes definition of bradycardia and its relation to apnea occurrence in preterm infants.

2.3.1 The preterm infants

Biologically, a fetus requires a number of weeks in the maternal uterus that its organs becomes ready for extrauterine life. If the infant is born before completing this cycle, it may encountered with an irritating environment. Premature birth is defined as any birth that occurs before 37 weeks counted from the first day of the last menstrual period. More than 11.5% of babies born in the U.S. each year are premature [Martin et al., 2013] and nearly 6% (nearly 55,000) of the births are premature in France. The factors of maturity in preterm infants are expressed by gestational age and birth weight.

Prematurity is one the leading causes of neonatal mortality. In 2009, 35.4% of newborn infant deaths in U.S. were preterm related [Mathews and MacDorman, 2013]. Advances in medicine and caring of premature infants in NICU improve their chances of survival [Martin and Fabes, 2008]. Indeed, the number of children born prematurely is increasing in France and worldwide. For example, in France, a study recently shows that the number of premature births has increased from 4.5% to 5.5% in singleton live birth and 40.5% to 42.1% in multiple live birth between 1997 and 2008 [Zeitlin et al., 2013]. The earlier the child is born, the more serious and frequent the problems of adaptation are and also the treatment would be more complex. Common problems are:

- Respiratory: respiratory distress syndrome, hyaline membrane disease and bronchopulmonary dysplasia,
- Neurological: apnea of prematurity, retinopathy, cerebral palsy and intraventricular cerebral hemorrhage
- Cardiovascular: patent ductus arteriosus,
- Infectious: sepsis, pneumonia, and urinary tract infection,
- Hematologic: anemia, jaundice and thrombocytopenia,
- Metabolic: hypoglycemia, hypocalcemia and hypothermia,
- Gastrointestinal: necrotizing enterocolitis and gastric residuals.

2.3.2 Apnea in preterm infants

Apnea is one of the most important complications of prematurity. Many researches have been conducted over recent years to try to explain the causes, consequences and identifying better treatments ([Rigatto and Brady, 1972], [Kurth et al., 1987], [Di Fiore et al.,

2001], [Henderson-Smart and Steer, 2004], [Zhao et al., 2011] [Chen, 2013], [Manzionna and Di Mauro, 2014] and [Haskova et al., 2013a]).

Apnea is defined as a prolonged interruption of ventilatory flow for more than 15-20 s associated with the instability of controlling rate of breathing, falling blood oxygen concentration (SpO_2 80% for 0.4 s) and bradycardia (heart rate $< 2/3$ of baseline for 0.4 s) in severe conditions in the preterm infant [Zhao et al., 2011]. Apnea of prematurity is found in 50% of premature infants and is almost universal in infants who are 1000 g at the time of birth [Finer et al., 2006]. Attacks of apnea are common consequences of premature birth, with the most severe essentiality of artificial ventilation and drug treatment. Preemies are supported in a NICU where they are kept in a incubator, which reproduces the conditions of fetal development by controlling temperature, humidity and limited exposure from the environment outside (virus, bacteria, ...). Continuous monitoring is carried out on vital parameters such as temperature, respiration, cardiac function, oxygenation, and brain activity. Therapies may include intravenous nutrition catheters, oxygen, mechanical ventilation, and supplemental medicines. The main cause of apnea incidence is known to be the immaturity of central control of respiration [Gerhardt and Bancalari, 1984]. Other factors which are implicated in apnea of prematurity are unregulated inhibitory neurotransmitters, decreased central chemosensitivity and some diseases like sepsis, cytokinesis and increased levels of bilirubin [Abu-Shaweesh and Martin, 2008].

Alternative theory explaining apnea of prematurity is the Polyvagal theory [Porges, 1995], which proposes a model for explaining the incidence of apnea as a result of changes in heart rate (HR). In normal and mature brainstem, vagal stimulation causes a rise in HR and respiratory sinus arrhythmia. However, in case of immaturity of the brainstem, increase in vagal tone may result in bradycardia, which in turn causes a decrease in SpO_2, and a decrease reflex in respiration rate. Hence dysfunction in the neural regulation of cardiopulmonary function results in apnea [Porges, 1996].

Apnea may be symptomatic type if it comes from a pathology (infection, anemia, hypoxia, hypoglycemia, hypothermia, etc.), idiopathic or if it is only due to the immaturity of the mechanisms controlling the breathing. Furthermore, based on the presence or absence of respiratory movements, apnea is classified into three categories:

- Central: occurring in the absence of respiratory movement;
- Obstructive: occurring in the presence of respiratory movement;
- Mixed: containing central and obstructive apnea elements.

The clinical observation of a preemie or continuous cardiorespiratory monitoring allows to detect only the central apnea unlike obstructive apnea which can be only detected by observing associated phenomena, such as bradycardia and hypoxemia. The longest and most frequent apnea is often mixed (50 to 70% of cases). Short apnea (less than ten seconds) is usually central type (10 to 25% of cases) and obstructive infrequent (12 to 20% case) [Laugel et al., 2000].

2.3.3 The apnea-bradycardia

The SA node is able to generate impulses at a rate of about 80 bpm in normal mature man and the stimulation signal is transmitted to the ventricles. Vagus nerve stimulation associated with increased parasympathetic stimulation (drug use) and also in the presence of heart block, ventricular rate and therefore reduced vestibular signals, lead to such complication. The HR less than 60 beats per minute is called bradycardia. Bradycardia in newborn infants is defined by a drop in HR, below 100 bpm or at least 33% from an average value, for 4 seconds or more [Poets et al., 1993]. Bradycardia usually occurs within the first seven seconds of apnea [Dorostkar et al., 2005]. The majority of preterm infants have AB during the first weeks. However, they gradually disappear during the maturation and usually after 40 weeks.

Short apnea are considered as non-pathological which do not usually accompany with bradycardia or hypoxemia. In contrast, those over 20 seconds are considered serious because they are followed by a drop in saturation of oxygen in blood, resulting to bradycardia due to hypoxemia, whose depth is proportional to the duration of the apnea. AB of prematurity is interpreted as it reflects a form of incompatibility of preterm infant to life out of uterine life.

2.4 Review on AB detection works

The first researches are conducted to detect apnea caused during sleep in adults. Most of them use respiration signals due to its simplicity. One of these researches is proposed in [Taha et al., 1997] where pulse oximetry is incorporated into the definitions of apnea and hypopnea. In proposed algorithm, the detection of desaturation as a fall in oxyhemoglobin saturation level has been achieved in order to separate apnea and hypopnea. It would be an apnea, if there is a period of no breathing, as indicated by sum of respiratory inductive plethysmography (RIP), lasting at least 10 seconds and coincident with the desaturation event. If there is breathing, a hypopnea is defined as a minimum of three breaths showing at least 20% reduction in sum of RIP magnitude from the immediately preceding breath followed by a return to at least 90% of "baseline" breath. Another work employs nasal flow apnea monitoring and pulse oximetry data and Principal Components Analysis (PCA) for processing and detecting apnea since the RIP signals are highly correlated. The PCA tracks the relative changes in respiration values and indicates the degree of signal asynchrony. For accurately measuring RIP phase differences, the cross-correlation function is calculated [Millard, 1999].

Aguirre et al. try to model and analyze three physiological time series: blood oxygen saturation, heart rate and respiration in order to distinguish between the dynamics corresponding to normal breathing and apnea. The estimated models are nonlinear autoregressive (AR), moving average (MA) and the regressors used to compose such models and are carefully chosen, among hundreds of candidates, by an automatic procedure. They find out that the underlying dynamics of the data are nonlinear and deterministic and can quantify the stability of the fixed point in phase space using the blood oxygen of time series as an index to detect apnea [Aguirre et al., 1999].

Varady et al. make use of signal processing to realize a data fusion based on empirical rules. This approach is applied on all possible features extracted from nasal and abdominal respiration signals, and an additional verification signal obtained from blood oxygen saturation to detect apnea. The output of this algorithm would be the triggering an alarm [Várady et al., 2000]. These researchers in another work, [Várady et al., 2000], use four different artificial neural networks in order to recognize three different patterns in the respiration signals. This method classifies normal breathing, hypopnea, and Heart rate and blood pressure signals in the apnea detection system since they show a well-defined performance during sleep apnea: a slight decrease after the onset of apnea and rapid decrease before its offset. However, the meaningful changes in cardiovascular signals are superimposed on other cardiovascular phenomena like fluctuations caused by the vascular baroreflex. Another problem is that the effect of apnea on cardiovascular changes occurs many seconds later and their exact characteristic is strongly dependent on the actual hemodynamic state of the patient. Hence, they only employ respiration signals and can achieve to a powerful algorithm to detect apnea by detection performance of 90%.

Much of the previous works are on extracting respiration information from the ECG follows an approach of using direct measurements of respiratory-induced characteristics of the time series ECG signal. Here, ECG is used to derive respiratory waveforms. In an definitive paper, Moody et al. [Moody et al., 1985] outline a method for deriving the respiratory waveform using two ECG leads. After subtracting the baseline, the area of each normal QRS complex in each of the two leads is obtained. Since the window width is fixed, the area is proportional to the mean amplitude of the signal, hence to the projection of the mean cardiac electrical vector onto the lead axis. Assuming that the leads are orthogonal, the arctangent of the ratio of the areas measured in the two leads gives the angle of the mean axis with respect to one of the axes. The axis direction measurement provides one sample of the respiration signal followed by interpolating using a cubic spline. Then derived signal is compared visually to chest Pneumatic Respiration Transducer measurements. Similarly, in another work, [Zhao et al., 1994], the power spectra of these two signals are obtained and compared statistically. Travaglini et al. also use such technique in the cardiac vector to derive respiration signal [Travaglini et al., 1998]. However in this work eight ECG leads are utilized.

Another approach to obtain the effect of respiration from ECG is simply to use the height of each R peak plotted as a step function to represent the respiratory signal which is compared visually against the reference respiratory signal [Felblinger et al., 1999]. There is also a well-known method that examines the power-spectra of the ECG to establish the presence of components at a respiratory frequency. In [Pallas-Areny et al., 1989], it is described that the cardiac vector contains respiratory information in the power spectrum of the ECG. The power spectra of leads I, II and III is obtained and the respiration harmonics are found to be present in the spectra.

The effect of respiration can be also derived from the photoplethysmography (PPG). Lindberg et al. employ a band-pass filter to eliminate all frequencies of the PPG except the range in which respiration is likely to be present The filtered signal is the derived respiratory waveform. The number of breaths (peaks) recorded by the reference method are compared with the number of breaths (peaks) in the derived respiratory waveform

over a ten-minute period [Lindberg et al., 1992]. Nakajima et al. also use a heart rate tunes band-pass filter approach. In order to assess the performance of this method, the median value of the breathing rate over five sequential breaths detected from the derived respiratory waveform is compared with the same measure from a reference transthoracic impedance plethysmogram [Nakajima et al., 1996].

Since the most appropriate sensor for acquiring bio-signals from neonates is surface electrodes of ECG, it is more convenient to detect AB by only monitoring the ECG. In [McNames and Fraser, 2000], the significancy of changes in various features including the time distance between R peaks (RR signal), amplitudes and width of each QRS complex are studied. They report HR ($1/RR$) as the most affected feature. The spectrogram of S pulses and the QRS energy also provide useful information about occurrence of apnea. In another work [Penzel et al., 2002], it is suggested that in compare to HR temporal characteristics, features extracted in frequency domain are more informative. Chazal et al. suggests to derive respiration effect from ECG which is integrated with HR features. Then a linear discriminant classifier is applied to them for AB detection [de Chazal et al., 2004]. Mendez et al use the features: beat-by-beat power spectral density of HR and the area under the QRS complex, in a supervised learning K-Nearest Neighbor (KNN) classifier separating apnea events from normal ones [Mendez et al., 2007]. More recent works processed the aforementioned features using BN approaches. An example of such approach is the work of Travieso et al (2014) which classifies the cepstrum of RR signals using HMM.

As discussed before, the type of apnea which we intend to detect in this project is the one caused by prematurity in pretem infants. As can be concluded, the origin of the revealment of the apnea of prematurity and sleep apnea in adults are totally different and the mechanisms involve for their generation can be separated from each other. Hence, we need to review some more studies which are especially presented for apnea of prematurity. One of the earliest studies on automatic detection of apnea of prematurity is presented by Johansson et al. where a PPG based device is used for extraction and monitoring the respiration and ECG signals. They claimed that the effect of cardiac and respiratory activities can be restored from the PPG signal using appropriate bandpass filters. Then, the number of peaks are counted by zero crossing approach in respiration signal and by detecting the positive slopes of the R peaks (using the derivative) in cardiac signal. The peak number and the power spectral density (PSD) of the extracted signals are compared with ECG and respiration signals acquired directly with surface electrodes and impedance sensor respectively. They concluded that their proposed produces signals of equal quality to these traditional methods, and is in some cases even better [Johansson et al., 1999].

Another study is presented in [Pravisani et al., 2003] where associates parameter measurements of RR Interval (RRI) series and using principal components analysis (PCA). They found RRI variability to be the appropriate tool to investigate AB, since RRI carries relevant information about Autonomic Nervous System (ANS) status. They claim that the power spectral analysis reveals two main components: the High Frequency (HF) component, which is mainly modulated by vagal tone, and the Low Frequency (LF) one, which is found to increase in presence of increased sympathetic tone. Moreover, it has been also shown that the RRI signal involves non-linear contri-

butions. Therefore in this paper, they investigate both linear and non-linear indexes of RRI variability using ratio of LF/HF components, Approximated Entropy (ApEn) and Spectral Entropy (SE) measurements. PCA is used to detect the first two factors used to provide a meaningful description of the data. Then, Hierarchical Ascending Classification (HAC) approach which is an useful tool able to evidence the presence of classes of individuals in a given dataset is employed to stratify the data obtained by PCA. HAC creates, at each step, a partition obtained by the aggregation of the two nearest elements. Finally, they are able to detect 10/13 AB episodes. These results suggest that the RRI contains information that can be employed to predict the onset of the bradycardia event.

An algorithm fusion approach is presented by Cruz et al. for early detection of AB in premature infants [Cruz et al., 2006]. In this work, two conventional methods, fixed and relative thresholds, have been implemented and an abrupt change detection based method, has been proposed. The fixed-threshold method compares a window-averaged RRI with a suitable fixed threshold. Most neonatal monitors are based on this simple detection scheme. The relative threshold method considers the fact that the mean heart rate of a patient can change through time, and implements a relative threshold of 33% of increase on the mean RR interval in a 4 seconds window. The third method models the bradycardia event as a statistically significant change in the mean RR interval. The Cumsum test is used to define an index which change abruptly while a peak happens in RR series. A weighted combination of local detectors is considered for the fusion of the algorithms which provides a higher weight to the more reliable detectors. In this work, they have used the optimality criterion proposed by Chair and Varshney [Chair and Varshney, 1986]. Similarly, Portet et al. apply just amplitude, amplitude-duration thresholds and an algorithm for sudden changes detection in HR. Then, they integrate them in a decision tree that exhibits better performance. The third algorithm searches for a sudden change in a moving window that if the difference between the maximum and minimum amplitudes is more than a threshold and this window is classified as a downward spike for which the minimum exceeds the normal range then, a bradycardia is detected. Each bradycardia previously found is then expanded forwards and backwards until it reaches the baseline to find the boundary of an AB episodes [Portet et al., 2007].

In [Altuve et al., 2009, Hernández et al., 2012], ECG data is acquired to be used for extracting the features including the RR interval, R-wave amplitude and QRS duration for periods at rest, before, during and after apnea-bradycardia episodes. Significant changes in the value of these features is reported. By analyzing their proposed features, it was observed a statistical significant modification in the amplitude of the R-wave and in the duration of the QRS complex, associated with the onset of the apnea-bradycardia episodes. For QRS complex detection, the ECG signal is processed by a cascade of low-pass and high-pass filters, followed by a double differentiator filter, an amplitude squaring process and a moving-window integrator. The final step is based on adaptive thresholds, which are continually adjusted by a set of heuristic rules whose constant parameters are optimized by evolutionary approach. These findings show the potential benefit of a multivariate approach to early AB detection and characterization.

Another work [Belal et al., 2011], presents three processing methods based on the feature signal extracted from ECG to detect the incidents of apnea of prematurity. The

slopes of the HR, RR and SpO2 during the marked-up events were calculated using a linear fit model, which was solved using the weighted least absolute residual (WLAR) method. Their first method is based on cumulative sum of HR, RR and SpO_2 along with the sum of their Shannon entropy. In the second one, the correlation between them is given to an artificial neural network. Finally, in third method the derivation of each data sample of the three parameters (HR, RR and SpO_2) is calculated. They claim that the third approach can be implemented effectively in monitoring devices. In one of the recent works [Camargo et al., 2014], a novel technology for detecting apnea of prematurity episodes is introduced which is based on cardiac pulse frequency (PF) and arterial oxygen saturation (SpO2) simultaneously. After the detection, vibrotactile stimulation is performed automatically to interrupt such episodes. The thresholds on PF and SpO2 signals had been established to identify the apnea episode through the proposed system.

A series of researches are presented by Altuve et al. ([Altuve et al., 2011a, Altuve et al., ,Altuve et al., 2015]) for AB detection using HSMM. In this works, Altuve et al. suggests an algorithm which includes allocating two HSMMs (or HMMs) models each of which are trained by normal periods or Bradycardia episodes in RR signals. In inference issue, a moving window goes through the test observation sample by sample while each model generates a value for likelihood during each segments within the window. Then the labels are determined according to the model with better likelihood at each time instant. In this thesis, we have used such strategy for applying our proposed Markovian frameworks in AB detection and discussed about it in details in following chapter. In addition, Masoudi et al. also benefited from this strategy and employ the CHMM proposed in [Rezek et al., 2000], instead of HSMM [Masoudi et al., 2013].

An AR model based method for apnea detection in preterm infants, is also proposed by Ge et al. [Ge et al., 2013]. They use RR series extracted from raw ECG signals and model it by an AR process with both time-varying coefficients and order. In the normal periods, it is reasonable to assume that the processes are quasi-stationary and the AR coefficients are varying slowly, thus the classical Kalman filtering algorithm can be applied to estimate them. Moreover, a HMM is imposed on the evolution of AR orders which update the smoothed distribution of the order at each time instant. Finally the distances of model orders in terms of their distribution functions are quantified by the Kullback-Leibler and the Kolmogorov-Smirnov distances. The applied distances act as two indices to detect AB. In addition, Ge et al., in their recent work [Ge et al., 2014], suggest another AR process based method where similarly the model order and coefficients are considered unknown. In this work, the joint posterior distribution of the AR coefficients and orders are updated on-line and recursively while the AR coefficients are assumed to be normally distributed. They propose the recursive updates of the posterior distribution of orders using the rank-1 Cholesky factor. The AR process noise level is used as an index to characterize the rapid shift in the AR model space. Their proposed method presents outstanding performance in AB detection.

2.5 Conclusion

In this chapter, the concept of prematurity and its complications are noted. AB is introduced as a common concern about premature newborns. Prematurity is with higher frequency as the gestational age is lower and indicating to an immaturity of the cardio-respiratory system. The preemies are kept in NICU as the oxygenation and tissue perfusion are monitored. They likely need long-term follow up, extending the period of hospitalization and sometimes home telemonitoring. Prevention of apnea and their consequences, require a rigorous and continuous cardiorespiratory monitoring which allows early intervention. Early detection of these episodes would reduce the associated risks, promote the development of premature newborns, improve their quality of life. Indeed, first medical treatment consists of soft and manual stimulation of the baby. Although this stimulation of the newborn at the time of AB remains the most common effective method. Experience shows that for babies in distress, the time of intervention remains as an important factor. The automatic approaches of AB detection can be helpful since the time of interventions are finally reduced slightly because the time required for diagnosis of bradycardia (10 seconds on average after beginning apnea) is not taken into account. One way to reduce this time is to employ a method for early detection or prediction of the occurrence of bradycardia. These objectives are the main theme of this thesis and solutions based on BN are proposed in chapters 3 and 4.

In addition, the basics of heart and ECG are reviewed. Understanding of cardiac physiologic changes associated with prematurity in previous medical and technical researches, we grasp the ability to interpret premature ECG based on its important features (duration of segments and intervals, axis deviations and morphologies of the waves) to find the coincidence of AB episodes.

Chapter 3

Markovian Frameworks

3.1 Introduction

The previous chapter has clearly reviewed the physiological indicators of AB episodes in premature infants. It is also stated that the early detection decreases dramatically the side-effects of prematurity. Using an automatic detector accelerates the process of detection and intervention. The occurrence of AB indicates that the dynamics of cardiorespiratory system has been changed which is expected to be distinguishable by monitoring the patterns of appropriate indicators as they construct feature signals over the time. As the main the subject of this thesis, we are interested in taking into account the dynamic of events in cardiorespiratory system using mathematical models. The methodology of this chapter is based on methods suggested by BN models, and in particular stochastic models such as HMM and its generalizations.

In following parts, firstly, HMM is described as a basic framework for other Markovian approaches. Secondly, some of the HMM generalizations are reviewed briefly, including HSMM, CHMM and CHSMM [Yu and Kobayashi, 2006] [Rezek et al., 2000] [Natarajan and Nevatia, 2007b]. The parameter definition and the approach for re-estimation are described for each method. Finally, our proposed frameworks for CHMM and CHSMM are discussed in detail.

3.2 HMM and the existing generalizations

3.2.1 HMM

In the world around us, there exists phenomena like speech or activity of the heart which can be described in the form of a sequence of static patterns which evolve according to finite certain statuses. HMM can be introduced as a tool to describe such time series.

3.2.1.1 Markov chain

Consider a system that can change its condition to one of M different states S_m, $m = 1, 2, \cdots, M$ in every moment. In other words, at each time instant, the system can enter to another state or remain in current state based on their probabilities. Define

q_t as the state of the system at time t, $t = 1, 2, \cdots, T$. Generally, in order to have a complete description of the system, it is needed to specify q_t and all the previous states. It is assumed that transition to a state depends only on previous state of the system. This is the first order Markov characteristic. Hence, the probabilistic dependency in state transitions can be simplified as follows:

$$\begin{aligned} &P(q_t|q_{t-1}, q_{t-2}, \cdots, q_1) = P(q_t|q_{t-1}) \\ &a_{nm} = P(q_t = S_m|q_{t-1} = S_n) \qquad n, m = 1, 2, \cdots, M \end{aligned} \tag{3.1}$$

where a_{nm} is the state transition probability. The basic assumption is that the output of this process is the state at each time instant. Hence the state are observable and this model, is called observable Markov model. Accordingly, it fails to model the system with the random stochastic observations.

3.2.1.2 Hidden Markov chain definition

HMM is a tool to assign a probability distribution on a sequence of observations. Due to the aforementioned limitations of the Markov chain, we can generalize this model in a way that the observations are the probabilistic function of states. This generalization makes the model stochastic in two directions: 1- in the transition state to the next one and 2- after the determination of current state, observations are generated with the corresponding probability distribution. Hence, the state sequence is hidden and can not be observed while it follows the properties of Markov processes. Fig. 3.1(a) illustrates a HMM sequence. The hidden states of HMM are considered to be discrete and the observations at each time instant depends only on the state at that time. o_t is defined as the observation at time t and is independent of the previous states and all observations in the past. Thus, the joint probability of corresponding BN is expressed according to eq. 3.2:

$$P(o_{1:T}, q_{1:T}) = P(q_1)P(o_1|q_1)\prod_{t=2}^{T} P(q_t|q_{t-1})P(o_t|q_t) \tag{3.2}$$

The state transition probability distribution is determined by matrix $A_{M\times M}$. It is also assume that the values of observation is chosen from a time invariant set of N symbols defined by matrix $B_{N\times N}$ which is called emission matrix. A HMM is described with respect to its parameters: $\lambda = \{A, M, B, N, \pi\}$, where $\pi_m = P(q_1 = S_m)$.

The reason that the observation probability is represented by a matrix is that its nature is a discrete random variable which is chosen from countable set of symbols. This means that this model can represent only the limited systems whose observations values are from a finite subset of the integers or concepts (eg, weather). However, if the observation values are infinite or uncountable finite, the dimensions of the matrix B will be unlimited. Hence, the alternative would be assigning continuous random variable to observation. In this case, the emission probability of observations is defined by a probability distribution function (pdf) according to eq. 3.3. A Gaussian kernel is usually considered as the probability of density function with mean vector and covariance matrix

(μ_m and σ_m respectively). We consider just one kernel for simplicity, however it is common to use more. For more information see [Rabiner, 1989b].

$$b_m(o_t) = N(o_t; \mu_m, \sigma_m) \tag{3.3}$$

3.2.1.3 Three main issues

In the context of HMM three main issues are usually stated which are studied in the following part.

Problem 1- Probability evaluation. The subject of this issue is to answer to a question that, given a model with parameter λ, how much the sequence of observations $o_{1:T}$, would be likely generated by the mentioned model. A straightforward solution is to evaluate all possible state transitions that can generate the sequence. Since one state can move to M other states at each time instant, there would be N^T possibilities for state sequence. Another approach called forward algorithm is commonly used by defining the forward parameter as:

$$\alpha_t(m) = P(q_t = S_m, o_{1:t}) \tag{3.4}$$

which can be calculated recursively by:

$$\alpha_t(m) = \sum_{n=1}^{M} \alpha_{t-1}(m) a_{nm} b_m(o_t) \tag{3.5}$$

Then, $P(o_{1:T})$ is equal to $\sum_{m=1}^{M} \alpha_T(m)$.

Problem 2- The most likely sequence of states. In this case, given a sequence of observations $o_{1:T}$ and the parameters of the HMM, we search for a state sequence that is the most probable one which have generated the observations. The importance of this issue is related to HMM application. One of the HMM applications is classification problem where the solution of this issue is the way to classify the test data while the parameters of the model is estimated from training data set. Using the solution of this issue reveals the most probable class/state for each sample of observation. In fact, we seek to understand the inference of the observations. A simple solution is to choose the most probable state at each time instant. In first vision, this would have the accuracy and optimality. However, a scenario can be taken into account that a state is not likely at the moment but given the the state of the next time instant, they are the most likely sequence which are then able to generate two sample of observation. Therefore, the probability of two or three observations could be studied together. One of the most popular methods is Viterbi which presents the best state for observation generation at each time instant. In this method, δ_t, the joint probability of a sequence of previous states till time $t-1$ and the observed state at time t, is calculated by Eq. 3.6:

$$\delta_t(m) = \underset{(q_1, q_2, \cdots, q_{t-1})}{\arg\max} \{P(q_1, q_2, \cdots q_{t-1}, q_t = S_m, o_{1:t}|\lambda)\} \tag{3.6}$$

So δ_t shows the sequence with maximum probability that starts from the initial state and ends with state m at time t, leading to the best partial route. This criterion can be defined as a recursion for time $t+1$:

$$\delta_{t+1}(n) = \big(max_m\{\delta_t(m)a_{mn}\}\big)b_m(o_{t+1}) \tag{3.7}$$

In order to find the sequence of states with a given criterion, it is necessary that the maximum partial probability for each time t and state m should be studied. To facilitate the search, an array, $\psi_t(m)$ is used to store the chosen state at each time instant. The Viterbi method is expressed as a step by step algorithm as follows. For each state m, the recursive algorithm is initialized as:

$$\delta_1(m) = \pi_m b_m(o_1)\psi_1(m) = 0 \tag{3.8}$$

In the next steps of the algorithm, the probability of the most likely sequence at time t is stored in ψ_t. Using relations of Eq. 3.7 for time t and Eq. 3.9 in each iteration, we have:

$$\psi_t(m) = \arg\max_m\{\delta_{t-1}(m)a_{mn}\} \tag{3.9}$$

Now the best sequence that produces the observations, after the Tth step, is chosen by Eq. 3.10:

$$P_t^* = \max_{1\leq m\leq M}\{\delta_T(m)\} \tag{3.10}$$

$$q_t^* = \arg\max_{1\leq m\leq M}\{\delta_T(m)\} \tag{3.11}$$

Using these values, it is possible to move backward and determine the rest of the nominated states at previous time instants:

$$q_t^* = \psi_{t+1}q_{t+1}^*, \qquad t = T-1, T-2, \cdots, 1 \tag{3.12}$$

Problem 3- Estimating the model parameters. The third issue arises about the model parameters. We are looking for a set of HMM parameters, λ, which maximizes $P(o|\lambda)$ given a series of observation. Generally, it seems to be complicated, and perhaps it is impossible to generally achieved the maximization of $P(o|\lambda)$, however, according to the famous Baum-Wlech method [Baum et al., 1970], we can achieve the local optimality. In this method, the backward parameters, $\beta_t(m)$, is introduced as:

$$\beta_t(m) = P(o_{t+1}, o_{t+2}, \cdots, o_T|q_t = m) \tag{3.13}$$

which can be calculated like forward parameter by a recursion, given as:

$$\beta_t(m) = \sum_{n=1}^{T} a_{mn}b_n(o_{t+1})\beta_{t+1}(n) \tag{3.14}$$

by considering the initial value: $\beta_T(m) = 1$. Defining the forward and backward parameters, a recursive algorithm can be used for estimation of HMM parameters. In this algorithm which is called the forward-backward (FB), at each step, a new values of λ^* replaces λ. Two other parameters, $\zeta_t(n,m)$ and $\gamma_t(m)$ which are defined as the number of transitions from state n to state m and the number of transitions from state m respectively, are calculated in terms of FB parameters:

$$\zeta_t(n,m) = P(q_{t-1} = S_n, q_t = S_m|o_{1:T}) \tag{3.15}$$

$$\gamma_t(m) = \sum_{n=1}^{M}\zeta_t(n,m) \tag{3.16}$$

Using Expectation-Maximization (EM) algorithm, the parameters, $\zeta_t(n,m)$ and $\gamma_t(m)$ are calculated in step E and then a_{nm}, π_m and $b_m(o_t)$ are estimated using Eq. 3.17 in step M:

$$a_{nm} = \frac{\sum_{t=1}^{T-1} \zeta_t(n,m)}{\sum_{t=1}^{T-1} \gamma_t(n)} \tag{3.17}$$

$$\pi_m = \gamma_1(m) \tag{3.18}$$

$$\hat{\mu}_m = \frac{\sum_{t=2}^{T} \gamma_t(m) o_t}{\sum_{t=2}^{T} \gamma_t(m)} \tag{3.19}$$

$$\hat{\sigma}_m = \frac{\sum_{t=2}^{T} \gamma_t(m)(o_t - \hat{\mu})^2}{\sum_{t=2}^{T} \gamma_t(m)} \tag{3.20}$$

In the context of classification, this algorithm is similar to the process of learning or training. So the third and then second issues, demonstrate a learning and inference approach for classification respectively.

3.2.2 HSMM proposed in [Yu and Kobayashi, 2006]

One of the famous generalizations of HMM is the hidden semi-Markov model (HSMM), in which the pdf of the amount of time that the system rests in a state (sojourn time) is not coerced into geometric distributions. (cf Fig. 3.1(b)). There are many works on HSMM like [Ferguson, 1980], [Levinson, 1986], [Yu and Kobayashi, 2006] and [Yu, 2010], and we review the framework proposed in [Yu and Kobayashi, 2006] due to its simplicity in implementation. The HSMM is called explicit-duration if the observations are supposed to be independent of sojourn time as considered in this work.

Consider $\{S_1, S_2, \cdots, S_M\}$ to be the states of HSMM with initial and transition probability π_m and $a_{nm} = p(q_{t-1} = S_n | q_t = S_m)$ where $q_t, t = 1, \cdots, T$, is defined as the state at time t. Also it is assumed that the resting time in a state is a random variable, generating value d, with $p_m(d)$ probability where $d \in \{1, 2, \cdots, D_{max}\}$. τ_t and o_t are the remaining time of current state to transit to another state and the observation at time t respectively. $\tau_{t_0} = d$ means that the system leaves the current state at $t = t_0 + d$.

3.2.2.1 FB recursions

The forward parameter can be defined as follows:

$$\alpha_{t|t-1}(m,d) = P(q_t = S_m, \tau_t = d | o_{1:t-1}) \tag{3.21}$$

Let $q_t = S_m, \tau_t = d$, so two conditions might happen. The state sequence may transit to S_m from any other states or continues the previous state, i.e. $q_{t-1} = S_m, \tau_{t-1} = d+1$.

Therefore, the following equation is straightforward:

$$\begin{aligned}\alpha_{t|t-1}(m,d) &= P(q_{t-1}=S_m,\tau_{t-1}=d+1|o_{1:t-1}) + P(q_t=S_m,\tau_t=d,q_{t-1}=S_n,\tau_{t-1}=1)\\ &= \alpha_{t|t-1}(m,d+1) + \sum_{n=1}^{M}\alpha_{t-1|t-1}(n,1)a_{nm}p_m(d)\\ &= \alpha_{t|t-1}(m,d+1)\tilde{b}_m(o_{t-1}) + \sum_{n=1}^{M}\alpha_{t|t-1}(n,1)\tilde{b}_n(o_{t-1})a_{nm}p_m(d)\end{aligned} \tag{3.22}$$

where

$$\tilde{b}_m(o_t) = \frac{\alpha_{t|t}(m,d)}{\alpha_{t|t-1}(m,d)} \tag{3.23}$$

and the initial condition of forward variable is given by:

$$\alpha_{1|0}(m,d) = \pi_m p_m(d) \tag{3.24}$$

Similarly, for backward parameter which is defined as:

$$\beta_t(m,d) = \frac{P(o_{t:T}|q_t=S_m,\tau_t=d)}{P(o_{t:T}|o_{1:t-1})}, \tag{3.25}$$

by assuming $q_t = S_m, \tau_t = d$, two conditions might be happen. While $d = 1$, at next sample of time, there would be a state transition, $q_{t+1} = S_n, \tau_{t+1} = d'$. Whereas for $d > 1$, the state sequence remains in current state ($q_{t+1} = S_m, \tau_{t+1} = d - 1$). Consequently, backward variable is defined as follows:

$$\beta_t(m,d) = \begin{cases} \tilde{b}_m(o_t)\beta_{t+1}(m,d-1) & d \neq 1 \\ \tilde{b}_m(o_t)\sum_{n=1}^{M}\sum_{d'=1}^{D_{max}}\beta_{t+1}(n,d')p_n(d') & d = 1 \end{cases}$$

with initial condition defined as:

$$\beta_{1|0}(m,d) = \tilde{b}_m(o_T) \tag{3.26}$$

3.2.2.2 Three main issues

Now we are ready to present the solution of three important and famous problems usually defined in HMM context.

Problem 1- Probability of specific sequence of observations subjected to the model parameters:

$$P(o_{1:T}) = P(o_1)\prod_{t=2}^{T}P(o_t|o_{1:t-1}) \tag{3.27}$$

where

$$P(o_t|o_{1:t-1}) = \sum_{m,d}\alpha_{t|t-1}(m,d)\tilde{b}_m(o_t) \tag{3.28}$$

Problem 2- Choosing the most probable sequence of states that generates the sequence of known observations (Inference): The solution is based on the criterion presented in the paper [Yu and Kobayashi, 2006] for maximum a posteriori (MAP) re-estimation of model parameters. Unlike the Viterbi algorithm, only one forward pass is

necessary based on the previously calculated F-B variables. First, Eq. 3.29 is calculated for each samples of time. Then starting from $t = 1$, assigning q_t^* to the states between time t to $t + \tau_t^*$, $q^*_{t+\tau_t^*+1}$ to the states of $t + \tau_t^* + 1$ to $t + \tau_t^* + \tau^*_{t+\tau_t^*+1}$ and so forth.

$$(q_t^*, \tau_t^*) = \arg\max_{m,d} P(S_m \ starts \ at \ t, \tau_t = d | o_{1:T}) \tag{3.29}$$
$$\arg\max_{m,d} \beta_t(m,d) p_m(d) \sum_n \alpha_{t-1|t-2}(n,1) b_N^*(o_{t-1}) a_{nm}$$

Problem 3- Finding the parameters of the models from the observation. The following Maximum Likelihood (ML) procedure should be repeated till the convergence of the likelihood :

$$\hat{a}_{nm} = \sum_{t=2}^{T} \frac{\alpha_{t-1|t-2}(n,1)\tilde{b}_n(o_{t-1})a_{nm}\sum_d p_m(d)\beta_t(m,d)}{\sum_{m'}\alpha_{t-1|t-2}(n,1)\tilde{b}_{m'}(o_{t-1})a_{nm'}\sum_d p_{m'}(d)\beta_t(m',d)} \tag{3.30}$$

$$\hat{p}_m(d) = \sum_{t=2}^{T} \frac{\sum_n \alpha_{t-1|t-2}(n,1)\tilde{b}_n(o_{t-1})a_{nm}p_m(d)\beta_t(m,d)}{\sum_{d'}\sum_n \alpha_{t-1|t-2}(n,1)\tilde{b}_n(o_{t-1})a_{nm}p_m(d')\beta_t(m,d')} \tag{3.31}$$

$$\hat{\pi}_m = \frac{\pi_m \sum_d p_m(d)\beta_1(m,d)}{\sum_{m'}\pi_{m'}\sum_d p_{m'}(d)\beta_1(m',d)} \tag{3.32}$$

$$\hat{\mu}_m = \frac{\sum_{t=2}^{T}\sum_d \alpha_{t|t-1}(m,d)\beta_t(m,d)o_t}{\sum_{t=2}^{T}\sum_{m'}\sum_d \alpha_{t|t-1}(m',d)\beta_t(m',d)} \tag{3.33}$$

$$\hat{\sigma}_m = \frac{\sum_{t=2}^{T}\sum_d \alpha_{t|t-1}(m,d)\beta_t(m,d)(o_t-\hat{\mu})^2}{\sum_{t=2}^{T}\sum_{m'}\sum_d \alpha_{t|t-1}(m',d)\beta_t(m',d)} \tag{3.34}$$

3.2.3 Coupling in HMM

There are variety of new HMM architectures that have been proposed to solve a specific class of problems and to overcome certain limitations in the traditional HMM. The characteristics coming along with each of the generalization of standard HMM are effective in special aspects. More over, combining the features of different generalizations to form a more powerful model seems to be a need to model some of the real systems. Although HMMs are, by definition, single-process models, different approaches have been proposed to use them for the analysis of multivariate data. One common approach is to consider the observation data as a multi-dimensional random variable (multivariate). Although these approaches provide interesting results in some applications ([Dumont et al., 2008], [Altuve et al., 2011a] and [Altuve et al., 2011b]), the single-process nature of HMM can not be adapted easily in other cases, where observations are indeed generated by distinct underlying processes, such as in vision, speech recognition [Brand et al., 1997b] or forensic analysis [Brewer et al., 2006].

In order to overcome these limitations, another BN, called CHMM is presented (Fig. 3.1(e)). There have been some variations of the fully coupled HMMs as an interesting new HMM architectures for which the model size and inference problem are more

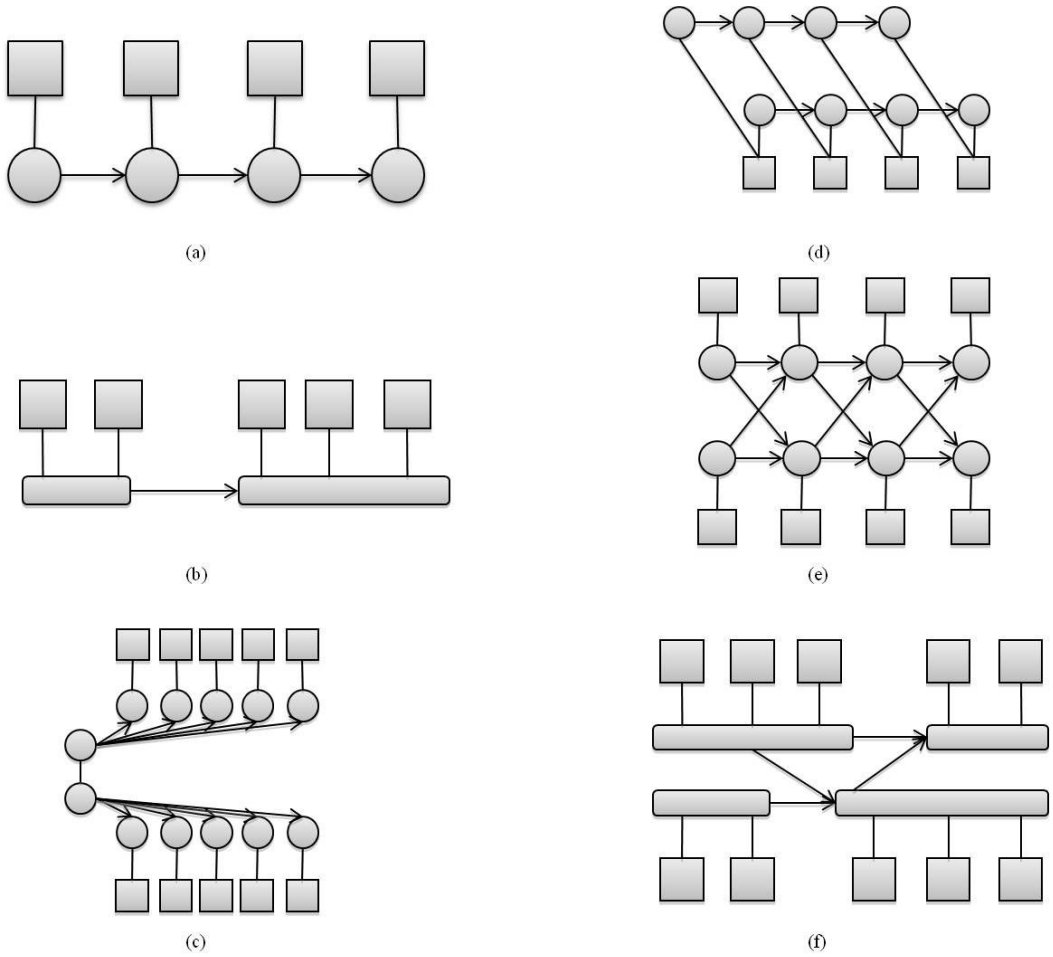

Figure 3.1: Structures of a) HMM, b) HSMM, c) ECHMM, d) FHMM, e) CHMM and f) CHSMM. Squares and circles/rectangular represent observations and states respectively. Rectangular shows a resting in a state.

flexible but more complicated than the standard full coupled HMM models. Among these variations, we can point out to Event Coupled HMM (ECHMM), Multirate-coupled CHMMs (MCHMM), Coupled Factorial HMM (CFHMM) (Fig. 3.1(d), Coupled observation Decomposed HMM (CDHMM) [Guo et al., 2012], Error Weighted semi-coupled HMM (EWSCHMM) [Lin et al., 2012] and Coupled Hidden Semi Markov Models (CHSMM). The former is depicted in Fig. 3.1(c), proposed in [Kristjansson et al., 2000]. The motivation of this variation is to model the processes with loosely coupling where the coupled parameter is the onset of the events. However, only limited range of processes can be modeled by ECHMM since its specific structure makes restrictions. MCHMMs are a multiscale generalization of HMMs for the joint modeling of scale-based observation sequences. Similar to CHMMs, hidden states are assumed to underlie the observations in a multi channel structure, but unlike standard HMMs, each channel is associated with a particular scale. This will force the sequences of observations to have different lengths naturally [Cetin et al., 2007]. The CFHMM is a factorial hidden Markov model (FHMM) with the coupling between the components shown as the dependencies between their corresponding independently evolving HMMs [Kodali

et al., 2010].

3.2.3.1 Previous studies on Coupling of HMM

CHMM has been proposed for the first time by Brand in [Brand, 1997, Brand et al., 1997b] as a generalization of HMM. In a CHMM, each channel (Markov chain) is associated with an univariate observation to represent its underlying generation process and transition probabilities depend on the current state of all channels. However, this structure implies that the state space grows exponentially with respect to the number of channels [Brand, 1997]. In order to cope with this complexity, Brand proposed a simplification, considering a factorization of the transition matrix:

$$P(S_t^c|S_{t-1}^1, S_{t-1}^2, \cdots, S_{t-1}^C) = \prod_{c=1}^{C} P(S_t^c|S_{t-1}^c) \tag{3.35}$$

where S_t^c denotes the state of channel c at time t and C is total number of channels. According to Brand's assumption, the state conditional probability in the left side of equation (3.35) is substituted by the product of all marginal conditional probabilities. This model has been successfully applied in the field of sensor fusion, such as in Forensic Electronics [Brewer et al., 2006], genetics [Zhao et al., 2011], audio-visual speech recognition systems [Nefian et al., 2002] and target tracking [Gai et al., 2007].

In another work, Rezek et al. computed the FB parameters without considering Brand's assumption in Eq. (3.35) and derived the ML estimators for the CHMM parameters using EM algorithm [Rezek et al., 2000]. They considered the CHMM as an one-channel HMM with an ordered C-fold state formed by $(S^1, S^2, \cdots, S^C)$, with a state space of $N = \prod_{c=1}^{C} n_c$ and a transition matrix of dimension $N \times N$. The algorithm's complexity reaches $O(TM^{2C})$ with $M = n_c$. Based on the idea of Rezek et al., an $N \times N$ transition matrix A is formed, whose indices represent probability of transition from one C-fold state to another state of the respective structure in large HMM-like. Practically there is no necessity to estimate the full $N \times N$ state transition matrix A. Each of its element can still be obtained as a factor of the elements of the original transition matrices A^c. In this framework, observation at each time instant can be chosen from a set of finite and countable set of symbols or may be generated from a pdf. Consequently solving the three aforesaid problems for a CHMM is equivalent to solving the same problems in a single-channel HMM context.

Rezek applied CHMM using Gibbs sampling and ML criterion in learning procedure of two different real data. First database, named Cheyne Stokes data, consisted of one EEG recording and a simultaneous respiration recording. A fractional spectral radius (FSR) measure was computed from consecutive non-overlapping windows. Respiration FSR and EEG FSR formed the first and second channel respectively. ML estimator in comparison with sampling method needs more restart to find the most dominant solution and also it mostly converges to the singularities. The second database is considered to model a full night EEG recording were fitted to a fifth order Bayesian auto regressive model whose coefficient were extracted priory from two EEG channels. Each of which is considered as an observation of each CHMM channel. The number of states in one channel is set to 3 corresponding to three sleep stages. The results

fully agreed with literature and show that the coupling is minimal during periods of wakefulness and maximal during deep sleep. ML approach is faster while runs in to singularities and demands several restarts and if the number of states grows, ML tries to visit the states of relatively low data evidence.

Zhong et al. proposed another approach in [Zhong and Ghosh, 2001,Fu et al., 2003], in which the transition probability is expressed under the form of a weighted sum of marginal conditional probabilities, associated with a rather complex estimation method for the model parameters considering the normalization of weights as follows:

$$P(S_t^{\varsigma}|S_{t-1}^1, S_{t-1}^2, \cdots, S_{t-1}^C) = \sum_{c=1}^{C} \theta_{c\varsigma} P(S_t^{\varsigma}|S_{t-1}^c) \tag{3.36}$$

where $\theta_{c\varsigma}$ is a coefficient, named coupling weight, which shows the amount of effect, S_{t-1}^c from channel c induced to the distribution of S_{t-1}^{ς} in channel c. Using the assumption, they could model the joint dependency as a linear combination of all marginal dependencies. Since the C channels are coupled together, the extended FB variables should be defined jointly across C channels as follows:

$$\alpha_t(m_1, m_2, \cdots, m_C) = P(O_1 : t, q_t^1 = S_{m_1}, \cdots, q_t^C = S_{m_C}) \tag{3.37}$$

$$= \sigma_{n_1,n_2,\cdots,n_C} \alpha_{t-1}(n_1, n_2, \cdots, n_C) \prod_c b_{m_c}^c(o_t^c) \sum_{c'} \theta_{c'c} a_{n_{c'}m_c}^{c'c} \tag{3.38}$$

$$\beta_t(m_1, m_2, \cdots, m_C) = P(O_1 : t|q_t^1 = S_{m_1}, \cdots, q_t^C = S_{m_C}) \tag{3.39}$$

$$= \sigma_{n_1,n_2,\cdots,n_C} \beta_{t+1}(n_1, n_2, \cdots, n_C) \prod_c b_{m_c}^c(o_t^c) \sum_{c'} \theta_{cc'} a_{m_c n_c'}^{cc'} \tag{3.40}$$

where $b_{m_c}^c(o_t^c)$ and $a_{n_{c'}m_c}^{c'c}$ are the observation probability subjected to the state of corresponding channel and transition matrix considering the impact of between or within channels. According to this formulation, both the joint FB variables cannot be easily decoupled into product of marginals the same as what we can see in Rezek's work. Zhong's algorithm failed to use Baum-Welch method for re-estimation part. This is the exact consequence of their inability to calculate the separate FB parameters so that, both the E and M steps become too complex to compute since they are entangled together. Here, lagrange multiplier optimization method is used to address the problem since all parameters are still restricted to stochastic constraints. They claimed that it would be straightforward to prove that and applying the transformation, the re-estimation of parameters will converge to a critical point of the likelihood function $P(O|\lambda)$. The only remaining complication is that the calculation of model parameters using likelihood can be stated as the analytical forms.

Zhou et al. apply the coupled HMM in the field of genetic, defining the nodes as the states which are affected from previous nodes called parents. They consider the conditional probability for a node with multiple parents as if node Y has m parents which is computed as follows [Zhou and Wong, 2007]:

$$P(S_t^{\varsigma}|S_{t-1}^1, S_{t-1}^2, \cdots, S_{t-1}^C) = \sum_{c=1}^{C} \frac{1}{C} P(S_t^{\varsigma}|S_{t-1}^c) \tag{3.41}$$

This equation shows that the probability of jumping to a node with multiple parents can be equal to the average of marginal probability.

In another study, Nefian et al. propose an idea on using the two-channel coupled HMM to model an audio-visual speech recognition system [Nefian et al., 2002, Xie and Liu, 2006]. Avoiding the problems of ML estimation, they present an efficient method for the initialization of the ML training which employs a Viterbi algorithm derived specially for the coupled HMM. The Viterbi algorithm determines the optimal sequence of states for the coupled nodes of the audio and video streams and maximizes the observation likelihood. CHMM can be tracked in many other applications like in Forensic Electronic [Brewer et al., 2006] and target tracking [Gai et al., 2007].

3.2.3.2 A study on CHSMM

CHSMM can be considered as another variation of CHMM as a combination of the HSMM and CHMM. It is first proposed by Natarajan et al. [Natarajan and Nevatia, 2007b] to include a more complex structure in both space (channel) and time (Fig. 3.1(f)). Although Brand's original formulation as the conditional independence assumption in Eq. 3.35 is not really a marginal conditional probability since the terms does not sum up to one, it is a necessary simplification to yield tractable calculations in CHSMM. Natarajan adopted this formulation to avoid excessive parameters appeared in [Zhong and Ghosh, 2001]. The normalization issue is handled by scaling and normalizing the FB variables.

CHSMM is a CHMM for which the sojourn durations of each state have an explicitly specified probability density function. Its implementation in [Natarajan and Nevatia, 2007b] suffers from numerical underflow commonly occurred in HMM and its generalizations, and is also a well known situation in Bayesian analysis when probabilities (smaller than 1 by definition) are multiplied for the forward and backward calculations. We propose a simple trick inspired from the work of [Yu and Kobayashi, 2006] to rescale FB probabilities in order to avoid the underflow problem. As mentioned previously, the idea is to replace the joint probabilities with conditional probabilities in both FB parameters definitions. We then derived the algorithmic procedures to re-estimate the model parameters using the ML criterion.

In following of this part, we briefly describe the CHSMM framework as stated in [Natarajan and Nevatia, 2007a]. In this study, the parameters of the model is defined as before just adding the duration variable as $\lambda = (Q^c, O^c, A^c, B^c, D^c, \pi^c)$. In this relation, the variables, A^c, B^c, D^c, π^c are possible states set, set of observation, observation probability distribution and initial state distribution respectively. D^c contains a set of parameters of the form $P(d_m^c = k|q_t^c = m)$, i.e. the probability of resting in state m in channel c for k samples of time. A^c is the transition matrix in channel c which is calculated based on Brand's assumption. As in the standard HMM, Baum-Welch algorithm used for parameter re-estimation. So calculation of FB parameters is sufficient to establish the framework. For this goal, a helpful parameter defined as they called it, $brick(c, m, tsc, tec)$, which denotes the probability of channel c staying in state m for a duration of $tec - tsc + 1$ and then observing the sequence $(O_{tsc}, \cdots, O_{tec})$. Then we can

have:

$$brick(c,m,tsc,tec) = P(d_m^c = tec - tsc + 1) \prod_{t=tsc}^{tec} P(O_t|q_t^c = m) \tag{3.42}$$

where the first term is the probability of channel c in state m having a duration $tec - tsc + 1$ and the second one denotes the probability of observing O_t in channel c and state m. Then they defined $\alpha_{m,ts,te}^c$ as the probability of channel c remaining in state i from time ts to te, given the observation O_1 to O_{te} on channel c. So the forward parameter can be calculated recursively using the following recursion:

$$\begin{aligned}\alpha_{m,ts,te}^c = brick(c,m,ts,te) \sum_{mc=1}^{M} \sum_{ts'=ts-1-Th}^{ts-1} P(q_{ts}^c = m|q_{ts-1}^c = mc)\alpha_{mc,ts',ts-1}^c \\ \times \prod_{c'=1,c'\neq c}^{C} \{ \sum_{mc'=1}^{M} \sum_{tsc'=ts-Th-1}^{ts-1} \sum_{tec'=ts-1}^{tsc'+Th} P(q_{ts}^c = m|q_{ts-1}^{c'} = mc')\alpha_{mc',tsc',tec'}^{c'} \}\end{aligned} \tag{3.43}$$

This recursion is true under the condition of the channels evolve independently and influence each other only in state transition. The second and third term of the equation correspond to the intra-channel and cross-channel effects, respectively. The backward parameter $\beta_{m,ts,te}^c$ can be calculated according to the following recursion:

$$\begin{aligned}\beta_{m,ts,te}^c = brick(c,m,ts,te) \sum_{mc=1}^{M} \sum_{te'=te-1}^{ts+1+Th} P^c(mc|m)\beta_{mc,te+1,te'}^c \\ \times \prod_{c'=1,c'\neq c}^{C} \{ \sum_{tsc'=te-Th}^{te} \sum_{tec'=te}^{tsc'+Th} P^c(m|mc')\alpha_{mc',tsc',tec'}^{c'} \}\end{aligned} \tag{3.44}$$

In this equation, they have used the forward parameter to weight the cross-channel effect since this parameter controls the state transition. By the means of FB parameters, the re-estimation parameter can be readily calculated as follows:

- Initial Probability

$$\hat{\pi}_m^c = \pi_m^c \sum_{te=1}^{te=Th} \beta_{m,1,te}^c \tag{3.45}$$

- Transition Probability

$$\begin{aligned}\hat{P}(q_{t+1}^c = m|q_{t+1}^{c'} = n) = \sum_{t=1}^{T} \{ \sum_{ts'=t-Th}^{t} \sum_{te'=t}^{c'} \alpha_{m,ts',te'}^{c'} \\ \times P(q_{t+1}^c = m|q_{t+1}^{c'} = n) \sum_{te=t+1}^{t+1+Th} \beta_{m,t+1,te}^c \}\end{aligned} \tag{3.46}$$

- Output Probability

$$\hat{P}(o_k^c|q_t^c = m) = \sum_{t=1}^{T} \sum_{te=t,o_{t'}^c=o_t^c,t\leq t'\leq te} \{\alpha_{m,t,te}^c \sum_{m=1}^{M} P(q_{te+1}^c = m|q_{te}^c = n) \sum_{te'=te+1}^{te+1+Th} \beta_{m,te+1,te'}^c \} \tag{3.47}$$

- Duration Probability

$$\hat{P}(d_t^c|q_t^c = m)=\sum_{ts=1}^{T}\sum_{ts=1}^{T}\{\alpha_{m,ts,ts+d}^c \sum_{m=1}^{M} P(q_{ts+d+1}^c = m|q_{ts+d}^c = n)\sum_{te=ts+d+1}^{ts+d+1+Th} \beta_{m,ts+d+1,te}^c\} \quad (3.48)$$

The most interesting part of this work is the decoding algorithm. To make an inference in a graphical model, it is needed to find the MAP path, given the observations. Let us denote highest probability path at that time instant with variable $\delta_{m,ts,te}^c$, which is defined as the log-likelihood of the maximum probability path such that channel c is in state m from time ts to te.

$$\begin{aligned}\delta_{n,ts,te}^c = \log(brick(c,m,ts,te)) + max_{n,ts'}(\delta_{n,ts',ts-1}^c + \log(P^c(n|m))) \\ + \sum_{c'=1}^{C} max_{m,ts'}(\delta_{n',ts',te'}^{c'} + log(P^c(n|n'))) \end{aligned} \quad (3.49)$$

where, $ts' \leq ts$ and $te' \geq te$. $max_{c=1}^{C}\{max_{m=1}^{M}(\delta_{m,ts,T}^c)\}$ gives the approximate maximum log-likelihood probability by storing the (m, ts') values that produce the maximum in Eq. 3.49, the most probable path through CHSMM can be reconstructed. We can criticize this algorithm in numerical underflow commonly happened in implementation of HMM and its generalization. However, in these equations, this problem empowered a great deal more severely due to considering channels. Hence, the problem should be systemically overcome in any algorithm which is proposed for a coupling issue. Other previous coupling methods in HMM context are based on the recursive relations of FB variables, for which the learning procedure can be complicated and time consuming, such as in [Zhong and Ghosh, 2001]. In following sections, we try to introduce a kind of equations using the approach in [Yu and Kobayashi, 2006] that is capable of solving the underflow problem and suggesting a new framework for CHMM with a considerable reduction in the complexity of the FB algorithm comparing to CHMM proposed by Rezek.

3.3 Proposed generalizations

3.3.1 CHMM

Let's note $\{S_1^c, S_2^c, \cdots, S_{M(c)}^c\}$ to be the state space of channel c in a CHMM and let q_t^c and o_t^c, $t = 1, 2, \cdots, T$ be the state and the observation of channel c at time t, respectively. C is the total number of channels in the CHMM model. Also, let $a_{nm}^{c'c} = P(q_t^c = S_m^c|q_{t-1}^{c'} = S_n^{c'})$ denote the probability of transition to state m in channel c at time t, subjected to being in state n in channel c' at time $t-1$. The probability of the observation is written as $b_m(o_t^c) = P(o_t^c|q_t^c = S_m^c)$, where o_t^c may be either discrete or continuous. We also define $o_t = \{o_t^1, o_t^2, \cdots, o_t^C\}$. In this study, a is assumed to be normally distributed, and characterized by its mean (μ_m^c) and standard deviation (σ_m^c). For simplicity, we also note $v_t^c(m) \equiv \{q_t^c = S_m^c\}$, so that $a_{nm}^{c'c} = P(v_t^c(m)|v_{t-1}^{c'}(n))$. The structure of the inter-channel coupling is depicted in Fig. 3.2 for the two-channel case.

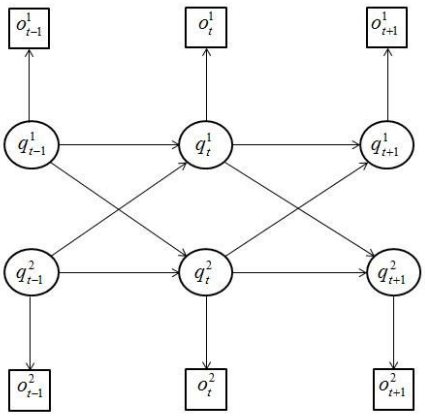

Figure 3.2: Bayesian Network representation for the proposed CHMM showing the probabilistic relations among the states and observations for the particular case of a two-channel CHMM.

3.3.1.1 FB recursions

In order to define the FB parameters, first consider the probability of observation of each channel at time t given all the previous observations.

$$P(o_t^\varsigma|o_{1:t-1}) = \sum_{m=1}^{M(\varsigma)} P(v_t^\varsigma(m), o_t^\varsigma|o_{1:t-1}) = \sum_{m=1}^{M(\varsigma)} P(o_t^\varsigma|v_t^\varsigma(m), o_{1:t-1})P(v_t^\varsigma(m)|o_{1:t-1})$$
$$= \sum_{m=1}^{M(\varsigma)} b_m(o_t^\varsigma)P(v_t^\varsigma(m)|o_{1:t-1}) \tag{3.50}$$

Following [Yu and Kobayashi, 2006], we define the forward parameter as:

$$\alpha_{t|x}^\varsigma(m) = P(v_t^\varsigma(m)|o_{1:x}) \tag{3.51}$$

where for $x = t-1, t, T$, the above quantity is termed as predicted, filtered and smoothed probability respectively. The forward recursion based on the predicted probability is determined by

$$\alpha_{t|t-1}^\varsigma(m) = P(v_t^\varsigma(m)|o_{1:t-1}) = \sum_{n_1=1}^{M(1)}\sum_{n_2=1}^{M(2)}\cdots\sum_{n_C=1}^{M(C)} P(v_t^\varsigma(m), v_{t-1}^1(n_1), v_{t-1}^2(n_2), \cdots, v_{t-1}^C(n_C)|o_{1:t-1})$$
$$= \sum_{n_1=1}^{M(1)}\sum_{n_2=1}^{M(2)}\cdots\sum_{n_C=1}^{M(C)} P(v_t^\varsigma(m)|v_{t-1}^1(n_1), v_{t-1}^2(n_2), \cdots, v_{t-1}^C(n_C), o_{1:t-1})$$
$$\times P(v_{t-1}^1(n_1), \cdots, v_{t-1}^C(n_C)|o_{1:t-1}) \tag{3.52}$$

where the first term can be simply calculated by Brand's assumption in Eq. (3.35). $o_{1:t-1}$ can be omitted since knowing all the previous states, it does not add any information

to estimate $v_t^\varsigma(m)$

$$P(v_t^\varsigma(m)|v_{t-1}^1(n_1),\cdots,v_{t-1}^C(n_C),o_{1:t-1}) = \prod_{c=1}^{C} P(v_t^\varsigma(m)|v_{t-1}^c(n_c)) \tag{3.53}$$

For the second term, we show in appendix A.3 that the states of channels given the observations $o_{1:t-1}$ are independent:

$$P(v_{t-1}^1(n_1),...,v_{t-1}^C(n_C)|o_{1:t-1}) = \prod_{c=1}^{C} P(v_{t-1}^c(n_c)|o_{1:t-1}) \tag{3.54}$$

Substituting Eq. (3.35), (3.53) and (3.54) in (3.52), the following recursion can be obtained:

$$\alpha_{t|t-1}^\varsigma(m) = \sum_{n_1=1}^{M(1)}\sum_{n_2=1}^{M(2)}\cdots\sum_{n_C=1}^{M(C)}\prod_{c=1}^{C} P(v_t^\varsigma(m)|v_{t-1}^c(n_c)) \times P(v_{t-1}^c(n_c)|o_{1:t-1}) \tag{3.55}$$

Note that $P(v_t^\varsigma(m)|v_{t-1}^c(n_c)) \neq P(v_t^\varsigma(m)|v_{t-1}^c(n_c),o_{1:t-1})$. Hence, the summations can be exchanged with the product.

$$\begin{aligned}\alpha_{t|t-1}^\varsigma(m) &= \prod_{c=1}^{C}\sum_{n_c=1}^{M(c)} P(v_t^\varsigma(m)|v_{t-1}^c(n_c))P(v_{t-1}^c(n_c)|o_{1:t-1}) = \prod_{c=1}^{C}\sum_{n_c=1}^{M(c)} a_{n_c m}^{c\varsigma}\alpha_{t-1|t-1}^c(n_c) \\ &= \prod_{c=1}^{C}\sum_{n_c=1}^{M(c)} a_{n_c m}^{c\varsigma}\alpha_{t-1|t-2}^c(n_c)\tilde{b}_{n_c}^c(o_{t-1})\end{aligned} \tag{3.56}$$

where $\tilde{b}_m^\varsigma(o_t)$ is defined as the ratio of the filtered probability, $\alpha_{t|t}^\varsigma(m)$, over the predicted probability $\alpha_{t|t-1}^\varsigma(m)$.

$$\tilde{b}_m^\varsigma(o_t) \triangleq \frac{\alpha_{t|t}^\varsigma(m)}{\alpha_{t|t-1}^\varsigma(m)} = \frac{P(o_t|v_t^\varsigma(m),o_{1:t-1})}{P(o_t|o_{1:t-1})} = \frac{P(o_t|v_t^\varsigma(m),o_{1:t-1})}{\prod_{c=1}^{C}P(o_t^c|o_{1:t-1})}. \tag{3.57}$$

The last equality is based on the observation decomposition (cf appendix A.2),

$$P(o_t|o_{1:t-1}) = \prod_{c=1}^{C} P(o_t^c|o_{1:t-1}) \tag{3.58}$$

Assuming the effects of other channels, the nominator of Eq. (3.57) is summarized as the product of distinct channel effects as follows:

$$P(o_t|v_t^\varsigma(m),o_{1:t-1}) = \underbrace{\sum_{n_1=1}^{M(1)}\cdots\sum_{n_C=1}^{M(C)}}_{C-1\text{ (except channel }\varsigma)} P(o_t,v_t^1(n_1),\cdots,v_t^C(n_C)|v_t^\varsigma(m),o_{1:t-1}) \tag{3.59}$$

$$\begin{aligned}&= \underbrace{\sum_{n_1=1}^{M(1)}\cdots\sum_{n_C=1}^{M(C)}}_{C-1\text{ (except channel }\varsigma)} P(o_t^1|v_t^1(n_1))P(o_t^2|v_t^2(n_2))\cdots P(o_t^C|v_t^C(n_C)) \\ &\quad \times P(v_t^1(n_1),\cdots,v_t^C(n_C)|v_t^\varsigma(m),o_{1:t-1})\end{aligned} \tag{3.60}$$

since the probability calculation of o_t^c only needs the corresponding state, i.e. $v_t^c(n_c)$, we can establish the following equation:

$$P(o_t|v_t^\varsigma(m), o_{1:t-1}) = b_m(o_t^\varsigma) \underbrace{\sum_{n_1=1}^{M(1)} \cdots \sum_{n_C=1}^{M(C)}}_{C-1 \text{ (except channel } \varsigma)} b_{n_1}(o_t^1) \cdots b_{n_C}(o_t^C) \times P(v_t^1(n_1), \cdots, v_t^C(n_C)|v_t^\varsigma(m), o_{1:t-1}) \quad (3.61)$$

For achieving a recursion, we utilize the following simplification (also necessary to prove Eq. (3.54)) :

$$P(v_t^1(n_1), \cdots, v_t^C(n_C)|o_{1:t-1}) = \prod_{c=1}^{C} P(v_t^c(n_c)|o_{1:t-1}) \quad (3.62)$$

Then we have:

$$P(v_t^1(n_1), \cdots, v_t^C(n_C)|v_t^\varsigma(m), o_{1:t-1}) = \frac{\prod_{c=1}^{C} P(v_t^c(n_c)|o_{1:t-1})}{P(v_t^\varsigma(m)|o_{1:t-1})} = \prod_{c=1, c\neq\varsigma}^{C} \alpha_{t|t-1}^c(n_c) \quad (3.63)$$

Hence, using Eq. (3.50), we have:

$$P(o_t|v_t^\varsigma(m), o_{1:t-1}) = b_m(o_t^\varsigma) \prod_{c=1, c\neq\varsigma}^{C} \sum_{n_c=1}^{M(c)} b_{n_c}(o_t^c) \alpha_{t|t-1}^c(n_c) = b_m(o_t^\varsigma) \prod_{c=1, c\neq\varsigma}^{C} P(o_t^c|o_{1:t-1}) \quad (3.64)$$

Accordingly, $\tilde{b}_m^\varsigma(o_t)$ can be calculated by:

$$\tilde{b}_m^\varsigma(o_t) = \frac{b_m(o_t^\varsigma)}{P(o_t^\varsigma|o_{1:t-1})} \quad (3.65)$$

Backward parameter is defined in a way that the probability of being in state m of channel c given all the samples of observations, i.e. smoothed probability, could be easily calculated by the product of forward and backward parameters.

$$\alpha_{t|T}^\varsigma(m) = \alpha_{t|t-1}^\varsigma(m) \times \beta_t^\varsigma(m) \quad (3.66)$$

which leads to :

The backward recursion is derived based on the next transition in channel ς as follows:

$$\begin{aligned} \beta_t^\varsigma(m) &= \tilde{b}_m^\varsigma(o_t) \sum_{n=1}^{M(\varsigma)} \frac{P(o_{t+1:T}, v_{t+1}^\varsigma(n)|v_t^\varsigma(m), o_{1:t})}{P(o_{t+1:T}|o_{1:t})} \\ &= \tilde{b}_m^\varsigma(o_t) \sum_{n=1}^{M(\varsigma)} \frac{P(o_{t+1:T}|v_{t+1}^\varsigma(n), v_t^\varsigma(m), o_{1:t})}{P(o_{t+1:T}|o_{1:t})} \times P(v_{t+1}^\varsigma(n)|v_t^\varsigma(m), o_{1:t}) \end{aligned} \quad (3.67)$$

Considering $P(o_{t+1:T}|v_{t+1}^{\varsigma}(n), v_t^{\varsigma}(m), o_{1:t}) = P(o_{t+1:T}|v_{t+1}^{\varsigma}(n), o_{1:t})$ and Eq. (3.67), the first term in the summation can be identified as $\beta_{t+1}^{\varsigma}(n)$. The second term writes :

$$P(v_{t+1}^{\varsigma}(n)|v_t^{\varsigma}(m), o_{1:t}) = \underbrace{\sum_{n_1=1}^{M(1)} \cdots \sum_{n_C=1}^{M(C)}}_{\text{except}\varsigma} P(v_{t+1}^{\varsigma}(n), v_t^1(n_1), \cdots, v_t^C(n_C)|v_t^{\varsigma}(m), o_{1:t})$$

$$= \underbrace{\sum_{n_1=1}^{M(1)} \cdots \sum_{n_C=1}^{M(C)}}_{\text{except } \varsigma} P(v_{t+1}^{\varsigma}(n)|v_t^1(n_1), \cdots, v_t^C(n_C), v_t^{\varsigma}(m), o_{1:t}) \times P(v_t^c(n_c), \cdots, v_t^C(n_C)|v_t^{\varsigma}(m), o_{1:t})$$

$$= a_{mn}^{\varsigma\varsigma} \prod_{c \neq \varsigma}^{C} \sum_{n_c=1}^{M(c)} a_{n_c n}^{c\varsigma} \alpha_{t|t-1}^{c}(n_c) \tilde{b}_{n_c}^{c}(o_t) \tag{3.68}$$

Thus the backward recursion is given by:

$$\beta_t^{\varsigma}(m) = \tilde{b}_m^{\varsigma}(o_t) \sum_{n=1}^{M(\varsigma)} \left\{ \beta_{t+1}^{\varsigma}(n) a_{mn}^{\varsigma\varsigma} \times \prod_{c=1, c\neq\varsigma}^{C} \sum_{n_c=1}^{M(c)} a_{n_c n}^{c\varsigma} \alpha_{t|t-1}^{c}(n_c) \tilde{b}_{n_c}^{c}(o_t) \right\} \tag{3.69}$$

The initial condition of parameters are $\beta_T^c(m) = \tilde{b}_m^c(o_T)$ and $\alpha_{1|0}^c = \pi_m^c b_m^c(o_1)$.

3.3.1.2 Three main issues

We resume hereby the three main issues treated in the context of hidden Markov models.

Problem 1: Evaluation of the likelihood of an observation sequence, given the model parameters:

$$P(o_{1:T}) = P(o_1) \prod_{t=2}^{T} P(o_t|o_{1:t-1}) \tag{3.70}$$

where

$$P(o_t|o_{1:t-1}) = \prod_{c=1}^{C} P(o_t^c|o_{1:t-1}) \tag{3.71}$$

the latter can be calculated by Eq. (3.50).

Problem 2: Finding the optimal state m_c^* that generates the observations. We can use the maximum a posteriori (MAP) estimation:

$$m_c^*(t) = \arg\max_m P(v_t^c(m)|o_{1:T}) = \arg\max_m \alpha_{t|T}^c \tag{3.72}$$

to construct the optimal sequence of channel c: $Q_c^* = \{m_c^*(1), m_c^*(2), \cdots, m_c^*(T)\}$.

Problem 3: Learning of model parameters $\lambda = \{a_{m'm}^{c'c}, \mu_m^c, \sigma_m^c, \pi_m^c\}$ that best fit the observation $\{o_{1:T}\}$. The Baum-Welch algorithm [Baum et al., 1970] can be applied, using the ML criterium. As a special case of the EM algorithm, the likelihood is estimated using the forward-backward parameters in the E-step while the M-step uses expected counts of transitions and observations to re-estimate the parameters λ (see Fig. 3.3). This procedure is performed iteratively until convergence.

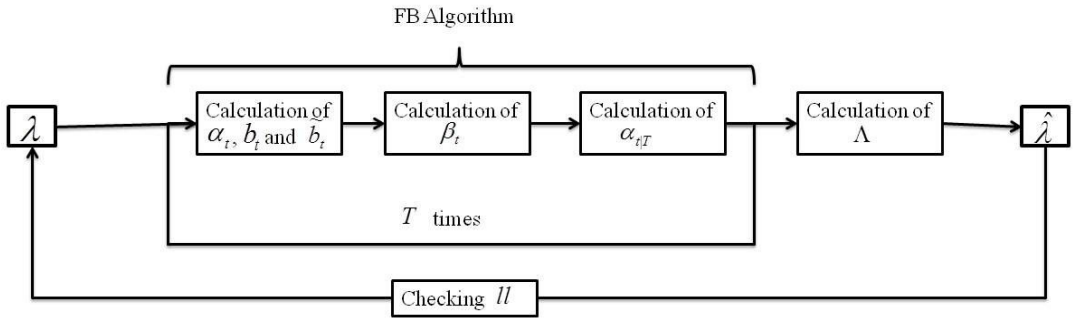

Figure 3.3: Overall EM algorithm structure depicting the use of FB algorithm.

Liporace generalizes the Baum-Welch algorithm to a larger class of distributions of observation probability than just Gaussian distribution [Liporace, 1982]. In the following, we present a generalization of the re-estimation method inspired from the Liporace paper. The likelihood probability of the observations $\{o_{1:T}\}$ given the model parameters λ writes:

$$P_\lambda(o_{1:T}) = \sum_{\mathcal{S}} P_\lambda(o_{1:T}, \mathcal{S}) \tag{3.73}$$

while $P_\lambda(o_{1:T}, \mathcal{S})$ can be written as:

$$P_\lambda(o_{1:T}, \mathcal{S}) = \prod_{c=1}^{C} \{\pi_1^c b_1(o_1^c) \times \prod_{t=2}^{T} \prod_{c'=1}^{C} a_{m_{t-1}m_t}^{c'c} b_{m_t}(o_t^c)\} \tag{3.74}$$

where $\mathcal{S}$ is the set of the state sequences of all channels. The objective is to maximize $P_\lambda(o_{1:T})$ over all parameters λ. The algorithm starts with an initial guess of λ_0, and then updates it to ensure that $P_{\lambda_{k+1}}(o_{1:T}) \geq P_{\lambda_k}(o_{1:T})$ in each iteration. As in the EM algorithm structure, an *auxiliary function* $Q(\lambda, \bar{\lambda})$ is used, defined by ([Liporace, 1982]):

$$Q(\lambda, \bar{\lambda}) = \sum_{\mathcal{S}} P_\lambda(o_{1:T}, \mathcal{S}) \log(P_{\bar{\lambda}}(o_{1:T}, \mathcal{S})). \tag{3.75}$$

The reason to use $Q(\lambda, \bar{\lambda})$ is related to the fact that $Q(\lambda, \bar{\lambda}) > Q(\lambda, \lambda)$ implies that $L_{\bar{\lambda}}(o_{1:T}) > L_\lambda(o_{1:T})$. This can be proved in case we notice that $\log x \leq x - 1$, with equality if and only if $x = 1$. Therefore

$$\begin{aligned} Q(\lambda, \bar{\lambda}) - Q(\lambda, \lambda) = L_\lambda(o_{1:T}) \log(\frac{L_{\bar{\lambda}}(o_{1:T})}{L_\lambda(o_{1:T})}) &\leq L_\lambda(o_{1:T})(\frac{L_{\bar{\lambda}}(o_{1:T})}{L_\lambda(o_{1:T})} - 1) \\ &= L_{\bar{\lambda}}(o_{1:T}) - L_\lambda(o_{1:T}) \end{aligned} \tag{3.76}$$

Extra normalization conditions include:

$$\sum_{m=1}^{M(c)} \bar{a}_{m'm}^{c'c} = 1, \text{ and } \sum_{m=1}^{M(c)} \bar{\pi}_m^c = 1 \tag{3.77}$$

for all channels (c) and states of channel (m'). These constraints are integrated using the Lagrange multiplier method to yield the maximization problem:

$$\lambda_{\text{new}} = \arg\max_{\bar{\lambda}} Q(\lambda, \bar{\lambda}) + \sum_{c=1}^{C} \theta_c (\sum_{m=1}^{M(c)} \bar{a}_{m'm}^{c'c} - 1) + \sum_{c=1}^{C} \varepsilon_c (\sum_{m=1}^{M(c)} \bar{\pi}_m^c - 1) \tag{3.78}$$

where θ_c, ε_c are the Lagrange parameters.

A description of the re-estimation procedure is detailed in the following part 3.3.1.2-3 [Liporace, 1982].

-Re-estimation of transition matrices: Differentiate Eq. (3.78) with respect to each $\bar{a}_{m'm}^{c'c}$ to obtain

$$\sum_{\mathcal{S}} P_\lambda(o_{1:T}, \mathcal{S}) \sum_{t \in \mathfrak{T}_{m'm}^{c'c}} \frac{1}{\bar{a}_{m'm}^{c'c}} - \theta_c = 0 \tag{3.79}$$

where Interchange the order of summations as in [Liporace, 1982] to get

$$\bar{a}_{m'm}^{c'c} \theta_c = \sum_{t=1}^{T} \sum_{\mathcal{S} \in \mathfrak{S}_{m'm}^{c'c}(t)} P(o_{1:T}, \mathcal{S}) = \sum_{t=1}^{T} P(o_{1:T}, q_{t-1}^{c'} = S_{m'}, q_t^c = S_m) \tag{3.80}$$

where $\mathfrak{S}_{m'm}^{c'c}(t) = \{\mathcal{S} : q_{t-1}^{c'} = S_{m'}, q_t^c = S_m\}$ denotes the set of state sequence having state m' at time $t-1$ and state m at time t. Summing over all states of channel c, we obtain θ_c.

$$\theta_c = \sum_{k=1}^{T} \sum_{m=1}^{M(c)} P(o_{1:T}, q_{t-1}^{c'} = S_{m'}, q_t^c = S_m) \tag{3.81}$$

Furthermore, we define $\Lambda_t^{c'c}(m', m) = P(q_{t-1}^{c'} = S_{m'}, q_t^c = S_m | o_{1:T})$ as the conditional smoothed transition probability:

$$\begin{aligned} \Lambda_t^{c'c}(m', m) &= P(v_{t-1}^{c'}(m'), v_t^c(m) | o_{1:T}) = \frac{P(v_{t-1}^{c'}(m'), v_t^c(m), o_{t:T} | o_{1:t-1})}{P(o_{t:T} | o_{1:t-1})} \\ &= \beta_t^c(m) P(v_t^c(m) | v_{t-1}^{c'}(m')) P(v_{t-1}^{c'}(m') | o_{1:t-1}) = \beta_t^c(m) a_{m'm}^{c'c} \alpha_{t-1|t-2}^{c'}(m') \tilde{b}_{m'}(o_{t-1}^{c'}) \end{aligned} \tag{3.82}$$

It is then straightforward to achieve the re-estimation for the transition probability:

$$\bar{a}_{m'm}^{c'c} = \frac{\sum_{t=1}^{T} \Lambda_t^{c'c}(m', m)}{\sum_{t=1}^{T} \sum_{m''=1}^{M(c)} \Lambda_t^{c'c}(m', m'')} \tag{3.83}$$

-Re-estimation of $\bar{\pi}_m^c$: Adding the appropriate constraint for initial probability of being in state m to the auxiliary function, we have

$$\frac{\partial}{\partial \bar{\pi}_m^c} \{Q(\lambda, \bar{\lambda}) - \varepsilon_c (\Sigma_{m=1}^{M(c)} \bar{\pi}_m^c - 1)\} = 0 \tag{3.84}$$

$$\sum_{\mathcal{S}} P_\lambda(o_{1:T}, \mathcal{S}) \frac{1}{\bar{\pi}_m^c} - \varepsilon_c = 0$$

$$\bar{\pi}_m^c \varepsilon_c = P(o_{1:T}, q_1^c = S_m)$$

$$\varepsilon_c = \sum_{m=1}^{M(c)} P(o_{1:T}, q_1^c = S_m)$$

Thus, using Eq. (3.66), the re-estimation of initial probability is straightforward

$$\bar{\pi}_m^c = \frac{\alpha_{1|T}^c(m)}{\sum_{m'=1}^{M(c)} \alpha_{1|T}^c(m')} \tag{3.85}$$

-Re-estimation of $\bar{\mu}_m^c$ and $\bar{\sigma}_m^c$: As previously mentioned, the density functions of observations $b_m(o_t^c)$ are assumed to be Gaussian functions. Applying the derivation on auxiliary function with respect to the components of $\bar{\mu}_m^c$, we can obtain:

$$\bar{\mu}_m^c = \frac{\sum_{t=1}^{T} \alpha_{t|T}^c(m) o_t^c}{\sum_{t=1}^{T} \sum_{m'=1}^{M(c)} \alpha_{t|T}^c(m')} \tag{3.86}$$

$$\bar{\sigma}_m^c = \frac{\sum_{t=1}^{T} \alpha_{t|T}^c(m)(o_t^c - \bar{\mu}_m^c)^2}{\sum_{t=1}^{T} \sum_{m'=1}^{M(c)} \alpha_{t|T}^c(m')} \tag{3.87}$$

3.3.2 CHSMM

Let $S_1^c, S_2^c, \cdots, S_{M(c)}^c$ be the states of a semi-Markov chain in channel c mentioned by the superscript in this paper, with the initial distribution $\{\pi_m^c\}$ and q_t^c the state of the model at time t with $t = 1, 2, \cdots, T$. o_t^c denotes the observation of channel c with the conditional probability $b_m(o_t^c) = P(o_t^c | q_t^c = S_m^c)$ and the transition probability matrix is defined as $a_{nm}^{c'c} = P(q_t^c = S_m^c | q_t^{c'} = S_n^{c'})$. The duration of a given state is a discrete variable, having value d with probability $p_m^c(d)$, where $d \in \{1, 2, \cdots, D(c)\}$.

Each semi-Markov chain remains in a state for d samples of time so $\tau_t^c = d$ defined as the remaining sojourn time of the current state ($q_t^c = S_m^c$) which denotes that cth semi-Markov chain stays in S_m^c from t to $t+d-1$ and transits to another state at time $t+d$. So a coupled hidden semi Markov model is specified by, say $\lambda = \{\{\pi_m^c\}, p_m^c(d), a_{nm}^{c'c}, b_m^c(o_t)\}$. All the probabilities which will be supposed in the following parts are conditioned on the parameter set that we will omit it due to simplicity.

In our model, we assumed that o_t^c is only depended on q_t^c. Also, we chose to use Brand's original formulation such that the interaction between channel i and j is independent of interaction between i and k. Another important assumption is that only when there is a state transition in a given channel, the states of all other channels affect the transition. Also, the sojourn durations in each channel are considered as independent random variables without specific superscript. We note that o_t denoted the set of observations of all channels at time t. In addition, let's define $v_t^\varsigma(m, d)$ equal to the ordered pair $(q_t^\varsigma = S_m^\varsigma, \tau_t^\varsigma = d)$. For the rest of the chapter, the superscript of the state variable is omitted due to the simplicity.

3.3.2.1 FB recursions

In order to generalize the Baum-Welch method to re-estimate the parameters of CHSMM, it is essential to define the forward and backward variables and derive the recursion formulas for the ease of calculations which are needed in solving the three standard

problems of HMM. It should be noted that when there is a state transition in a channel, the other channels affect it, Otherwise, we just consider the intra-channel effect. The Forward parameter is defined as the probability of being in state m of channel ς at time t and resting in it till time $t+d$ subjected to the observations of all channels from the beginning to time x.

$$\alpha^{\varsigma}_{t|x}(m,d) = P(v^{\varsigma}_t(m,d)|o_{1:x}) \tag{3.88}$$

where if $x = t-1, t$ and T, *predicted*, *filtered* and *smoothed* definition of forward parameter are achieved respectively. We can write the recursion based on predicted form of forward parameter at time t ($\alpha^{\varsigma}_{t|t-1}(m,d)$) by considering the forward parameter at time $t-1$ ($\alpha^{c}_{t-1|t-2}(m,d)$). In order to follow the transitions correctly, the value of τ^{ς}_{t-1} should be inspected. Since Two possible situations can happen, we define $\mathcal{A}^{\varsigma}_t(m,d)$ and $\mathcal{B}^{\varsigma}_t(m,d)$ which represent resting and transition effect respectively:

Case 1: $\tau^{\varsigma}_{t-1} = d_{\varsigma} \neq 1$ *consequently we have only the intra channel effect.* Resting parameter is defined as:

$$\begin{aligned}\mathcal{A}^{\varsigma}_t(m,d) &= P(v^{\varsigma}_t(m,d), \tau^{\varsigma}_{t-1} > 1|o_{1:t-1}) = P(v^{\varsigma}_{t-1}(m,d+1)|o_{1:t-1}) \\ &= \alpha^{\varsigma}_{t-1|t-1}(m,d+1) \\ &\qquad \text{cf the definition in (2)}\end{aligned} \tag{3.89}$$

To achieve the recursion, we can use $\tilde{}b_M$ parameter as follows:

$$\tilde{b}^{\varsigma}_m(o_t) \triangleq \frac{\alpha^{\varsigma}_{t|t}(m,d)}{\alpha^{\varsigma}_{t|t-1}(m,d)} = \frac{P(o_t|v^{\varsigma}_t(m,d), o_{1:t-1})}{P(o_t|o_{1:t-1})} = \frac{P(o_t|v^{\varsigma}_t(m,d), o_{1:t-1})}{\prod_{c=1}^{C} P(o^c_t|o_{1:t-1})} \tag{3.90}$$

where the last derivation is based on the mean field assumption for the observation of different channels,

$$P(o_t|o_{1:t-1}) = \prod_{c=1}^{C} P(o^c_t|o_{1:t-1}) \tag{3.91}$$

Assuming the effects of other channels states, the nominator of (3.90) is summarized as the product of distinct channel effects:

$$P(o_t|v^{\varsigma}_t(m), o_{1:t-1}) = \underbrace{\sum_{n_1=1}^{M(1)} \sum_{d_1=1}^{D(1)} \cdots \sum_{n_C=1}^{M(C)} \sum_{d_C=1}^{D(C)}}_{C-1 \text{ (except channel } \varsigma)} P(o_t, v^1_t(n_1,d_1), \cdots, v^C_t(n_C,d_C)|v^{\varsigma}_t(m,d), o_{1:t-1}) \tag{3.92}$$

$$= \underbrace{\sum_{n_1=1}^{M(1)} \sum_{d_1=1}^{D(1)} \cdots \sum_{n_C=1}^{M(C)} \sum_{d_C=1}^{D(C)} \prod_{c=1}^{C}}_{C-1 \text{ (except channel } \varsigma)} P(o^c_t|v^c_t(n_c,d_c)) \times P(v^1_t(n_1,d_1), \cdots, v^C_t(n_C,d_C)|v^{\varsigma}_t(m,d), o_{1:t-1}) \tag{3.93}$$

since the probability calculation of o^c_t only needs the corresponding state, i.e. $v^c_t(n_c,d_c)$, we can substitute $P(o^c_t|v^c_t(n_c,d_c))$ by $b_{n_c}(o^c_t)$. In the appendix, we will deal with the predicted independency of the states given the observations:

$$P(v^1_t(n_1,d_1), \cdots, v^C_t(n_C,d_C)|o_{1:t-1}) = \prod_{c=1}^{C} P(v^c_t(n_c,d_c)|o_{1:t-1}) \tag{3.94}$$

Based on this assumption, we have:

$$P(v_t^1(n_1,d_1),\cdots,v_t^C(n_C,d_C)|v_t^\varsigma(m,d),o_{1:t-1}) = \frac{\prod_{c=1}^{C} P(v_t^c(n_c,,d_c)|o_{1:t-1})}{P(v_t^\varsigma(m,d)|o_{1:t-1})}$$
$$= \prod_{c=1,c\neq\varsigma}^{C} \alpha_{t|t-1}^c(n_c,d_c) \tag{3.95}$$

Accordingly, we can establish the following equation:

$$P(o_t|v_t^\varsigma(m),o_{1:t-1}) = b_m(o_t^\varsigma)\prod_{c=1,c\neq\varsigma}^{C}\sum_{n_c=1}^{M(c)}\sum_{d_c=1}^{D(c)} b_{n_c}(o_t^c)\alpha_{t|t-1}^c(n_c,d_c)$$
$$= b_m(o_t^\varsigma)\prod_{c=1,c\neq\varsigma}^{C} P(o_t^c|o_{1:t-1}) \tag{3.96}$$

Hence, $\tilde{b}_m^\varsigma(o_t)$ is given by:

$$\tilde{b}_m^\varsigma(o_t) = \frac{b_m(o_t^\varsigma)}{P(o_t^\varsigma|o_{1:t-1})} \tag{3.97}$$

Note that the value of the $\tilde{b}_m^\varsigma(o_t)$ is nearly 1 if the model learns the observation accurately. Therefore, it is easy to show that:

$$\mathcal{A}_t^\varsigma(m,d) = \alpha_{t-1|t-2}^\varsigma(m,d+1)\tilde{b}_m^\varsigma(o_{t-1}) \tag{3.98}$$

Case 2: $\tau_{t-1}^\varsigma = d_\varsigma = 1$ *so both cross and intra channel effect constitute the transition parameter.* knowing that there is a transition in channel ς, we consider the effect of other channels as follows:

$$\mathcal{B}_t^\varsigma(m,d) = P(v_t^\varsigma(m,d),\tau_{t-1}^\varsigma = 1|o_{1:t-1}) \tag{3.99}$$
$$= \sum_{n_1=1}^{M(1)}\sum_{d_1=1}^{D(1)}\cdots\sum_{n_C=1}^{M(C)}\sum_{d_C=1}^{D(C)} P(v_t^\varsigma(m,d),v_{t-1}^1(n_1,d_1),\cdots,v_{t-1}^C(n_C,d_C)|o_{1:t-1}) \tag{3.100}$$

Applying the Bayes theorem gives:

$$P(v_t^\varsigma(m,d),v_{t-1}^1(n_1,d_1),\cdots,v_{t-1}^C(n_C,d_C)|o_{1:t-1})$$
$$= P(v_t^\varsigma(m,d)|v_{t-1}^1(n_1,d_1),\cdots,v_{t-1}^C(n_C,d_C),o_{1:t-1})$$
$$\times P(v_{t-1}^1(n_1,d_1),\cdots,v_{t-1}^C(n_C,d_C)|o_{1:t-1}) \tag{3.101}$$

and using Brand's assumption over the first term, we derive the following equation:

$$\mathcal{B}_t^\varsigma(m,d) = \sum_{n_1=1}^{M(1)}\sum_{d_1=1}^{D(1)}\cdots\sum_{n_C=1}^{M(C)}\sum_{d_C=1}^{D(C)}\prod_{c=1}^{C} P(v_t^\varsigma(m,d)|v_{t-1}^c(n_c,d_c))$$
$$\times P(v_{t-1}^1(n_1,d_1),\cdots,v_{t-1}^C(n_C,d_C)|o_{1:t-1}) \tag{3.102}$$

Note that the term $P(v_t^{\varsigma}(m,d)|v_{t-1}^{c}(n_c,d_c))$ is substitute by $a_{n_c m}^{c\varsigma}$. We also use

$$P(v_{t-1}^{1}(n_1,d_1),\cdots,v_{t-1}^{C}(n_C,d_C)|o_{1:t-1}) = \prod_{c=1}^{C} P(v_{t-1}^{c}(n_c,d_c)|o_{1:t-1}) \tag{3.103}$$

due to the proves given in appendix. In addition, we organize Eq. 3.102 according to the parameters of double summations corresponded to each channel as follows:

$$\mathcal{B}_t^{\varsigma}(m,d) = \prod_{c=1}^{C}\sum_{n_c=1}^{M(c)}\sum_{d_c=c}^{D(c)} P(v_t^{\varsigma}(m,d)|v_{t-1}^{c}(n_c,d_c)) \times P(v_{t-1}^{c}(n_c,d_c)|o_{1:t-1}) \tag{3.104}$$

Hence, $\mathcal{B}_t^{\varsigma}(m,d)$ can be established as follows:

$$\begin{aligned}\mathcal{B}_t^{\varsigma}(m,d) = p_m^{\varsigma}(d)&\sum_{n_\varsigma=1}^{M(\varsigma)}\{a_{n_\varsigma m}^{\varsigma\varsigma}\alpha_{t-1|t-2}^{\varsigma}(n_\varsigma,1)\tilde{b}_{n_\varsigma}^{\varsigma}(o_{t-1})\}\\ &\times\prod_{c=1,c\neq\varsigma}^{C}\sum_{n_c=1}^{M(c)}\sum_{d_c=1}^{D(c)}\{a_{n_c m}^{c\varsigma}\alpha_{t-1|t-2}^{c}(n_c,d_c)\tilde{b}_{n_c}^{c}(o_{t-1})\}\end{aligned} \tag{3.105}$$

and finally, an obvious summation gives the forward parameter recursion:

$$\alpha_{t|t-1}^{\varsigma}(m,d) = \mathcal{A}_t^{\varsigma}(m,d) + \mathcal{B}_t^{\varsigma}(m,d) \tag{3.106}$$

By considering the initialization of this recursion as $\alpha_{1|0}^{\varsigma}(m,d) = \pi_m^{\varsigma} p_m^{\varsigma}(d)$, we can complete the calculation of the forward variable. Moreover, $P(o_t^c|o_{1:t-1})$ can be calculated by the means of this variable:

$$P(o_t^c|o_{1:t-1}) = \sum_{m=1}^{M(c)}\sum_{d=1}^{D(c)} \alpha_{t|t-1}^{c}(m,d) b_m^c(o_t) \tag{3.107}$$

The Backward parameter is defined as a normalized form of the probability of observation at time t to T of channel ς subjected to staying at state m for d more samples of time. Similarly, the state transitions in a channel result to different equations for backward parameter. $d = 1$ is indicating a transition so $\beta_t^{\varsigma}(m,d)$ is obtained by considering all channels impact. It is generally defined as:

$$\begin{aligned}\beta_t^{\varsigma}(m,d) &\triangleq \frac{P(v_t^{\varsigma}(m,d)|o_{1:T})}{P(v_t^{\varsigma}(m,d)|o_{1:t-1})}\\ &= \frac{P(o_{t:T}, v_t^{\varsigma}(m,d)|o_{1:t-1})}{P(o_{t:T}|o_{1:t-1})P(v_t^{\varsigma}(m,d)|o_{1:t-1})}\\ &= \frac{P(o_{t:T}|v_t^{\varsigma}(m,d),o_{1:t-1})}{P(o_{t:T}|o_{1:t-1})}\\ &= \tilde{b}_m^{\varsigma}(o_t)\frac{P(o_{t+1:T}|v_t^{\varsigma}(m,d),o_{1:t})}{P(o_{t+1:T}|o_{1:t})}\end{aligned} \tag{3.108}$$

Case 1: $\tau_t^{\varsigma} = d \neq 1$ consequently we have only the intra channel effect.

In this case, it is simple to show as follows:

$$\begin{aligned}\beta_t^{\varsigma}(m,d) &= \tilde{b}_m^{\varsigma}(o_t)\frac{P(o_{t+1:T}|v_{t+1}^{\varsigma}(m,d-1),o_{1:t})}{P(o_{t+1:T}|o_{1:t})} \\ &= \tilde{b}_m^{\varsigma}(o_{t-1})\beta_{t+1}^{\varsigma}(m,d-1)\end{aligned} \tag{3.109}$$

Case 2: $\tau_t^{\varsigma} = d = 1$ *other channels impact on current channel.*

The current state $v_t^{\varsigma}(m,1)$ transits to a new one $v_{t+1}^{\varsigma}(n_{\varsigma},d_{\varsigma})$, so backward recursion is derived based on the next transition in channel ς as follows:

$$\beta_t^{\varsigma}(m) = \tilde{b}_m^{\varsigma}(o_t)\sum_{n_{\varsigma}=1}^{M(\varsigma)}\sum_{d_{\varsigma}=1}^{M(\varsigma)}\frac{P(o_{t+1:T},v_{t+1}^{\varsigma}(n_{\varsigma},d_{\varsigma})|v_t^{\varsigma}(m,1),o_{1:t})}{P(o_{t+1:T}|o_{1:t})} \tag{3.110}$$

$$= \tilde{b}_m^{\varsigma}(o_t)\sum_{n_{\varsigma}=1}^{M(\varsigma)}\sum_{d_{\varsigma}=1}^{M(\varsigma)}\frac{P(o_{t+1:T}|v_{t+1}^{\varsigma}(n_{\varsigma},d_{\varsigma}),v_t^{\varsigma}(m,1),o_{1:t})}{P(o_{t+1:T}|o_{1:t})}\times P(v_{t+1}^{\varsigma}(n_{\varsigma},d_{\varsigma})|v_t^{\varsigma}(m,1),o_{1:t})$$

The first term after summation can be interpreted as $\beta_{t+1}^{\varsigma}(n_{\varsigma},d_{\varsigma})$ according to 3.108 and $P(o_{t+1:T}|v_{t+1}^{\varsigma}(n),v_t^{\varsigma}(m),o_{1:t}) = P(o_{t+1:T}|v_{t+1}^{\varsigma}(n),o_{1:t})$. The last term should be calculated correctly as follows including the effects of other channels:

$$\begin{aligned}&P(v_{t+1}^{\varsigma}(n_{\varsigma},d_{\varsigma})|v_t^{\varsigma}(m,1),o_{1:t}) \\ &=\underbrace{\sum_{n_1=1}^{M(1)}\sum_{d_1=1}^{D(1)}\cdots\sum_{n_C=1}^{M(C)}\sum_{d_C=1}^{D(C)}}_{C-1\text{ (except channel }\varsigma)}P(v_{t+1}^{\varsigma}(n_{\varsigma},d_{\varsigma}),v_t^1(n_1,d_1),\cdots,v_t^C(n_C,n_C)|v_t^{\varsigma}(m,1),o_{1:t}) \\ &=\underbrace{\sum_{n_1=1}^{M(1)}\sum_{d_1=1}^{D(1)}\cdots\sum_{n_C=1}^{M(C)}\sum_{d_C=1}^{D(C)}}_{C-1\text{ (except channel }\varsigma)}P(v_{t+1}^{\varsigma}(n_{\varsigma},d_{\varsigma})|v_t^1(n_1,d_1),\cdots,v_t^C(n_C,d_C),v_t^{\varsigma}(m,1),o_{1:t}) \\ &\qquad\times P(v_t^c(n_c,d_c),\cdots,v_t^C(n_C,d_C)|v_t^{\varsigma}(m,1),o_{1:t})\end{aligned} \tag{3.111}$$

which can be simply written as follows:

$$P(v_{t+1}^{\varsigma}(n_{\varsigma},d_{\varsigma})|v_t^{\varsigma}(m,1),o_{1:t}) = a_{mn_{\varsigma}}^{\varsigma\varsigma}p_{n_{\varsigma}}^{\varsigma}(d_{\varsigma})\prod_{c=1,c\neq\varsigma}^{C}\sum_{n_c=1}^{M(c)}a_{n_cn_{\varsigma}}^{c\varsigma}p_{n_c}^{c}(d_c)\alpha_{t|t-1}^{c}(n_c)\tilde{b}_{n_c}^{c}(o_t) \tag{3.112}$$

Thus the backward recursion is given by:

$$\beta_t^{\varsigma}(m,1) = \tilde{b}_m^{\varsigma}(o_t)\sum_{n_{\varsigma}=1}^{M(\varsigma)}\sum_{d_{\varsigma}=1}^{D(\varsigma)}\left\{\beta_{t+1}^{\varsigma}(n_{\varsigma})a_{mn}^{\varsigma\varsigma}p_{n_{\varsigma}}^{\varsigma}(d_{\varsigma})\prod_{c=1,c\neq\varsigma}^{C}\sum_{n_c=1}^{M(c)}\sum_{d_c=1}^{D(c)}a_{n_cn}^{c\varsigma}p_{n_c}^{c}(d_c)\alpha_{t|t-1}^{c}(n_c)\tilde{b}_{n_c}^{c}(o_t)\right\} \tag{3.113}$$

Finally a recursion to calculate backward parameter with the initial value $\beta_t^{\varsigma}(m,1) = \tilde{b}_m^{\varsigma}(o_T)$ for all d is achieved.

3.3.2.2 Three main issues

Similar to CHMM, in addition to FB parameter, we further introduce four more variables for convenience in re-estimation part. Firstly, The conditional probability of transition to state m in channel ς while the channel c is in state n given the observation of all channels writes:

$$\begin{aligned}\Upsilon_t^{c\varsigma}(n,m) =& \sum_{d=1}^{D(\varsigma)} P(q_{t-1}^c = S_n, v_t^\varsigma(m,d), \tau_{t-1}^\varsigma = 1 | o_{1:T}) \\ =& \sum_{d=1}^{D(\varsigma)} \beta_t^\varsigma(m,d) a_{mn}^{c\varsigma} p_m^\varsigma(d) \sum_{d'=1}^{D(c)} \alpha_{t-1|t-2}^c(n,d') \widetilde{b}_n^c(o_{t-1})\end{aligned} \tag{3.114}$$

while ignoring the impact of other channels. Then, the conditional probability of transition to state m at time t and rest in till $t+d$ given the observation of all time and channels. Separating $o_{t:T}$ gives us a trick to calculate this variable.

$$\delta_t^\varsigma(m,d) = P(\tau_{t-1}^\varsigma = 1, v_t^\varsigma(m,d) | o_{1:t-1}) = \frac{P(\tau_{t-1}^\varsigma = 1, v_t^\varsigma(m,d), o_{t:T} | o_{1:t-1})}{P(o_{t:T} | o_{1:t-1})} \tag{3.115}$$

hence, it is straightforward to obtain:

$$\begin{aligned}\delta_t^\varsigma(m,d) = \beta_t^\varsigma(m,d) p_m^\varsigma(d) \times & \sum_{n=1}^{M(\varsigma)} \alpha_{t-1|t-2}^\varsigma(n,1) \widetilde{b}_n^\varsigma(o_{t-1}) a_{nm}^{\varsigma\varsigma} \\ & \times \prod_{c=1}^{C} \sum_{n_c=1}^{M(c)} \sum_{d_c=1}^{D(c)} \alpha_{t-1|t-2}^c(n_c,d_c) \beta_{t-1}^c(n_c,d_c) a_{n_c m}^{c\varsigma} \qquad (3.116) \\ = & \beta_t^\varsigma(m,d) p_m^\varsigma(d) \mathcal{B}_t^\varsigma(m,d) \qquad (3.117)\end{aligned}$$

Furthermore, we can readily calculate the smoothed version of the forward variables as a product of forward and backward variables:

$$\alpha_{t|T}^\varsigma(m,d) = P(v_t^\varsigma(m,d) | o_{1:T}) = \alpha_{t|t-1}^\varsigma(m,d) \beta_t^\varsigma(m,d) \tag{3.118}$$

The summation on d defines another variables:

$$\gamma_t^\varsigma(m) = \sum_d \alpha_{t|T}^\varsigma(m,d) \tag{3.119}$$

Problem 1: Evaluation of the likelihood of an observation sequence, given the model parameters:

$$P(o_{1:T}) = P(o_1) \prod_{t=2}^{T} P(o_t | o_{1:t-1}) \tag{3.120}$$

where

$$P(o_t | o_{1:t-1}) = \prod_{c=1}^{C} P(o_t^c | o_{1:t-1}) \tag{3.121}$$

where the latter can be calculated by Eq. 3.91.

Problem 2: Finding the optimal state m_c^* that generates the observations. We can use the MAP estimation for inference of CHSMM. First, we calculate $\delta_t^s(m,d)$ for $t = 1, 2, \cdots, T$ and over all states and durations in all channels. Then we begin from $t_0 = 1$ and find m and d which gives us the maximum of $\delta_{t_0}^s(m,d)$. Consequently, the optimal state sequence assigned as $Q_{max}(t_0 : t_0 + d)$ and we continue for the rest of samples of time. This algorithm is illustrated for more detail.

Algorithm 3.4: Inference

Inputs: $\delta_t^s(m,d) = P(\tau_{t-1}^s = 1, q_t^s = S_m, \tau_t^s = d|o_{1:T})$.
Outputs: likelihood probability $ll^c = \log(P(o_{1:T}^c))$ and the series of states $Q_{max}(t)$

1: For each state (m) and duration (d), compute $\alpha_{t|t-1}^c(m,d)$ (Eq. 3.106),
2: For each state (m) and duration (d), compute $\beta_t^c(m,d)$ (Eq. 3.109, 3.113),
3: For each state (m) and duration (d), compute $\delta_t^c(m,d)$ (Eq. 3.116),
4: Start from $t = 1$
5: **repeat**
6: Find the state $(m^*)$and duration (d^*) which give the maximum $(m^*, d^*) = \arg\max_{m,d}\{\delta_t^c(m,d)\}$.
7: $Q_{max}^c(t : t + d^*) = m^*$.
8: $t = t + d^* + 1$.
9: **until** $t \geq T$
10: Calculate ll^c from $\log\left(\Pi_{t=1}^T b_{Q_{max}(t)}^c(o_t)\right)$.

Problem 3: Learning of model parameters $\lambda = \{a_{m'm}^{c'c}, \mu_m^c, \sigma_m^c, \pi_m^c\}$ that best fit the observation $\{o_{1:T}\}$. The Baum-Welch algorithm [Baum et al., 1970] can be applied, using the ML criterium.

To learn the parameters of HMM according to ML criterium from the given observations $\{o_{1:T}\}$, we use similarly our approach in proposed CHMM learning procedure.

We assume that the set of model parameters λ is given and there are K_c transitions happened in the Markov chain of channel c. The likelihood probability of the observations $\{o_{1:T}\}$ given λ in all channels can be written as follows:

$$P_\lambda(o_{1:T}) = \sum_{\mathcal{S}} P_\lambda(o_{1:T}, \mathcal{S}) \tag{3.122}$$

and $P_\lambda(o_{1:T}, \mathcal{S})$ can be written as:

$$P_\lambda(o_{1:T}, \mathcal{S}) = \prod_{c=1}^{C}\{\pi_{m_0}^c p_{m_0}^c(d_0) \prod_{t_0=1}^{d_0}(b_{m_0}^c(o_{t_0})) \times \prod_{k=1}^{K_c}(p_{m_k}^c(d_k) a_{m_{k-1}m_k}^{cc} \prod_{c_k=1, c_k\neq c}^{C}(a_{n_k m_k}^{c_k c}) \prod_{t_k=d_0+\cdots+d_{k-1}+1}^{d_0+\cdots+d_k}(b_{m_k}^c(o_{t_k})))\} \tag{3.123}$$

where $\sum_{k=0}^{K_c}(d_k) = T$ in channel c.

In addition to the conditions in CHMM, a constraint on resting time parameter is maintained as:

$$\sum_{d=1}^{D(c} \bar{P}_m^c(d) = 1, \tag{3.124}$$

-Re-estimation of transition matrices: Considering 3.105, a transition probability $\bar{a}^{c\varsigma}_{m'm}$ satisfies

$$\frac{\partial}{\partial \bar{a}^{c\varsigma}_{m'm}}\{Q(\lambda,\bar{\lambda}) - \theta(\Sigma_{m=1}^{M(\varsigma)}\bar{a}^{c\varsigma}_{m'm} - 1)\} = 0 \tag{3.125}$$

where θ is the Lagrange parameter. Applying the derivation we obtain as follows:

$$\sum_{\mathcal{S}} P_\lambda(o_{1:T},\mathcal{S}) \sum_{t\in\mathfrak{T}^{c'c}_{m'm}} \frac{1}{\bar{a}^{c\varsigma}_{m'k}} - \theta = 0 \tag{3.126}$$

where $\mathfrak{T}^{c'c}_{m'm} = \{t : q^{c'}_{t-1} = S_{m'}, q^c_t = S_m\}$. If we change the order of summations similar to [Liporace, 1982], 3.126 becomes

$$\bar{a}^{c'c}_{m'm}\theta = \sum_{t=1}^{T} \sum_{\mathcal{S}\in\mathfrak{S}^{c'c}_{m'm}(t)} P(o_{1:T},\mathcal{S}) \tag{3.127}$$

$$\bar{a}^{c'c}_{m'm}\theta = \Sigma_{t=1}^{T} P(o_{1:T}, q^{c'}_{t-1} = S_{m'}, q^c_t = S_m) \tag{3.128}$$

where $\mathfrak{S}^{c'c}_{m'm}(t) = \{\mathcal{S} : q^{c'}_{t-1} = S_{m'}, q^c_t = S_m\}$, the set of state sequence which contains state m' at time $t-1$ and state m at time t. Summing over all state of channel c, we obtain θ

$$\theta = \Sigma_{k=1}^{T}\Sigma_{m''=1}^{M(c)} P(o_{1:T}, q^{c'}_{t-1} = S_{m'}, q^c_t = S_{m''}) \tag{3.129}$$

and the re-estimation equation for transition probability is achieved.

$$\bar{a}^{c'c}_{m'm} = \frac{\Sigma_{t=1}^{T}\Upsilon^{c'c}_t(m',m)}{\Sigma_{t=1}^{T}\Sigma_{m''=1}^{M(c)}\Upsilon^{c'c}_t(m',m'')} \tag{3.130}$$

-Re-estimation of $\bar{p}^c_m(d)$: The same manipulations that we perform in previous part, are used for estimation of duration probability densities.

$$\frac{\partial}{\partial \bar{p}^c_m(d)}\{Q(\lambda,\bar{\lambda}) - \theta(\Sigma_{d=1}^{D_{max}(c)}\bar{p}^c_m(d) - 1)\} = 0$$

$$\sum_{\mathcal{S}} P_\lambda(o_{1:T},\mathcal{S}) \sum_{t\in G^c_{m,d}}^{M(c)} \frac{1}{\bar{p}^c_m(d)} - \theta = 0 \tag{3.131}$$

$$\bar{p}^c_m(d)\theta = \Sigma_{t=1}^{T} P(o_{1:T}, \tau^c_{t-1} = 1, v^c_t(m,d))$$

where $G^c_{m,d} = \{t : \tau^c_{t-1} = 1, v^c_t(m,d)\}$.

Summing over all values of d in both sides, we obtain θ and $\bar{p}^\varsigma_m(d)$ as follows

$$\theta = \Sigma_{t=1}^{T}\Sigma_{d'=1}^{D(c)} P(o_{1:T}, \tau^c_{t-1} = 1, q^c_t = S_m, \tau^c_t = d')$$

$$\bar{p}^c_m(d) = \frac{\Sigma_{t=1}^{T}\delta^c_t(m,d)}{\Sigma_{t=1}^{T}\Sigma_{d'=1}^{D(c)}\delta^c_t(m,d')} \tag{3.132}$$

-Re-estimation of $\bar{\pi}_m^c$: Adding the appropriate constraint for initial probability of being in state m to the auxiliary function, we have

$$\frac{\partial}{\partial \bar{\pi}_m^c}\{Q(\lambda,\bar{\lambda}) - \theta(\Sigma_{m=1}^{M(c)}\bar{\pi}_m^c - 1)\} = 0$$

$$\sum_{\mathcal{S}} P_\lambda(o_{1:T},\mathcal{S})\frac{1}{\bar{\pi}_m^c} - \theta = 0 \tag{3.133}$$

$$\bar{\pi}_m^c\theta = P(o_{1:T}, q_1^c = S_m)$$

$$\theta = \Sigma_{m'=1}^{M(c)} P(o_{1:T}, q_1^c = S_{m'}) \tag{3.134}$$

Thus using 3.118, the re-estimation of initial probability is straightforward

$$\bar{\pi}_m^c = \frac{\Sigma_{d=1}^{D(c)}\alpha_{1|T}^c(m,d)}{\Sigma_{m'=1}^{M(c)}\Sigma_{d=1}^{D(c)}\alpha_{1|T}^c(m',d)} \tag{3.135}$$

-Re-estimation of $\bar{\mu}_m^s$ and $\bar{\sigma}_m^s$: In this work, the density functions of observations are assumed to be Gaussian function. Each $b_m^s(o_t)$ has the form, $b_m^s(o_t) = N(o_t;\mu_m^s,\sigma_m^s) = \frac{1}{\sqrt{2\pi}\sigma_m} exp\{-(x-\mu_m^s)^2/(2(\sigma_m^s)^2)\}$. Applying the derivation on auxiliary function with respect to the components of $\bar{\mu}_m^s$, we can obtained:

$$\frac{\partial Q(\lambda,\bar{\lambda})}{\partial \bar{\mu}_m^c} = 0 \qquad \frac{\partial Q(\lambda,\bar{\lambda})}{\partial \bar{\sigma}_m^c} = 0 \tag{3.136}$$

$$\bar{\mu}_m^c = \frac{\sum_{t=1}^T \gamma_t^c(m) o_t^c}{\sum_{t=1}^T \gamma_t^c(m)}$$

$$\bar{\sigma}_m^c = \frac{\sum_{t=1}^T \gamma_t^c(m)(o_t^c - \bar{\mu}_m^c)^2}{\sum_{t=1}^T \gamma_t^c(m)}$$

Considering the changes of likelihood during the execution of learning algorithm shows that after each iteration, this value is rising until it reaches near the ideal solution. Then, if the iterations continue, the likelihood probability gradually decreases and the algorithm will diverge. Also, the more the initial condition is closer to reality, the less iteration is needed to converge to the optimal answer.

3.4 The methodology of application

In our laboratory, HMM and HSMM are used in a particular algorithm to detect AB [Altuve et al., 2011a]. In this section, the initialization of the model parameters is detailed as well as the evaluation of the likelihood of each model which is generated by a sequence of observation. The mentioned Markovian frameworks can be used in classification and detection of events of interest by analyzing the likelihood. For this approach, K models are used to represent K different dynamics of all observations being analyzed (univariate or multivariate). The underlying assumption is that these dynamics are associated with states or distinct pathophysiological events, which should be discriminated. An initialization is required to obtain the initial values for model parameters prior to the learning procedure. First a K-means based method is used to

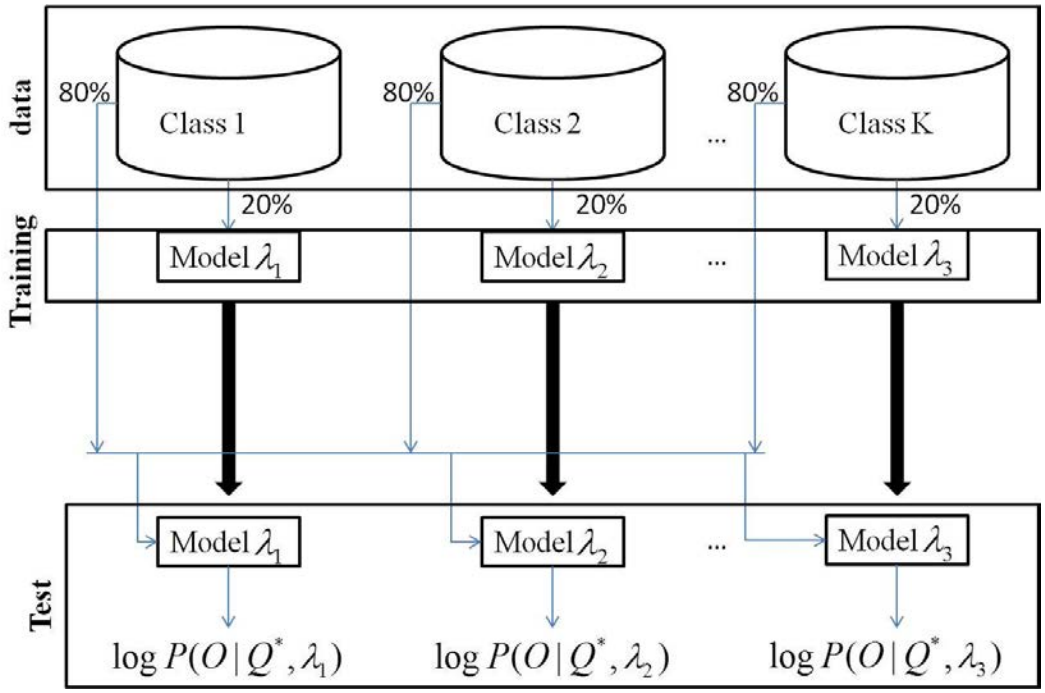

Figure 3.5: Classification algorithm.

estimate the mean (μ_m) and variance (σ_m) of the learning data corresponding to a state (m). Note that each class is signified by a state. The initial values of the rest of the model parameters are defined randomly. In following parts, we describe the scenarios of classification (on simulated data) and detection (on both simulated and real data).

3.4.1 Classification

In previous works like in [Dumont et al., 2008], the problem of classifying a time-series into one of K classes has been addressed by defining a set of K competing models, for which a learning dataset corresponding to each class is used to estimate the parameters of each model $(\lambda_1, \cdots, \lambda_K)$. Then, in the test phase, each time-series of the test dataset (o) is analyzed by calculating the log-likelihood using each model: $\log P(o \mid Q^*, \lambda_k), j \in \{1, 2, \cdots, K\}$, where Q^* represents the optimal state sequence. The classification result is obtained by choosing the class corresponding to the model presenting the maximum log-likelihood.

A similar approach is applied in this research for the proposed frameworks. However, in order to cope with the multichannel nature of the Coupled frameworks (CHMM and CHSMM), the overall log-likelihood corresponding to each class will be obtained by summing the log-likelihoods in all channels with their optimal state sequence Q_c^*:

$$ll_k = \sum_{c=1}^{C} \log\{P(o^c|Q_c^*, \lambda_k)\} \tag{3.137}$$

The classification algorithm is depicted in block diagram in Fig. 3.5.

A classification is performed on simulated data, by defining three classes: a_1, a_2 and a rest condition. We further assume that the number of states of the two competing models are equal and more than that of the model for learning the rest condition.

3.4.2 On-line Detection

The classification application can also be extended to on-line detection, as in [Dumont et al., 2008]. The data can be divided into overlapping moving windows, and the classification procedure is applied to each window. One of the K classes (i.e. class k) is defined to represent the event of interest. The difference of log-likelihood of class k from other classes can be obtained:

$$ll_{total}^{kj}(t) = ll_k(t) - ll_j(t) \tag{3.138}$$

where $j \in \{1, 2, \cdots, K\} - \{k\}$. In case of multiple channels as in our proposed coupled frameworks, the above equation can be rewritten as:

$$ll_{total}^{kj}(t) = \sum_{c=1}^{C} (ll_k^c(t) - ll_j^c(t)) \tag{3.139}$$

An event corresponding to class k takes place in a window containing time t if the following condition is satisfied

$$ll_{total}^{kj}(t) > \delta_{kj} \tag{3.140}$$

where δ_{kj} is a constant threshold that should be optimized.

In order to implement a system for monitoring premature babies, developed detection method must be adapted to the processing line. Indeed, the proposed implementation in [Dumont, 2008] is restricted to block processing and is not suited for this case. Altuve has proposed an adaptive preprocessing of AB observation series [Altuve, 2011]. He multiplies sequences of observations by a significant and arbitrary value, to overcome numerical problems during the learning phase. Furthermore, in order to increment the likelihood of a particular condition and its path in the Viterbi algorithm, the amplitude variability of observations should be minimized from the first sample. He has proposed a procedure to narrow the observation sequences around the average which is determined in a window of size L:

$$\hat{o}_{1:T} = o_{t-T+1:t} - \frac{1}{L} \sum_{t'=t-T-L+1}^{t-T} o_{t'} \tag{3.141}$$

where $\hat{o}_{1:T}$ is now the observation to take into account by the model and it is chosen by the moving window. In this case, the log-likelihood for the model k at time t is now determined by:

$$ll_k(t) = logP(\hat{o}_{1:T}|Q_k^*, \lambda_k) \tag{3.142}$$

where Q_k^* is the best state sequence obtained from inference procedure. The detection is determined by using the Eq. 3.140. Fig. 3.6 summarizes the method of on-line detection of apnea-bradycardia used in this thesis, using a sliding window on the observation.

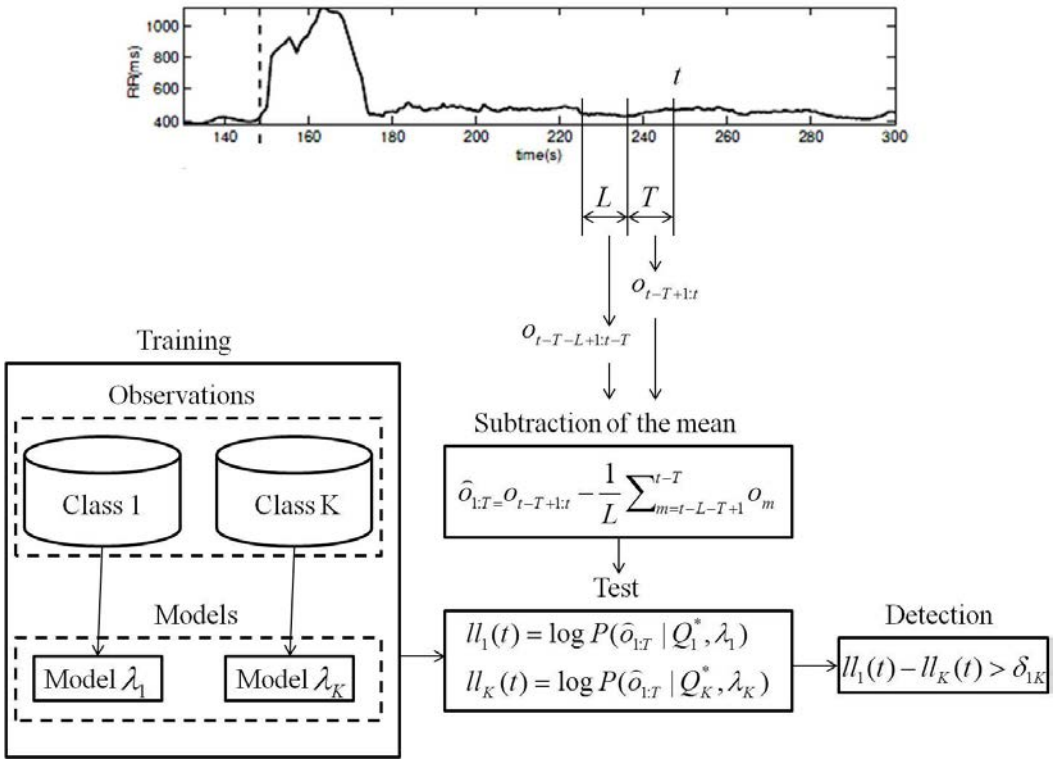

Figure 3.6: Algorithm for event of interest detection.

3.5 Conclusion

In this chapter, we review HMM framework and some of its generalizations which are related to our proposed methods. Among these generalizations, HSMM as stated in [Yu and Kobayashi, 2006] inspired the CHMM and CHSMM frameworks which are proposed in this thesis. Yu's approach to solve the HSMM issue is simple and well-defined that results in logical relations for FB algorithm. The same idea is employed in simplification of coupling issue in HMM. The effect of different channels in the state transition in one channel is not a straightforward computation in CHMM. In addition, considering the resting time parameter for the states makes the CHSMM more complex. According to this approach, the problem is broken into two situations: remaining in a state or transiting to other one. Then, for each of the cases, the FB parameters are presented. Furthermore, by defining $\tilde{b}$ parameter, the problem of overflow can be overcome and the FB parameters are rescaled. At the end of this chapter, the methodology in order to apply a Markovian framework for AB detection is also presented.

Chapter 4

Switching methods

4.1 Introduction

The morphology and dynamic of ECG signals are studied extensively using the combination of dynamical and state space models (SSM) [Sameni et al., 2007, Akhbari et al., 2012]. Moreover, they are also studied based on the sequential models like HMM [Coast, 1993]. In this chapter, we study the ECG signal containing AB episodes from the switching SSM point of view which can be accounted as a mixture of SSM and sequential models.

As discussed, apnea can greatly affect the activity of the heart. The subtle changes in the ECG patterns can be interpreted as a change happened in the dynamic of ECG beat generation process. In order to study such systems, one can propose to use SKF which is widely used for modeling systems with changeable dynamics [Ghahramani, 1998, Marculescu et al., 1998]. In a simple form of SKF, it is assumed that the model has linear dynamic at each time instant but it is time variant and switches among several linear subsystems over the time; each linear subsystem can be described by linear dynamical equations of continuous states (x_k) as *state equations* and a linear relation between states and observation (y_k) as *observation equation*:

$$\begin{aligned} x_k &= Ax_{k-1} + \gamma_k \\ y_k &= Mx_k + r_k \end{aligned} \tag{4.1}$$

γ_k and r_k are state and measurement noises. A and M are state and observation coefficients respectively. This type of modeling is generally referred to SSM. In SKF, the states are estimated by several KFs with different state or observation equations at each time instant. Meanwhile, a hidden discrete state variable called *switch* (s) is considered, so according to Markov characteristics, the status of this state changes over the time and indicates to the KF with the best state estimation performance. In general structure of SKF, both state and observation equations depend on the switch variable as follows:

$$\begin{aligned} x_k &= A(s)x_{k-1} + \gamma_k(s) \\ y_k &= M(s)x_k + r_k(s) \end{aligned} \tag{4.2}$$

where $\Gamma(s)$ and $R(s)$ are covariance matrices of γ_k and r_k respectively which are related to switch s. The inference and learning algorithms of SKF are fully described in [Murphy, 1998]. In SKF, an EM algorithm is applied to estimate the parameters of the linear models. The switching approach is used previously in other contexts such as tracking meteorological features over the time [Manfredi et al., 2005] and in a computer vision application for event detection and data collection at traffic intersections [Veeraraghavan et al., 2005, Veeraraghavan et al., 2006]. In [Pavlovic et al., 1999], an approximate Viterbi inference approach is proposed by defining a cost function based on KF parameters. They applied their method on figure motion analysis with convincing results. This algorithm is also applied in [Zheng and Hasegawa-Johnson, 2003] for segmentation of vowels, nasal, frication and silence in an acoustic signal. Depending on SKF application and the BN structure, switch may set on some of the continuous state variables rather than all of them, so that unaffected states are simply estimated by standard KF algorithm [Wu et al., 2003].In some other works where the dynamic that has generated the observation is not specified, the switch is only allocated on observation equation [Wu et al., 2004].

In this chapter, in order to model dynamic changes caused by AB occurrence in ECG, we assume two models for normal and bradycardia dynamics. Two approaches are also considered in this strategy and the results are compared with each other. The first approach is based on a research proposed by [Sameni et al., 2007], an Extended Kalman Filter (EKF) framework using McSharry's dynamical model. Hence, the method is a waveform based SEKF. This model uses the ECG and RR signals to include the HR information as well as ECG intra-beat characteristics. The second approach is a R-based SKF model, inspired from [Wu et al., 2004] by assigning the AR dynamics to each subsystem and using only RR signal as the observation.

The chapter is organized as follows. Section 4.2 provides relevant background on the ECG dynamical model based on McSharry's model and two algorithm EKF2 ([Sameni et al., 2007]) and EKF3 ([Akhbari et al., 2012]). The approaches of [Murphy, 1998] and [Wu et al., 2004] are presented in section 4.3. In section 4.4, we describe extensively our contribution including two different SKF based methods: 1- a wave-based SEKF and 2- a R-based SKF approaches for AB detection in real data.

4.2 ECG Dynamical Model

In this section, in order to be self-contained, a dynamical model of ECG and two EKF-based algorithms which are proposed previously for ECG denoising are presented.

4.2.1 McSharry's Differential Equations for ECG

Every heartbeat on an ECG record consists of finite number of characteristic waveforms (typically the P wave, QRS complex and T wave), each of which can be easily modeled by sum of Gaussian kernels. This idea is presented by [McSharry et al., 2003] in order to propose a nonlinear dynamic model for generating synthetic ECG. This model consists of three differential equations in Cartesian coordinates whose solution generates

a trajectory in 3D space around a circular limit cycle that upward and downward deviations from the plane of the circle create waves like the waveforms of an ECG beat.

4.2.2 EKF2 Algorithm

The EKF2 algorithm introduced by [Sameni et al., 2007], includes the simplified discrete form of McSharry's model in cylindrical coordinate system as follows:

$$\begin{aligned} \varphi_k &= (\varphi_{k-1} + \omega_k \delta) mod(2\pi) \\ z_k &= z_{k-1} - \sum_{n \in \{P,Q,R,S,T\}} \delta \frac{\alpha_n \omega_k}{b_n^2} \Delta\varphi_n \exp(-\frac{\Delta\varphi_n^2}{2b_n^2}) + \eta_k \end{aligned} \tag{4.3}$$

where $\Delta\varphi_n = (\varphi_{k-1} - \theta_n)$, φ_k is the phase of ECG signal, ω_k is angular velocity (proportional to the inverse of the RR-interval), z_k is the ECG signal considered as the sum of some Gaussian functions and η_k is a random additive white zero-mean noise that models the baseline wander. δ is the sampling period and a_n, b_n and θ_n are the parameters of the Gaussian kernels (amplitude, angular spread, and location of the Gaussian functions, respectively). A simple interpretation of this model is to map each heartbeat over the time on a unit circle lying on XY plane in a way that the first and last samples of a beat take phase $-\pi$ and π respectively. Accordingly the R peak is expected to be located around phase zero. Changing the parameters of Gaussian kernels results in ECG generation with different morphology. The parameter ω_k tunes the HR of generated ECG and determines velocity of moving on unit circle. The proposed differential equations can be integrated in KF model if we consider appropriate observations and relate them to the state variables. Eq.4.4 shows the observation equation of EKF2 model:

$$\begin{aligned} \Phi_k &= \varphi_k + r_{1k} \\ Z_k &= z_k + r_{2k} \end{aligned} \tag{4.4}$$

where $r_k = [r_{1k}, r_{2k}]$ is the observation noise of the model. Φ_k is the phase observation and Z_k represents the measured ECG signal. Since the state equation of this model is nonlinear, EKF is applied for state estimation. Following the notation of Eq.4.1, the state vector, state noise vector and observation noise vector are defined as:

$$\begin{aligned} x_k &= [\varphi_k, z_k] \\ \gamma_k &= [\alpha_P, \cdots, \alpha_T, b_P, \cdots, b_T, \theta_P, \cdots, \theta_T, \omega_k, \eta_k] \\ y_k &= [\Phi_k, Z_k] \\ r_k &= [r_{1k}, r_{2k}] \end{aligned} \tag{4.5}$$

The main application of EKF2 is denoising and compression of ECG signal [Sayadi and Shamsollahi, 2008].

4.2.3 EKF3 Algorithm

In the aforementioned algorithms, the angular velocity, ω_k, is one of the state noises. However, it can also be considered as a state variable ([Lin et al., 2011]). This is the

main subject of EKF3 algorithm proposed by [Akhbari et al., 2012]. In this algorithm, ω_k is considered as the third state variable and an appropriate corresponding observation is introduced. Due to the small changes of the distance between different waveforms during several consecutive beats, a simple autoregressive (AR) model is used for angular velocity as:

$$\omega_k = \omega_{k-1} + \beta_k \tag{4.6}$$

where β_k is a white zero-mean Gaussian noise with variance, σ_β^2, and represents the uncertainty in angular velocity estimation. In this model, angular velocity observation, Ω_k, is obtained using RR signal and it is assumed to be almost constant during each beat but may be contaminated by noise. Hence the corresponded observation equation may be considered as $\Omega_k = \omega_k + r_{3k}$.

4.3 SKF

According to Eq.4.2, a two-class classification problem (or an event detection) can be stated in the form of a SKF model and the estimation of Markovian discrete state practically performs the classification. A SKF algorithm includes two main issues. 1-Training: estimation of the model parameters from labeled training data, 2-Inference: estimation of state variables (Kalman continues states and Markov discrete state) from observations. As discussed, the classification problem is solved by monitoring the discrete state variable. In the following parts, in order to be self-contained, we review the inference problem for general structure of SKF and parameter re-estimation of two proposed SKF structures in ([Murphy, 1998] and [Wu et al., 2004]) for AB detection.

4.3.1 State Estimation (Inference)

The inference problem of KF deals with finding aposteriori mean $\hat{x}_k = E\{x_k|y_{1:k}\}$ that minimizes the mean square error $E\{(x_k - \hat{x}_k)^2|y_{1:k}\}$. For SKF, this can be simply achieved by taking into account the status of the switch. This statement can be expanded as follows:

$$\begin{aligned} p(x_k|y_{1:k}) &= \sum_{i=1}^{N} p(x_k, s_k = i|y_{1:k}) = \sum_{i=1}^{N} K_k^i p(x_k|y_{1:k}, s_k = i) \\ &= \sum_{i=1}^{N} K_k^i \sum_{j=1}^{N} g_k^{j|i} p(x_k|y_{1:k}, s_k = i, s_{k-1} = j) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} K_k^{ji} p(x_k|y_{1:k}, s_k = i, s_{k-1} = j) \end{aligned} \tag{4.7}$$

where $K_k^i = P(s_k = i|y_{1:k})$, $g_k^{j|i} = P(s_{k-1} = j|y_{1:k}, s_k = i)$ and $K_k^{ji} = K_k^i g_k^{j|i} = P(s_k = i, s_{k-1} = j|y_{1:k})$ for $i, j = 1, \cdots, N$. Knowing that the system is in dynamics j and i at time $k-1$ and k respectively and assuming that all the noises and the initial values are Gaussian random values, $p(x_k|y_{1:k}, s_k = i, s_{k-1} = j)$ is a Gaussian function, according

to standard KF assumption. Hence we have:

$$p(x_k|y_{1:k}, s_k = i, s_{k-1} = j) = \mathcal{N}(x_k^{ji}, P_k^{ji}) \tag{4.8}$$

where $x_k^{ji} = E\{x_k|y_{1:k}, s_k = i, s_{k-1} = j\}$ and $P_k^{ji} = Cov\{x_k|y_{1:k}, s_k = i, s_{k-1} = j\}$ and the standard KF algorithm can be utilized for calculation of its mean and covariance. According to Eq.(4.8) and the last expression of Eq.(4.7), $p(x_k|y_{1:k})$ is calculated by the mixture of N^2 Gaussians. So, in order to calculate $p(x_{k+1}|y_{1:k+1})$, N^3 Gaussians should be considered. Hence the number of Gaussians increases exponentially by the time. The suggested solution in previously proposed SKFs is to approximate N^2 Gaussians by N Gaussians at each step by considering $p(x_k|y_{1:k}, s_k = i) \simeq \mathcal{N}(x_k^i, P_k^i)$, where $x_k^i = E\{x_k|y_{1:k}, s_k = i\}$ and $P_k^i = Cov\{x_k|y_{1:k}, s_k = i\}$. Finally, the desired mixture approximation for the aposteriori probability of the states are achieved as follows:

$$p(x_k|y_{1:k}) = \sum_{i=1}^{N} K_k^i \mathcal{N}(x_k^i, P_k^i) \tag{4.9}$$

Another issue is to calculate the weights of Gaussian mixture. It is straightforward to show K_k^{ji} can be given as follows:

$$K_k^{ji} = \frac{L_k^{ji} c_{ji} K_{k-1}^j}{\sum_{i=1}^{N} \sum_{j=1}^{N} L_k^{ji} c_{ji} K_{k-1}^j} \tag{4.10}$$

where $L_k^{ji} = p(y_k|y_{1:k-1}, s_k = i, s_{k-1} = j)$, the likelihood of observation given states at time k and $k-1$, is calculated within the KF algorithm execution.

Therefore, the SKF algorithm based on the given interpretation can be summarized in three stages defined properly in appendix as follows:

1-$[x_k^{ji}, P_k^{ji}, L_k^{ji}] = FilteringKF(x_{k-1}^j, P_{k-1}^j, A_i, \Gamma_i, M_i, R_i)$

2-$[K_k^{ji}, K_k^i, g_k^{j|i}] = StatesProbability(L_k^{ji}, c_{ji}, K_{k-1}^j)$

3-$[x_k^i, P_k^i] = Collapsing(x_k^{ji}, P_k^{ji}, K_k^i, g_k^{j|i})$

where $..._i = ...(s_k = i)$ and the switching discrete state s is assumed to have a first order Markov chain with matrix transition $c_{ji} = P(s_k = i|s_{k-1} = j)$. First stage is dedicated to perform the filtering based on standard KF algorithm. For evolution of KF from time $k-1$ to k, this three-step algorithm is executed N^2 times in order to assess the probability of all possible transitions from i to j and calculate K_k^i. Note that K_k^i is a parameter whose monitoring gives us the result of inference of the switch variable. During the collapsing step, N^2 Gaussian pdfs are approximated by N Gaussians.

4.3.2 Parameters Re-estimation (training)

In re-estimation, the parameters of the models are obtained during training phase. SKFs parameter re-estimation can be set up in various forms according to the impact of the switch on the state equation or observation equation or both. In the following parts, we review the training procedure of two different SKF structures. These structures are interesting for our proposed methods for early detection of AB.

4.3.2.1 SKF I

Let us consider a SKF whose the switch affects only on state equation and determines which linear model should be used for state evolution at each time instant. Hence, the model equations are given as follows:

$$\begin{aligned} x_k &= A(s)x_{k-1} + \gamma_k(s) \\ y_k &= Mx_k + r_k \end{aligned} \tag{4.11}$$

In practice, the parameters of the matrix of states (A_i), covariance matrix of corresponding process noise (Γ_i) and the transition matrix of the switch (C) are required to be trained from the training data for executing the inference algorithm. In order to estimate the unknown parameters of this SKF structure, first we consider the joint probability of states $x_{1:T}$, observations $y_{1:T}$ and switch $s_{1:T}$:

$$p(\{x_{1:T}, y_{1:T}, s_{1:T}\}) = p(s_0)p(x_0|s_0)\prod_{k=1}^{T} p(s_k|s_{k-1})p(x_k|x_{k-1}, s_k)p(y_k|x_k) \tag{4.12}$$

Then log-likelihood of the joint probability is given by:

$$\begin{aligned} L = log(p\{x_{1:T}, y_{1:T}, s_{1:T}\}) = -\frac{1}{2}\sum_{k=1}^{T}\Big\{\{(x_k^i - A_i x_{k-1}^j)\Gamma_i^{-1}(x_k^i - A_i x_{k-1}^j)^\intercal + \log|\Gamma_i|\}K_k^i \\ +(y_k - Mx_k^i)R^{-1}(y_k - Mx_k^i)^\intercal + \log|R| + \log c_{s_{k-1}s_k}\Big\} + \log(P(s_0)P(x_0|s_0)) \end{aligned} \tag{4.13}$$

The criterion we maximize is as follows:

$$\begin{aligned} \arg\max_{A_i,\Gamma_i,c_{ji}} p(\{s_{1:T}, y_{1:T}, x_{1,T}\}) &= \arg\max_{A_i,\Gamma_i,c_{ji}} p(s_{1:T})p(x_{1:T}|s_{1:T})p(y_{1:T}|x_{1:T}) \\ &= \arg\max_{A_i,\Gamma_i,c_{ji}} p(s_{1:T})p(x_{1:T}|s_{1:T}) \end{aligned} \tag{4.14}$$

The details of the maximization process can be found in (Murphy 1998). We only show closed-form solution for re-estimation as follows:

$$\begin{aligned} A_i &= \frac{\sum_{k=2}^{T} K_k^i x_k^i x_{k-1}^i}{\sum_{k=2}^{T} K_k^i x_{k-1}^i x_{k-1}^i} \\ \Gamma_i &= \frac{\sum_{k=2}^{T} K_k^i x_k^i x_k^i - A_i \sum_{k=2}^{T} K_k^i x_k^i x_{k-1}^i}{\sum_{k=2}^{T} K_k^i} \\ c_{ji} &= \frac{\sum_{k=2}^{T} K_k^{ji}}{\sum_{k=1}^{T-1} K_k^i} \end{aligned} \tag{4.15}$$

Note that the parameter M in observation equation, is known by considering the relation between observation and state variables. R can be set by estimation of noise covariance from the observation.

4.3.2.2 SKF II

In the second SKF, the switch affects only the observation equation:

$$\begin{aligned} x_k &= Ax_{k-1} + \gamma_k \\ y_k &= M(s)x_k + r_k(s) \end{aligned} \tag{4.16}$$

In this structure all the model parameters (A, Γ, M_i, R_i and c_{ji}) can be estimated optimally. Considering the structure of the SKF, the optimization criterium based on the joint probability of states, switch and observation is as (4.17) ([Wu et al., 2004]):

$$\arg \max_{A,\Gamma,M_i,R_i,c_{ji}} p(\{x_{1:T}, s_{1:T}, y_{1:T}\}) = \arg \max_{A,\Gamma} p(x_{1:T}) \arg \max_{M_i,R_i,c_{ji}} p(s_{1:T}, y_{1:T}|x_{1:T}) \tag{4.17}$$

First term is related to states which are continuous variables and its maximization is straightforward by EM algorithm.

$$L_1 = -\frac{1}{2}\sum_{k=1}^{T}\{(x_k - Ax_{k-1})\Gamma^{-1}(x_k - Ax_{k-1})^\intercal + \log|\Gamma|\} \tag{4.18}$$

Taking the derivation of L_1 based on the parameters A and Γ, we have:

$$\begin{aligned} A &= (\sum_{k=2}^{T} x_k x_k^\intercal)(\sum_{k=2}^{T} x_{k-1}x_{k-1}^\intercal)^{-1} \\ \Gamma &= \frac{1}{T-1}\sum_{k=2}^{T} x_k x_k^\intercal - A\sum_{k=2}^{T} x_{k-1}x_k^\intercal \end{aligned} \tag{4.19}$$

Using second term, we can estimate parameters of observation equation and transition matrix of the switch which is a discrete variable. Therefore Baum-Welch algorithm should be applied to estimate the transition matrix of the switch. This algorithm uses expectation of the log likelihood for the discrete parameters estimation:

$$L_2 = -\frac{1}{2}\sum_{k=1}^{T}\left\{(y_k - M_i x_k)R_i^{-1}(y_k - M_i x_k)^\intercal + \log|R_i| + \log c_{s_{k-1}s_k}\right\}K_k^i \tag{4.20}$$

So we have

$$\begin{aligned} M_i &= \frac{\sum_{k=2}^{T} K_k^i y_k x_k^\intercal}{\sum_{k=2}^{T} K_k^i x_k x_k^\intercal} \\ R_i &= \frac{\sum_{k=2}^{T} K_k^i (y_k y_k^\intercal - M_i x_k y_k)}{\sum_{k=2}^{T} K_k^i} \\ c_{ji} &= \frac{\sum_{k=2}^{T} K_k^{ji}}{\sum_{k=1}^{T-1} K_k^i} \end{aligned} \tag{4.21}$$

4.4 Proposed Methodologies for AB detection

4.4.1 Waveform based SEKF model

In order to detect AB from normal beats in ECG, some modifications can be applied on EKF3 algorithm to incorporate dynamical variations. In wave-based approach, we try to employ inter-beat as well as intra-beat characteristics. This can be achieved using a dynamical model for the waveforms of a beat. Inspired by existing studies, the famous dynamical model of ECG beats suggested by McSharry is used. Hence, for normal beats and those associated with bradycardia, two dynamic models are considered. At this stage, we assumed that bradycardia complication does not change the morphology of the waveforms; hence the properties of parameters of Gaussian kernels (α_i, b_i and θ_i) which are considered as state noise variables during EKF2 and EKF3 algorithms are constant. The only parameter that can be affected by bradycardia is the temporal distance between the peaks of these waveforms. To characterize such phenomenon, we take the main idea of EKF3 algorithm which gives a continuous state variable to angular velocity (ω_k) which represents the speed of beat generation on the unit circle in McSharry's model. Fig 4.1 depicts two segments of synthesized ECG generated based on McSharry's model with various angular velocity. The parameters of the Gaussian kernels are kept equal in these traces. As can be seen, the radial trajectories are equal, whereas the temporal expansion of beats are infected by their angular velocity.

If ω_k is low, the generated beats elongate in time samples and vice versa. This parameter was mainly applied as a noise with constant properties during a beat generation [Sayadi et al., 2010]. However, using EKF3 idea, we relax this parameter to be assigned at each instant of algorithm evolution. The state variable representing ω_k is added to the other state variables in EKF2 to form EKF3 state equations. Consequently, any subtle changes in ω_k can be detected. In our proposed approach, a two-value switch s is set on the state equation corresponding to ω_k, in order to change it whenever the dynamic changes and to judge between normal and bradycardia conditions. This switch indicates the most probable one which the observation may be generated by. s is a discrete state and assumed to have a first order Markov chain with matrix transition C with $c_{ji} = P(s_k = i|s_{k-1} = j)$. Note that $..._i = ...(s_k = i)$ can be considered for all the parameters of SKF model (A, M, Γ and R) which the switch is allocated on them. The modified EKF3 is given as follows:

$$\begin{aligned} \omega_k &= A(s)\omega_{k-1} + \beta_k(s) \\ \varphi_k &= (\varphi_{k-1} + \omega_{k-1}\delta) mod(2\pi) \\ z_k &= z_{k-1} - \sum_{n\in\{P,Q,R,S,T\}} \delta \frac{\alpha_n \omega_{k-1}}{b_n^2} \Delta\varphi_n \exp(-\frac{\Delta\varphi_n^2}{2b_n^2}) + \eta_k \end{aligned} \tag{4.22}$$

The observation can be related to the states of Eq.4.22 as the following equations:

$$\begin{aligned} \Omega_k &= \omega_k + r_{1k} \\ \Phi_k &= \varphi_k + r_{2k} \\ Z_k &= z_k + r_{3k} \end{aligned} \tag{4.23}$$

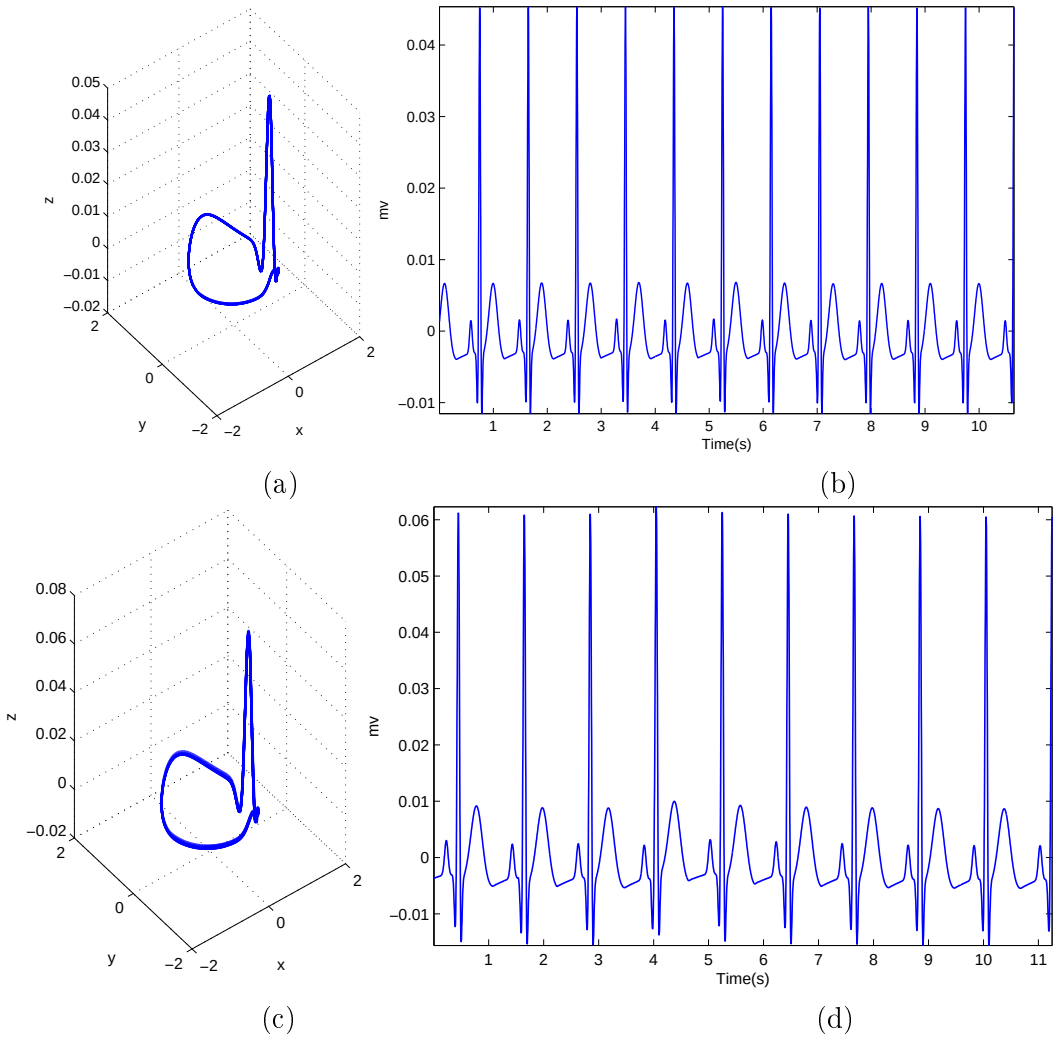

Figure 4.1: A benchmark on the effect of angular velocity in beat generation of ECG using McSharry's model. (a,c): the trajectory of beats around unite circle. (b,d): temporal expansion of trajectories in (a) and (c). (b): Generated beats with angular velocity of 5 rad/s. (d): Generated beats with angular velocity of 7 rad/s.

where $r_k = [r_{1k}, r_{2k}, r_{3k}]$ is the measurement noise. Moreover, we define a physiological envelope on the amplitude of the angular velocity corresponding to normal condition (i.e. $\omega_k(s = 1)$) that spans between the upper and lower ranges of Ω_{env} defined as:

$$\Omega_{env} = \overline{\Omega}_{normal} \pm 3\sigma_{\Omega_{normal}} \tag{4.24}$$

where $\overline{\Omega}_{normal}$ is the mean of the observation of angular velocity related to normal ECG and its standard deviation is $\sigma_{\Omega_{normal}}$. Ω_{env} is set by the training data. This constraint is defined since, practically, the value of $\omega_k(s = 1)$ is nearly constant in normal beats. In case of bradycardia, the value of angular velocity decreased due to beats elongation. So $\omega_k(s = 1)$ is set to the nearest endpoint. Similarly, in normal circumstances, the value of the state of angular velocity related to AB, $\omega_k(s = 2)$, is equal to the mean of Ω_k during the intervals that an AB is observed in training data. As mentioned, due to nonlinearity of the third equation in Eq.4.22, our proposed model is called as a SEKF.

Following the notation of Eq.4.1, the state, observation and their corresponding noise vectors are defined as:

$$\begin{aligned} x_k &= [\omega_k\ \varphi_k\ z_k] \\ \gamma_k &= [\alpha_n\ b_n\ \theta_n\ \beta_k\ \eta_k] \quad n \in \{P, Q, R, S, T\} \\ y_k &= [\Omega_k\ \Phi_k\ Z_k] \\ r_k &= [r_{1k}\ r_{2k}\ r_{3k}] \end{aligned} \tag{4.25}$$

where Γ, the covariance matrix of states γ_k which is a 17-dimensional matrix as we have considered three parameters for five Gaussian kernels in addition to two additive noises.

Since the second process equation is nonlinear, a linear approximation of Eq.4.22 near a desired point $(\hat{x}_k, \bar{\gamma}_k)$ is required for estimating the state vector according to EKF algorithm. Linearization can be performed as follows:

$$x_k = f(x_{k-1}, \gamma_k, k) = f(\hat{x}_{k-1}, \bar{\gamma}_k, k) + G_k(x_{k-1} - \hat{x}_{k-1}) + F_k(\gamma_k - \bar{\gamma}_k) \tag{4.26}$$

where f represents the nonlinear state evolution function. G_k and F_k are state linearization coefficients which are given by:

$$\begin{aligned} G_k &= \frac{\partial f(x_{k-1}, \bar{\gamma}_k, k)}{\partial x_{k-1}}|_{x_{k-1}=\hat{x}_{k-1}, s_k=i, s_{k-1}=j} \\ F_k &= \frac{\partial f(\hat{x}_{k-1}, \gamma_k, k)}{\partial \gamma_k}|_{\gamma_k=\bar{\gamma}_k, s_k=i, s_{k-1}=j} \end{aligned} \tag{4.27}$$

According to Eq.4.22 and Eq.4.27, we have:

$$\begin{aligned} &G_k(1,1) = A_i \qquad G_k(1,2) = G_k(1,3) = 0 \\ &G_k(2,1) = \delta \qquad G_k(2,2) = 1 \qquad G_k(2,3) = 0 \\ &G_k(3,1) = -\sum \delta \frac{\alpha_n}{(b_n)^2} \Delta\theta_n \exp(-\frac{\Delta(\theta_n)^2}{2(b_n)^2}) \\ &G_k(3,2) = -\delta \frac{\alpha_n \omega_{k-1}^j}{b_n^2}(1 - \frac{(\Delta\theta_n)^2}{b_n^2}) \exp(-\frac{\Delta\theta_n^2}{2b_n^2}) \qquad G_k(3,3) = 1 \end{aligned} \tag{4.28}$$

$$\begin{aligned} &F_k(1,1) = 1 \qquad F_k(1,2:17) = 0 \qquad F_k(2,1;17) = 0 \qquad F_k(3,1) = 0 \\ &F_k(3,2:6) = -\delta \frac{\omega_{k-1}^j \Delta\theta_n}{b_n^2} \exp(-\frac{\Delta\theta_n^2}{2b_n^2}) \\ &F_k(3,7:11) = 2\delta \frac{\alpha_n \omega_{k-1}^j \Delta\theta_n}{b_n^3}(1 - \frac{\Delta\theta_n^2}{2b_n^2}) \exp(-\frac{-\Delta\theta_n^2}{2b_n^2}) \\ &F_k(3,12:16) = \delta \frac{\alpha_n \omega_{k-1}^j}{b_n^2}(1 - \frac{\Delta\theta_n^2}{b_n^2}) \exp(-\frac{\Delta\theta_n^2}{2b_n^2}) \\ &F_k(3,17) = 1 \qquad M = I_3 \end{aligned} \tag{4.29}$$

where I_3 is the 3-dimension identity matrix. Hence, the SEKF approach is comprised of three stages : *FilteringSEKF*, *StatesProbability* and *Collapsing* (see appendices B), the same algorithms as in 4.2 and 4.3 are used for training and inference procedures.

Algorithm 4.2: Wave based: Training

Inputs: x_0^{ji}, P_0^{ji}, initial values of A_i, Γ_i, M, R and C, $i, j = 1, 2$.
Outputs: Trained values of A_i, Γ_i and C

repeat
  for $k = 1, 2, \cdots, T$ **do**
    Compute G_k and F_k from Eq. 4.27.
    $[x_k^{ji}, P_k^{ji}, L_k^{ji}]$ =FilteringSEKF$(x_{k-1}^{j}, P_{k-1}^{j}, A_i, \Gamma_i, M, R, G_k, F_k)$
    $[K_k^{ji}, K_k^{i}, g_k^{j|i}]$ =StatesProbability$(L_k^{ji}, c_{ji}, K_{k-1}^{j})$
    $[x_k^{i}, P_k^{i}]$ =Collapsing$(x_k^{ji}, P_k^{ji}, K_k^{i}, g_k^{j|i})$
  end for
  Calculate A_i, Γ_i and C using Eq. 4.15.
until Convergence

Algorithm 4.3: Wave based: Inference

Inputs: x_0^{ji}, P_0^{ji}, A_i, Γ_i, M, R and C, $i, j = 1, 2$.
Outputs: K_k^i,

for $i = 1, 2, \cdots, M$ **do**
  for $k = 1, 2, \cdots, T$ **do**
    Compute G_k and F_k from Eq. 4.27.
    $[x_k^{ji}, P_k^{ji}, L_k^{ji}]$ =FilteringSEKF$(x_{k-1}^{j}, P_{k-1}^{j}, A_i, \Gamma_i, M, R, G_k, F_k)$
    $[K_k^{ji}, K_k^{i}, g_k^{j|i}]$ =StatesProbability$(L_k^{ji}, c_{ji}, K_{k-1}^{j})$
    $[x_k^{i}, P_k^{i}]$ =Collapsing$(x_k^{ji}, P_k^{ji}, K_k^{i}, g_k^{j|i})$
  end for
end for
Applying a threshold on K_k^i

Initial values of our proposed SEKF model are set similarly to ([Sameni et al., 2007]) while the unknown parameters of switching dynamics (i.e. A_i, Γ_i and C) are updated iteratively in the re-estimation procedure. Therefore, the function *FilteringSEKF* (cf appendix B.1), is performed followed by *StatesProbability* and *Collapsing* functions over training data. Then using Eq. 4.15, we obtain the parameters of SEKF. Note that the parameter M in observation equation, is known by considering the relation between observation and state variables. R can be set by estimation of noise covariance from the observation. The re-estimation repeats itself till convergence.

After the training procedure to obtain the parameters of each model, we used the inference procedure over each test signal to achieve the probability of the dynamic i for each time instants $(K_k^i, i = 1, 2)$. If $K_k^1 > K_k^2$, the observation at time k is likely generated by dynamic 1. Moreover, since bradycardia is an event happening in a beat not in a time instant, it would be more accurate to relate each beat to a dynamic. Therefore, if more than $\xi\%$ of the temporal samples of an ECG beat are related to the

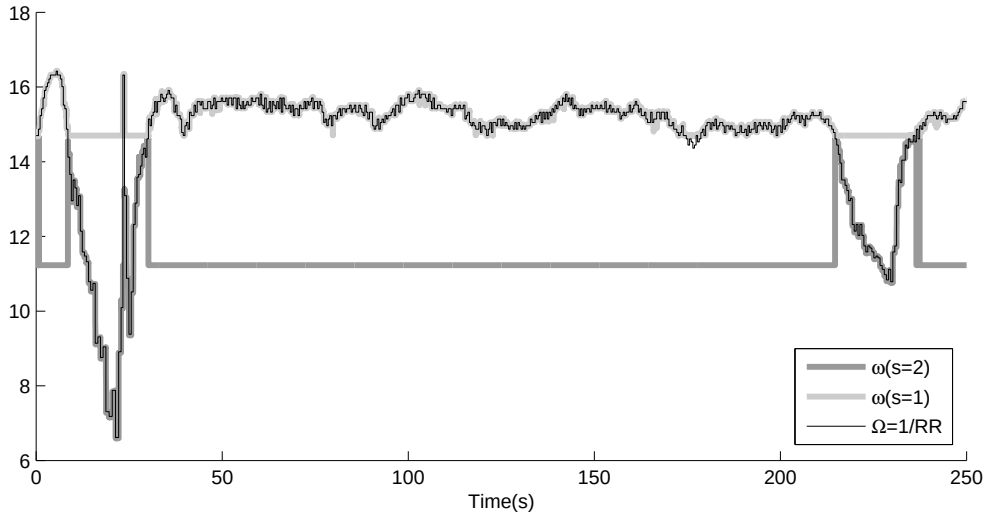

Figure 4.4: Estimated angular velocity (state variable controlled by switch) and the observation of the angular velocity for wave-based method.

dynamic of bradycardia, it is supposed that the AB begins from the first sample of that beat. The estimated state variable corresponding to angular velocity of wave-based model is illustrated in Fig.6.1. Physiological envelope (Eq. 4.24) restricts the range of this state variable value (ω_k^i).

Physiologically speaking, the elongation of a beat does not occur only at the end of the beat and it is accompanied by increase of temporal distances between different waveforms. Hence, we try to give an intra beat interpretation to angular velocity using dynamical model to detect the changes not only in the R-R distances but also in temporal distance between the waveforms peaks in a beat. We also investigated two methods in order to compute local mode on the label of consecutive samples.

In the first method, mode on window (MOW), we calculated the mode of labels fall in a window moving through the signal as the SEKF algorithm is applied on a sample basis. The length of the moving window is set to the average length of ECG beats in training data set. If more than ξ% of the temporal samples of an ECG beat are labeled as AB, it is supposed that the AB begins from the first sample of that beat. ξ can be in range of 0 t0 100%. Whereas in the second method, mode on beat (MOB), we can consider the label of samples of a beat and calculate their mode. Note that in MOB method, the label of the samples in a beat located in the beginning of AB event is determined at the end of the beat. Therefore, even with all samples labeled as AB, the detection delay of an AB event is one heart beat. There is an ambiguity in the borders of a beat, so it is required to choose a logic for defining the beginning and ending points. We have used Sameni's definition [Sameni et al., 2007] where the beginning of a beat is considered on a point placed in a equal distance between the corresponding R peak and the previous one.

4.4.2 R-based SKF model

The alternative and simple method that we propose for AB detection, consists of a SKF structure where the switch affects just the observation equation. In this structure, the observation includes the RR signals obtained from raw ECG. The normal RR signals are modeled by an AR process. A bradycardia episode commits a rise in RR signal and can also be independently fitted by an AR model which differs from the AR model corresponding to normal HR. Normal and bradycardia AR models whose parameters are obtained during training phase configure the state equation of this SKF. Using higher order AR models would be more accurate, but with more complexity. Therefore, we restrict our study to the first order AR models. Each of the first order AR model is equivalent to a 1-D SSM. The states of these two models evolve independently and estimate the amplitude of RR based on their given dynamic. Hence, x_{1_k} and x_{2_k} are RR signals corresponding to normal and AB respectively resulting from two independent AR models. They generate various K_k^1 and K_k^2 which are used to classify the samples of RR signal into one of the normal and bradycardia classes. It can be concluded that the approach employs just the inter beat information since it employs RR signal as the only observation. The state and observation equations are given as:

$$\begin{aligned} x_k &= [x_{1_k} \quad x_{2_k}]^\intercal = A_{2\times 2}[x_{1_{k-1}} \quad x_{2_{k-1}}]^\intercal + \gamma_k \\ y_k &= M(s)[x_{1_k} \quad x_{2_k}]^\intercal + r_k(s) \end{aligned} \tag{4.30}$$

where y_k is the calculated RR from recorded ECG that can be considered as a noisy version of x_{1_k} or x_{2_k}. A is diagonal matrix. In this structure all the model parameters (A, Γ, M_i, R_i and c_{ji}) can be estimated optimally. If we consider the structure of the SKF, the optimization criterion based on the joint probability of states, switch and observation is as Eq. 4.17 which is proposed by [Wu et al., 2004]. We perform the training of parameters A, Γ, M_i, R_i and c_{ji} using Eq.4.19 and 4.21.

The training and inference of this method are performed according to algorithms 4.5 and 4.6 using *FilteringKF* which is defined in appendices and includes the standard KF algorithm.

Algorithm 4.5: R-based: Training

Inputs: x_0^{ji}, P_0^{ji}, initial values of A, Γ, M_i, R_i and C, $i, j = 1, 2$.
Outputs: Trained values of A, Γ, M_i, R_i and C

repeat
  for $k = 1, 2, \cdots, T$ **do**
    $[x_k^{ji}, P_k^{ji}, L_k^{ji}]$ =FilteringKF$(x_{k-1}^j, P_{k-1}^j, A, \Gamma, M_i, R_i)$
    $[K_k^{ji}, K_k^i, g_k^{j|i}]$ =StatesProbability$(L_k^{ji}, c_{ji}, K_{k-1}^j)$
    $[x_k^i, P_k^i]$ =Collapsing$(x_k^{ji}, P_k^{ji}, K_k^i, g_k^{j|i})$
  end for
  Calculate A and Γ using Eq. 4.19 and M_i, R_i and C using Eq. 4.21.
until Convergence

Algorithm 4.6: R-based: Inference

Inputs: x_0^{ji}, P_0^{ji}, A, Γ, M_i, R_i and C, $i, j = 1, 2$.
Outputs: K_k^i,

1: **for** $i = 1, 2, \cdots, M$ **do**
2: **for** $k = 1, 2, \cdots, T$ **do**
3: $[x_k^{ji}, P_k^{ji}, L_k^{ji}]$ =FilteringKF$(x_{k-1}^j, P_{k-1}^j, A, \Gamma, M_i, R_i)$
4: $[K_k^{ji}, K_k^i, g_k^{j|i}]$ =StatesProbability$(L_k^{ji}, c_{ji}, K_{k-1}^j)$
5: $[x_k^i, P_k^i]$ =Collapsing$(x_k^{ji}, P_k^{ji}, K_k^i, g_k^{j|i})$
6: **end for**
7: **end for**
8: Applying a threshold on K_k^i

4.5 Conclusion

In this chapter, the algorithms which are previously proposed in literature and can be described based on dynamical and EKF modeling of ECG beat generation system, are reviewed. Then, the SKF inference is remarked followed by considering two different switching structure and their relevant parameters estimation in training phase. Finally, having the vital tools for construction of the novel methods to detect AB events, we introduce a wave-based and R-based methods. The former, according to its name, directly processes the morphology of the wave-forms and defines a state variable for the term angular velocity. The second method, fits readily two different AR models to the normal and bradycardia segments and evaluate the probability of each class for the inference.

Chapter 5

Datasets and Evaluation Metrics

5.1 Introduction

According to the previous chapters, we can divide the proposed methods in this thesis into Markovian and Switching approaches where the main interest is to suggest methods which take into account the dynamic of time series in the detection of pathophysiological events. For evaluation of the methods in both approaches, a real database consisting of raw ECG and three feature signals extracted directly from the ECGs is used. It includes AB episodes in preterm neonates. These features were previously exploited from the raw ECG in past studies conducted in our laboratory whose complete description can be found in [Altuve, 2011]. The database is used for on-line detection of AB episodes; however in various configuration in Markovian and Switching approaches. A part of database (Data 1) which includes just the feature signals are applied in Markovian approach, while their corresponding raw ECG segments together with one of the feature signals (Data 2) are employed to evaluate the Switching methods. Moreover, in Markovian approach, FitzHugh-Nagumo model [FitzHugh, 1961] is used to generate simulated bivariate signals. This model is particularly suited to our problem because its state variables have heterogeneous dynamics which can be altered by adjusting a parameter. For the simulated database, two studies have been performed and evaluated: i) three-channel classification of segments of multivariate time series including two classes corresponding to different dynamics of disturbance and a class for rest condition, ii) the on-line detection of disturbances. The methods which are proposed in Switching approach are specified by the physiological characteristics and present algorithms to solve directly the AB detection issue. Hence, their evaluation using simulated data gives no information. In the following parts, the simulated and real databases are detailed. Then, the evaluation procedure of Markovian and Switching methods are described.

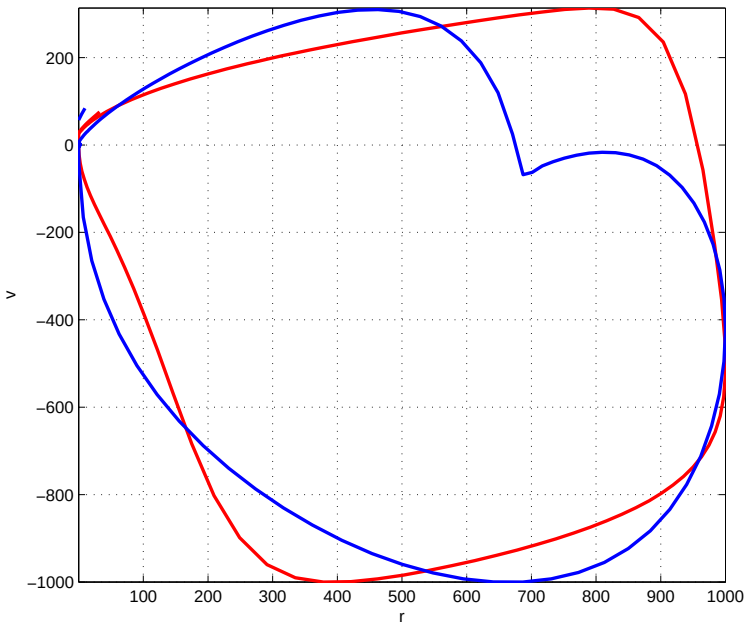

Figure 5.1: v (blue line) versus r (red line) trace.

5.2 Markovian Approach

5.2.1 Simulated database

The simulated data is generated with the FitzHugh-Nagumo model, defined by the following differential equations:

$$\begin{aligned} \frac{dv}{dt} &= 3(v - \frac{1}{3}v^3 + r + I) \\ \frac{dr}{dt} &= -\frac{1}{3}(v - a + 0.8r) \end{aligned} \tag{5.1}$$

where the variables r and v are set to their "rest" values (fixed point) and disturbances are injected to the system by changing the value of I from 0 to 1. Fig. 5.2 shows different dimensions of simulated data (v versus r). The dynamics of the system depend on the value of parameter a, which is assumed in this test to be a random variable with a uniform probability density function $a_1 \sim \mathcal{U}(0.58, 0.62)$ for the first class and $a_2 \sim \mathcal{U}(0.78, 0.82)$ for the second class. The impact of changing parameter a in the simulated data is illustrated in Fig. 5.2. The resemblance between the time series in class 1 and class 2 makes the classification a difficult issue. An appropriate classifier for such problem needs to correctly differentiate the dynamics of these time series and not only their instantaneous amplitudes. In order to evaluate methods of Markovian approach, 200 sequences of 400s duration at a sampling frequency of 10 Hz are generated with a disturbance introduced during 300-305s (Fig. 5.2(a)). 40 segments of the interval $300 - 310$s are used for training (cf Fig. 5.2(b)) and the rest of them for evaluation.

Concerning on-line detection, train data is similar to that used for the classification case. However, the test data contain the whole 400s-duration signals and are processed

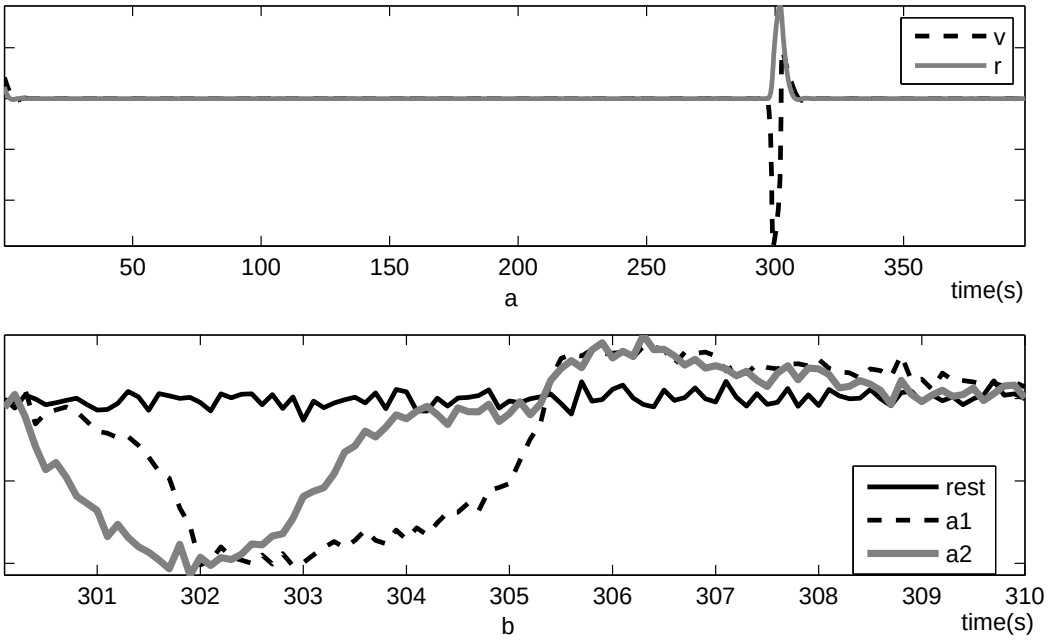

Figure 5.2: a: Simulated signals obtained from the FitzHugh-Nagumo model (state variables v and r) with a disturbance applied in 300-305(s) without noise. b: focusing on v dimension with 5 db additive noise, in the rest condition and during activations with different dynamics (a_1 and a_2). Note that different dynamics are obtained when using parameters a_1 and a_2, while the amplitude of the responses is similar.

with 10 s sample by sample sliding window. Finally, a white Gaussian noise was added to achieve SNR value of 5 dB to all the database.

5.2.2 Real database: Data 1

For real data analysis, we use a database consisting of 236, 1-lead ECG segments acquired from 32 preterm infants, hospitalized in the NICU at the University Hospital of Rennes, France [Altuve et al., 2011a]. ECG signals were acquired at 400Hz. This observational study was approved by the ethical committee of the University Hospital of Rennes and a written consent was obtained from the parents of each infant. All data were anonymized at acquisition. As a summary, the algorithm for data preparation comprises: i) preprocessing of raw ECG for discarding of base line drift and effect of 50/60 Hz of electricity. ii) detecting the ECG beats from the raw ECG signals using Pan and Tompkins method [Pan and Tompkins, 1985], iii) extracting beats: each beat was detected and extracted from the ECG in a defined time window around the QRS complex. It was then readjusted to beat prototypes that represent different types of morphologies. Prototype beats were created by average of the latest beats detected in 10 seconds ago. Only beats which had a normalized cross-correlation greater than 0.96 were used to update prototype beats. iv) segmenting each detected beat by a wavelet-based algorithm and v) constructing the RR time series (successive differences R waves), RAMP (differences between the peak of the R wave and the isoelectric level)

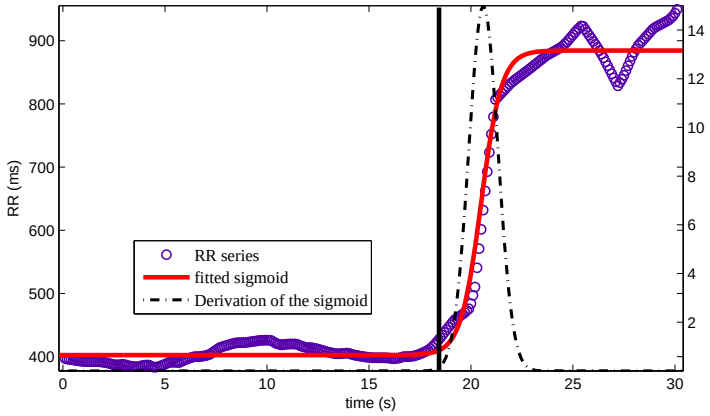

Figure 5.3: Approximation of the RR time series (shown by circles) for an episode of bradycardia, a sigmoid function (red continuous curve). The black dotted curve represents the derivation of the sigmoid function and the vertical line corresponds to annotation of the bradycardia onset.

and QRSd (the difference between the end and the beginning of the QRS complex) for each subject in the database [Altuve, 2011]. All the needed parameters for preprocessing, ECG segmentation and time series generation were optimized using evolutionary analysis proposed in [Altuve et al., 2011b, Hernández et al., 2012]. All the obtained time series which we call them as feature signals, were uniformly oversampled at the frequency of 10 Hz, using well-known interpolation techniques to get sufficient temporal resolution for the application of Markovian models. In addition, the amplitude of RAMP series were reduced by dividing on the maximum absolute value found in the first two minutes.

In this study, the onset of bradycardia were annotated manually by a clinician, but before that, a curve fitting procedure has been proposed to locate, with the best reproducibility, the beginning of the bradycardia. A sigmoid function in Eq. 5.2 is used to approximate the RR time series around the onset of bradycardia:

$$f(t) = A + \frac{B}{1 - exp\frac{t-C}{D}} \tag{5.2}$$

The beginning of the bradycardia is the point where the first derivative of the sigmoid function is greater than 1. An example of approximating the RR time series is shown Fig. 5.3. A total of 148 sets of time series are integrated in the database from 32 preterm infants (each consisting of the RR series, Ramp and QRSd). This database has 233 episodes of bradycardia which have been annotated with 21.48 ± 16.07 seconds duration.

Fig. 5.4 depicts the change of these features before, during and after an apnea-bradycardia episode. Two datasets were constructed for the learning phase, consisting of segments of 7 s duration: LS1: composed of 30 segments taken randomly and beginning with a bradycardia event. LS2: consisting of 30 segments taken randomly from the normal parts of the series (without any apnea-bradycardia event). The length of these

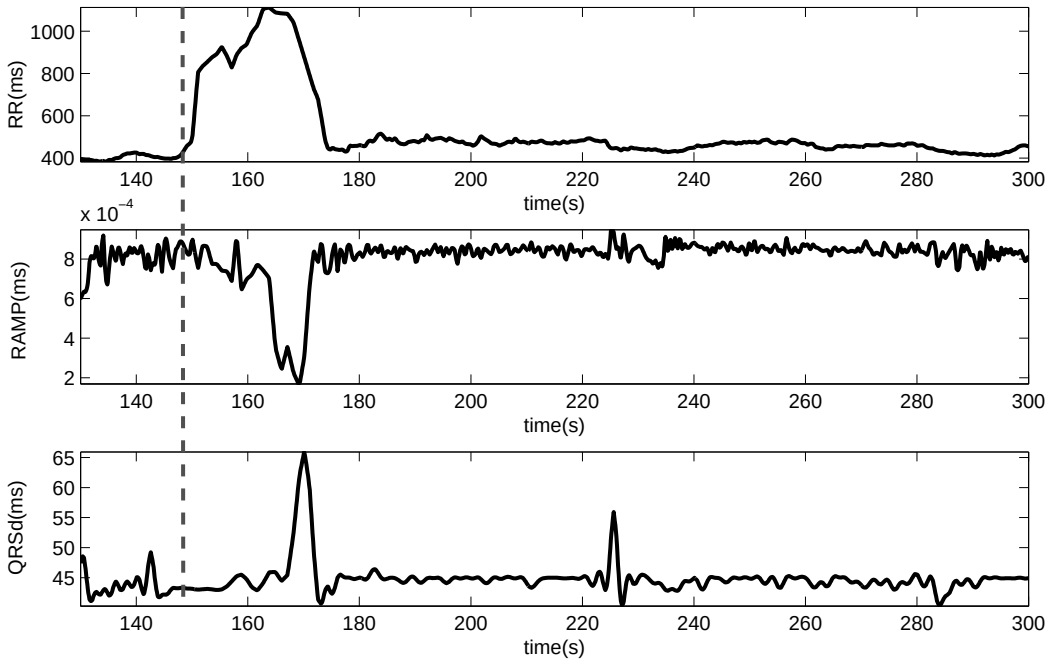

Figure 5.4: Example of time series extracted from test data. The onset of the apnea-bradycardia event is shown by dashed grey line.

segments (7 s) corresponds to the average time measured from the beginning of the bradycardia to the peak RR value within the bradycardia episodes [Altuve et al., 2011a]. After learning procedure, the evaluation of the trained models for detecting the onset of bradycardia is applied with a sliding window of size T = 7 s. The test dataset includes 40 sets of RR-RAMP-QRSd time-series including mostly normal activity and one or more apnea-bradycardia events.

In real data, since the RAMP data is modulated by respiration activities [Haskova et al., 2013b], the effect of apnea will first appear in the RAMP feature. We apply a synchronization time delay (τ), whose possible values are observed around 4.5 s. The best value of τ is investigated and reported in chapter 6.

5.2.3 Evaluation of Markovian approach

We evaluate the performance of a detector using the most common metrics found in the literature. To express how successfully a detector recognizes events without missing them, sensitivity is used. Likewise, specificity measures how exclusively it does not detect a wrong event.

$$SEN = TP/(TP + FN) \tag{5.3}$$

$$SPC = TN/(TN + FP) \tag{5.4}$$

where TP, FP, TN and FN denote the number of true positives, false positives, true negatives and false negatives, respectively which are counted over the samples of observations of test data. False positive rate error can also be defined as $1 - SPC$. The

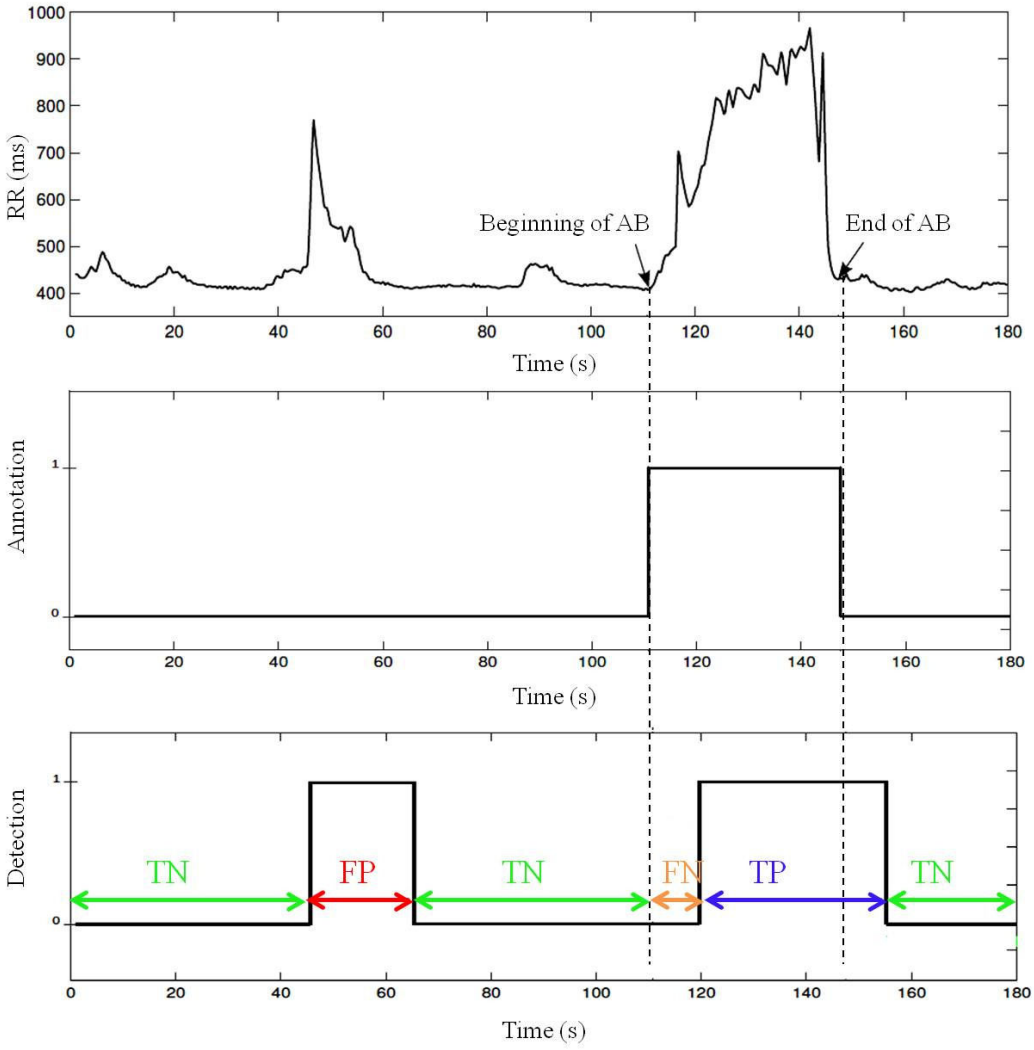

Figure 5.5: Method of claculation of TP, FP, TN and FN in detection scenario in order to

accuracy can also be quantified as:

$$AC = (TP + TN)/(N + P) \tag{5.5}$$

where $N = TN + FN$ and $P = TP + FP$. Moreover, time delay (TD) is another crucial metric for the overall system performance evaluation in the case of on-line detection of a desired event, and is defined as the elapsing time between the detected onset and the annotated onset of the event. The definition of the metrics is illustrated in Fig. 5.5.

In an on-line detection, performance is further analyzed by determining the value of threshold on total likelihoods obtained by Eq. (3.140) and are assessed from a Receiver Operating Characteristic (ROC) curves, based on the values of SEN and SPC. ROC is traced using different threshold values in Eq. (3.140). The point with optimum detection performance and related threshold are chosen using a criterion called perfect detection (PD) defined as:

$$PD = \arg\max\{SEN \times SPC\}. \tag{5.6}$$

Table 5.1: Confusion table.

		Classification		
		class 1	$\cdots$	class K
Reference	class 1	x_{11}	$\cdots$	x_{1K}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	class K	x_{K1}	$\cdots$	x_{KK}

In the cross-validation approach, the ROC curves are obtained by averaging the SENs and SPCs achieved by an identical threshold. Furthermore, three other metrics are also calculated: i- the distance to PD defined as $\sqrt{(1-SEN)^2+(1-SPC)^2}$, ii- AUC of the averaged ROC curve and iii- Positive Windows (PW) which is defined as the ratio of the number of detections occurring after the annotated onset, over the number of detected AB episodes, for all executions. PW illustrates the ability of an algorithm in AB prediction. Lower values of PW (below 50%) is an indication of the detection performance that can be reached by a given detection method, when configured optimally (optimal threshold). Similarly, lower values of this metric indicates a better performance.

The classification performance is evaluated using confusion tables, by calculating the SEN, SPC and AC. A confusion table is constructed by setting the rows for reference labels and columns for labels obtained by the classification algorithm, as shown in table 5.1. The SEN and SPC which show the ability of the method in finding the event of interest and in separating this event from others respectively, can be determined using the Eq.s 5.7 for class k and according to table 5.1.

$$\begin{aligned} SEN &= \frac{x_{kk}}{\sum_{i=1}^{K} x_{ki}} \\ SPC &= \frac{\sum_{i=1,i\neq k}^{K} \sum_{j=1,j\neq k}^{K} x_{ij}}{\sum_{i=1,i\neq k}^{K} \sum_{j=1}^{K} x_{ij}} \end{aligned} \tag{5.7}$$

In a classification task, the accuracy (AC) is defined as Eq. 5.8:

$$AC = \frac{\sum_{i=1}^{K} x_{ii}}{\sum_{i=1}^{K} \sum_{j=1}^{K} x_{ij}} \tag{5.8}$$

In order to cross-validate the results, the procedure of detection is repeated in 5 rounds, each of which involves a hold-out cross-validating and calculation of metrics. In each round, 20% of the data used for training and the rest remains for test data. Each segment includes at least an AB episode; therefore, various number of ABs are used for training in each round.

5.3 Switching approach

5.3.1 Real database: Data 2

As discussed, this data is derived from the same raw ECGs of Data 1 which is used in evaluation of Markovain models; however, to ensure the consistency of the results on various subjects and reducing the number of normal temporal samples, the full-length of the records were divided in to segments with lower duration. As discussed, the wave-based approach requires ECG, phase and angular velocity signals to be observed. Data 1 includes RR, QRSd and RAMP feature signals. Therefore, RR signal is used in order to generate the angular velocity observation using Eq. 5.9:

$$\Omega(t) = \frac{2\pi f}{RR(t)} \tag{5.9}$$

where f is the sampling frequency of the RR signal. Phase observation is constructed according to EKF2 algorithm [Sameni et al., 2007]. Generally, 105 groups of observation consisting of ECG, RR and angular velocity with 250 seconds duration and 400 Hz sampling rate construct the Data 2. This database is used to study the performance of the proposed wave-based method. For R-based approach, we used only the RR feature signal.

5.3.2 Evaluation of Switching approach

For the evaluation of this approach in detection issue, quantitative results are reported using common metrics: SEN and SPC, defined as in 5.2.3. In R-based method, the extracted RR signal is interpolated to reach 400 Hz sampling rate to be matched with annotations. Then, in order to have AR models with more convenient parameters and avoiding the algorithm to be overtrained, tedious and time-consuming, it is downsampled to 10 Hz. The metric, TD, is also defined in this approach to show the speed of an algorithm to detect an AB. It is calculated similarly as the elapsing time between the annotated onset of occurring AB and the onset which the algorithm determines. The wave-based method is expected to find the onset by 2.5 ms time resolution while R-based method has 0.1 s time resolution according to their sampling rate. The accuracy of detection in terms of events (AC_event) is also computed as the ratio of detected AB episodes over the total number of AB episodes.

Similar to the Markovian approaches, we perform cross-validation. All the methods are executed 5 times with different records in the train and test sets. Therefore, the mean and variance over all executions are reported.

5.4 Conclusion

In this chapter, the real database which is acquired from preterm neonates monitored in NICU and suffer from apnea associated with prematurity, was characterized. This data was employed in both Markovian and Switching approaches. In addition, in former approach, a simulated data which had a similar dynamics as AB events, was introduced

in order to study the performance of the proposed frameworks. After describing the data applied in each approach, we demonstrated our strategy for the evaluation issue.

Chapter 6

Experimental Results

6.1 Introduction

Throughout the previous chapters, we have already justified and explained our interest to study the dynamic of time series in the context of a pathophysiological event detection, AB, from the ECG signal recorded from preterm infants. In this chapter, the performance of the proposed algorithms presented in chapters 3 and 4 are studied in two main parts: 1- Markovian and 2- Switching approaches.

In Markovain approach, the methods are evaluated for classification and detection scenarios on simulated and for detection scenario on real signals. In classification, each segment of test data is classified in to one of the classes including dynamics and rest condition. An off-line procedure is used for classification task where the specific time of the events arrival is not important. Detection is performed using a specific sliding procedure which aims to estimate the precise time of events of interest incident. The algorithms are based on analyzing the likelihoods of observation given the models of normal condition and the dynamics. Experiments on simulated data are presented in sections 6.2.1 and 6.2.2 which demonstrate the feasibility of our proposed CHMM and CHSMM and optimize their structure for the analysis of the dynamic variables. The simulated bivariate signals which includes pattern looks like an AB episode are also used for evaluation. Section 6.2.3 presents the application of the proposed frameworks on real multivariate time series. They consists of the features extraction from the ECG of preterm neonates, including RR, RAMP and QRSd, as described in chapter 5 of this manuscript. The evaluation is performed in on-line detection of AB events. In this approach, total likelihood signals are calculated and assessed as indices whose alteration can indicate to AB incidence.

In part two, the performance of Switching approach including R-based and wave-based methods is presented using the same real database which is employed in Markovian methods, however, the raw ECG data and RR feature signals are only used. Qualitative and quantitative results are presented in section 6.3. During the evolution of SKF algorihtm in both structures, K_k^1 and K_k^2 (defined in Eq. 4.10) are computed. They can be considered as indices that by their monitoring during inference, we can receive meaningful information about the performance of each model.

6.2 Evaluation of Markovian approach

In this section, we proceed to evaluate the performance of models proposed in chapter 3 on synthetic signals, and then, examine them over real signals. The procedure for generating synthetic noisy signals based on FitzHugh-Nagumo model is described in section 5.2.1. Bivariate time series, with two slightly different dynamics are produced using this model by modifying one of the parameters. In addition, an amplitude reduction procedure was applied to the synthesized time series in order to make the subtraction of instantaneous values of the likelihood corresponding to each class easy and feasible. The process of classification and detection in this type of signals are particularly difficult since although the dynamics are different, however, their range of amplitude have overlap. The same detectin procedure is applied for AB detection on a real database. In all parts, the performance of the proposed frameworks are compared to HMM and some of the existing generalizations. The performance evaluation of Markovian approach is carried out in four following experiments:

1- Studies on simulated data:

- *3-class classification of simulated data:* The objective here is to evaluate the ability of these models to classify segments of observation based on their dynamics. The results have confirmed the superiority of coupling models.

- *Detection a specific event based on the dynamics of observations:* The interest here is to assess the impact of changes on the values of sensitivity, specificity and time delay of detection, whose simulated signals dynamics are finely modified by changing the model parameters of FitzHugh-Nagumo model.

2- Studies on real data:

- *Detection of AB episodes based on only the dynamics of observations:* The purpose of this experiment is to firstly analyze the impact of different values of synchronization time delay, because as reported in chapter 5.2.2, the impact of apnea appears immediately in the RAMP signal while apnea influences the RR and QRSd features couple of seconds later. Furthermore, in this experiment, we intend to finally illustrate the performance of our proposed frameworks in AB detection in preterm infants comparing to other identical existing models.

6.2.1 Classification on synthetic data

A classification is performed on simulated data, by defining three classes: a_1, a_2 and rest condition whose performance is assessed by contingency tables and calculating the accuracy per class. We further assume that the number of states of the two competing models is equal and is more than that of the model corresponding to rest condition. The SEN, SPC and AC metrics are calculated for various numbers of states to find their optimum number of states. Classification performance is presented in table 6.1, including the results of univariate observation of HMM using just dimension v. All results are obtained with optimal state numbers. According to this table, coupling methods are more efficient in distinguishing the two dynamics from each other compared with HMMs. Although our proposed CHMM apparently shows superior performance in terms of sensitivity and accuracy to separate dynamics related to a_1 and a_2, it fails

Table 6.1: Comparison of the cross-validation results corresponding to Markovian methods for optimal state number for dynamics classification

Method	Class	♯ states	SEN(%)	SPC(%)	AC(%)
	a_1	7	55.63 ± 37.30	91.56 ± 18.69	79.58±10.13
HMM	a_2	7	83.13±37.39	77.50±18.63	79.38± 9.96
(v)	*rest*	3	99.38±0.77	100±0	99.79±0.26
	a_1	3	86.25±5.23	90.63±2.30	89.17±1.59
HMM	a_2	3	81.25±4.59	92.94±2.52	89.04±1.55
(v/r)	*rest*	2	99.63±0.34	100±0	99.88±0.11
	a_1	7	67.00 ± 12.67	84.50 ± 3.81	78.67 ± 4.74
HSMM	a_2	7	69.00 ± 7.62	80.50 ± 6.82	76.67 ± 4.64
(v)	*rest*	3	94.00 ± 3.79	100 ± 0	98.00 ± 1.26
	a_1	3	100±0	74.44±1.26	82.96±0.84
HSMM	a_2	3	48.88±2.52	100±0	82.96±0.84
(v/r)	*rest*	2	100±0	100±0	100±0
	a_1	5	89.56 ± 31.49	96.75 ± 8.65	94.35 ± 11.26
CHMM	a_2	5	99.69 ± 0.53	93.25 ± 17.80	95.40 ± 11.85
(v/r)	*rest*	3	90.75 ± 16.99	100 ± 0	96.92 ± 5.66
our	a_1	5	96.04 ± 9.69	97.19 ± 4.04	96.81 ± 3.43
CHMM	a_2	5	93.98 ± 6.26	95.38 ± 4.49	95.58 ± 4.00
(v/r)	*rest*	3	95.13 ± 2.98	100 ± 0	98.38 ± 0.99
our	a_1	3	72.13 ± 10.71	92.06 ± 3.33	85.42 ± 2.32
CHSMM	a_2	3	84.13 ± 6.65	86.06 ± 5.36	85.42 ± 2.32
(v/r)	*rest*	2	100 ± 0	100 ± 0	100 ± 0

to detect dynamics from rest condition accurately. Fig. 6.1 illustrates the source of the problem. This figure depicts two dimensions of an observation, as well as their fitted Gaussian pdfs, each of which corresponds to a state m of channel c of a model ($b_m(o_t^c)$). Focusing on dark circles in the figure, we can observe that for a few samples of the rest condition, the probability of observation generated by one of the dynamic models is more than its probability in rest model. This problem can be suppressed by increasing the number of states in rest model, which leads to slow execution rate. So we accept some errors in classifying rest class in order to separate dynamics from each other accurately and rapidly. Whereas, the results show that CHSMM is capable of separating the dynamics from rest condition accurately (100% in all metrics), although its poor performance in classification of the dynamics. Similar performance can be observed in results of multivariate HSMM. We can conclude that coupling in HSMM perfectly separates dynamics from rest condition, however, coupling in HMM is able

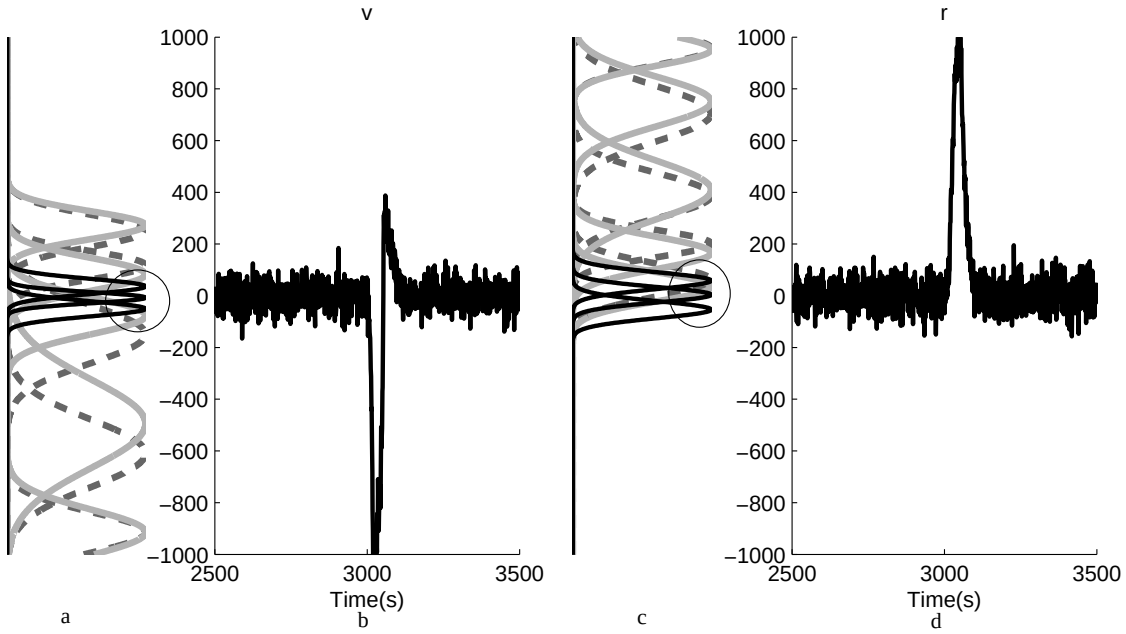

Figure 6.1: Examples of simulated responses for state variables of the FitzHugh-Nagumo model (observations) v (b) and r (d), and the fitted Gaussian pdf $b_m(o_t^c)$ obtained for each state (a and c, respectively) after the training phase. m different Gaussians can be observed in each case, each one characterizing one state of the corresponding model. In a and c traces of this figure, black, light grey and dashed dark grey kernels belong to states of rest, a_1 and a_2 models respectively. Circles show the source of error around rest conditions.

to sense subtle differences of characteristics between two dynamics. $D(c)$ is considered equal to 5 for simulated data analysis.

6.2.2 Detection on synthetic data

For on-line detection, performance of proposed methods is further analyzed by determining the optimum value of sensitivity and specificity using ROC curves which are traced based on the values of sensitivity, specificity for various values of threshold on total likelihood. The detection of dynamic a_1 is accomplished with first training three models corresponding to a_1, a_2 and rest. Fig.s 6.2 and 6.3 depict the relation of states to the range of observation amplitude which is obtained after training of the three models using CHMM and CHSMM respectively in disturbance detection task. In these figures, all signals used for training are plotted. Moreover, the states are shown according to their mean and variance of their Gaussian kernel. These characteristics determine intervals in the observation range of amplitude which are related to the states. In another word, each state can be represented by a mean and a variance (we assume just one Gaussian kernel for simplification of such a complex issue as coupling.) which is interpreted as an origin in the range of amplitudes of observation.

Fig. 6.4 and 6.5 illustrate the CHMM based processing of likelihoods obtained according to Eq.s (3.137) and (3.139). Fig. 6.4(a,b) shows 2-D observations generated by

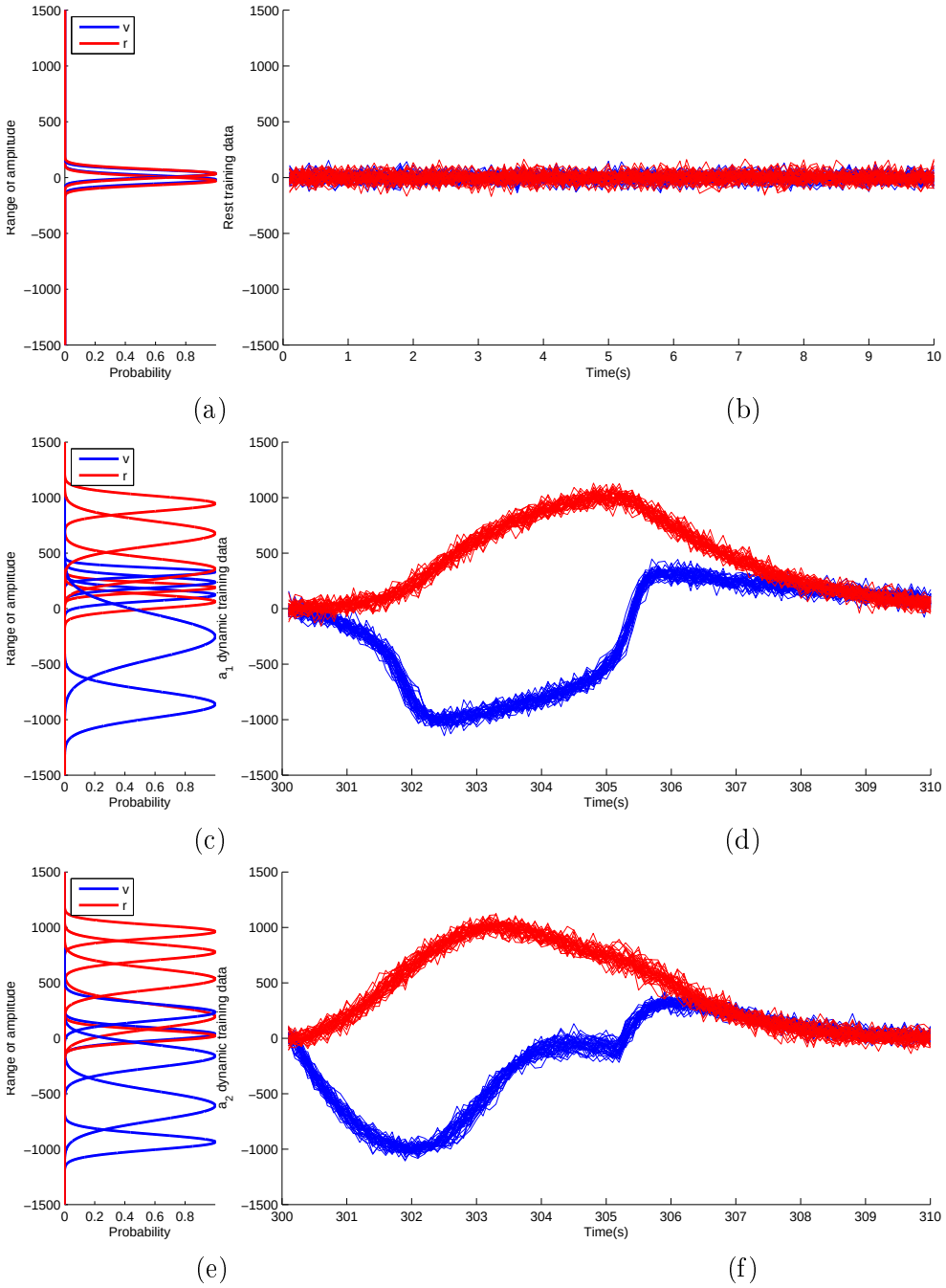

Figure 6.2: The state allocation in range of amplitude of training data in simulated disturbance detection using CHMM model. The optimal number of states for rest and disturbance classes are 2 and 5 respectively. (a) States of rest model. (b) Training data of rest condition including first 10 seconds of 2-D observation. (c) States of a_1 model. (d) Training data of a_1 dynamic including 300-310 seconds of 2-D observation. (e) States of a_2 model. (f) Training data of a_2 dynamic including 300-310 seconds of 2-D observation.

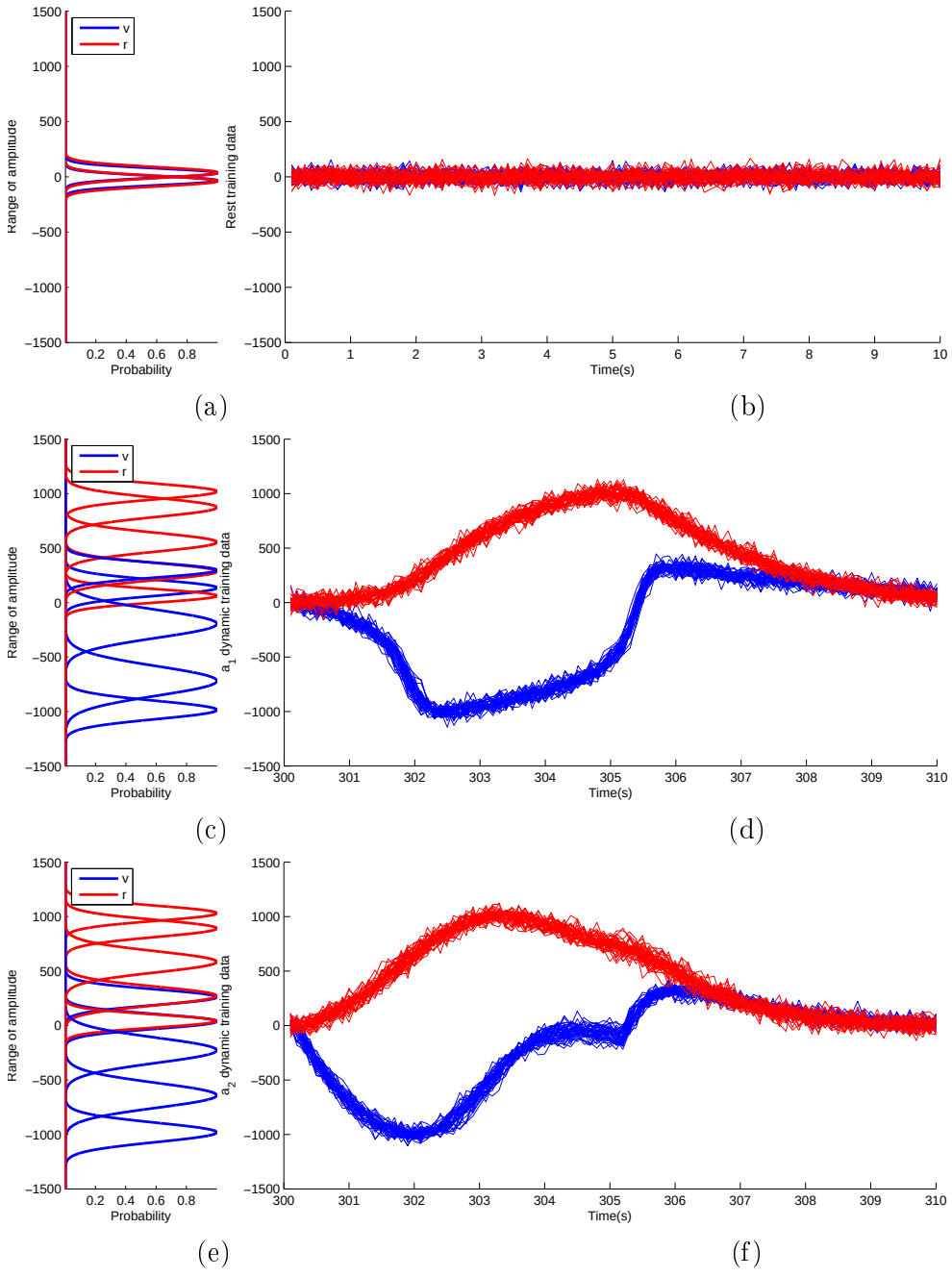

Figure 6.3: The state allocation in range of amplitude of training data in simulated disturbance detection using CHSMM model. The optimal number of states for rest and disturbance classes are 2 and 5 respectively. (a) States of rest model. (b) Training data of rest condition including first 10 seconds of 2-D observation. (c) States of a_1 model. (d) Training data of a_1 dynamic including 300-310 seconds of 2-D observation. (e) States of a_2 model. (f) Training data of a_2 dynamic including 300-310 seconds of 2-D observation.

a_1 and a_2 dynamics. The tree next rows of traces (Fig. 6.4(c-h)) depict ll_k^c, the likelihood of channel c resulted by kth model. The summations of likelihoods over channels and ll_{total}^{kj} are shown in Fig. 6.5(a-b) and Fig. 6.5(c-f) respectively. The process of a_1 dynamic detection is accomplished by first detecting the disturbance from rest condition using a threshold on $ll_{total}^{\{a_1\}\{rest\}}$. If a disturbance is detected, then by using second threshold on $ll_{total}^{\{a_1\}\{a2\}}$, we are able to determine which dynamic (a_1 or a_2) has generated the observation. The same traces using CHSMM framework are illustrated in Fig. 6.6 and Fig. 6.7. As can be seen in Fig. 6.5(e-f) and 6.7(e-f), by choosing an appropriate threshold, the observations generated by dynamics a_1 and a_2 can be successfully separated. The alterations of $ll_{total}^{\{a_1\}\{rest\}}$ comparing to $ll_{total}^{\{a_1\}\{a2\}}$ are enough intensive for detection of disturbance, as in Fig. 6.5(c-d) and Fig. 6.7(c-d).

The detection of disturbances generated by dynamic a_2 are not analyzed since it does not provide any additional information. Similar to classification, we assume the same number of states for dynamic models while the model corresponding to rest condition has less states. $D(c)$ is considered equal to 5 for simulated data analysis. The results corresponding to the optimal number of states based on BIC algorithm are reported in table 6.2 where all the metrics are related to the optimal threshold value (PD) of the corresponding ROC. The results of the proposed approaches are compared to conventional HMM-based detection techniques (HMM and HSMM). Furthermore, they are compared with a CHMM framework previously proposed by Rezek [Rezek et al., 2000]. An interesting result happens while using more than one dimension of observation, as can be observed, multivariate HMM achieves better SEN and SPC than univariate HMM. Moreover, coupling the dimensions leads to better results than even multivariate HMM because in CHMM, each dimension is processed by a Markov chain and the coupling between them can extract the information more efficiently than using these dimensions in multivariate way in a single HMM. The SEN and SPC of our proposed CHMM reached 94.32% and 99.34% respectively and the time delay is 0.98 ± 0.16 s which are superior than even CHMM of Rezek. According to the results of this table, the proposed CHSMM demonstrates marginally better performance in terms of SEN (96.67%), SPC (98.98%) and mean time delay (1.58 s), comparing to univariate and multivariate HSMM which implies that the coupling approach is able to perform far better than even multidimensional observation as discussed about the comparison of HMM and CHMM. Furthermore, although the proposed CHSMM achieves the best sensitivity and its specificity is nearly the same as the others, its mean time delay value is higher than methods like (univariate/multivariate) HMM and CHMM. This fact can be seen in methods based on HSMM and it is likely due to the effect of parameter d which allows the model to rest in a state for several samples; and hence the model has more inertia, whereas more dynamical characteristics are considered.

6.2.3 Detection of AB on real data

For real data, only the issue concerning the detection of AB is studied. As discussed in part 5.2.2, the real data used for Markovian approach include three time series (feature signals) extracted from ECG signal: RAMP, RR and QRSd. In order to analyze such data, we construct LS1 and LS2 training data for obtaining the parameters of the models corresponding to normal and bradycardia episodes and calculate the metrics on

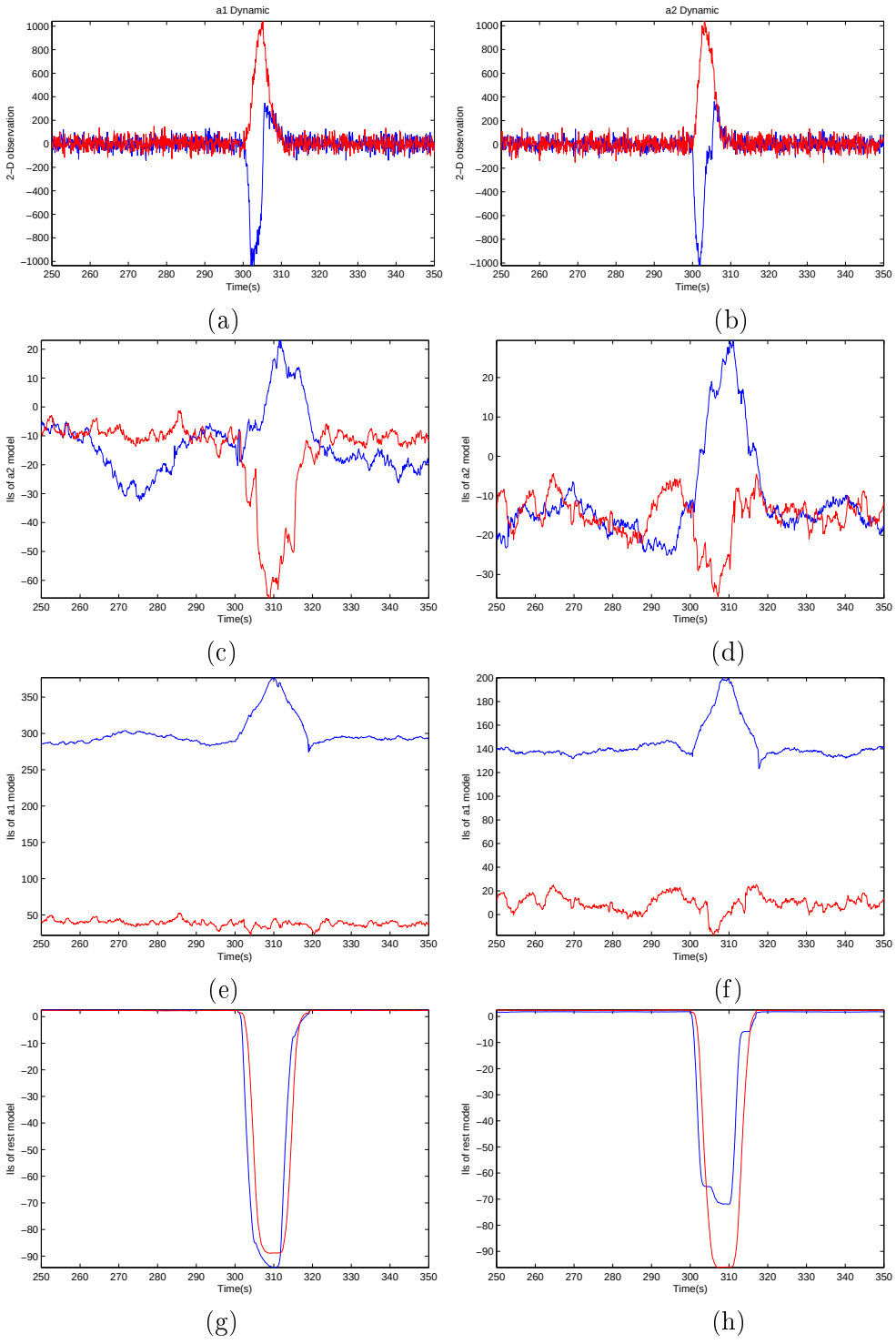

Figure 6.4: The likelihoods obtained after performing CHMM inference on two 2-D observations generated by different dynamics. ll denotes likelihood. (a) 2-D observation generated by dynamic a_1. (b) 2-D observation generated by dynamic a_1. (c,d) Likelihoods obtained for each channel corresponding to model a_2. (e,f) Likelihoods obtained for each channel corresponding to model a_1. (g,h) Likelihoods obtained for each channel corresponding to rest model. Likelihoods obtained from rest model are multiplied by 0.01 to be displayed appropriately.

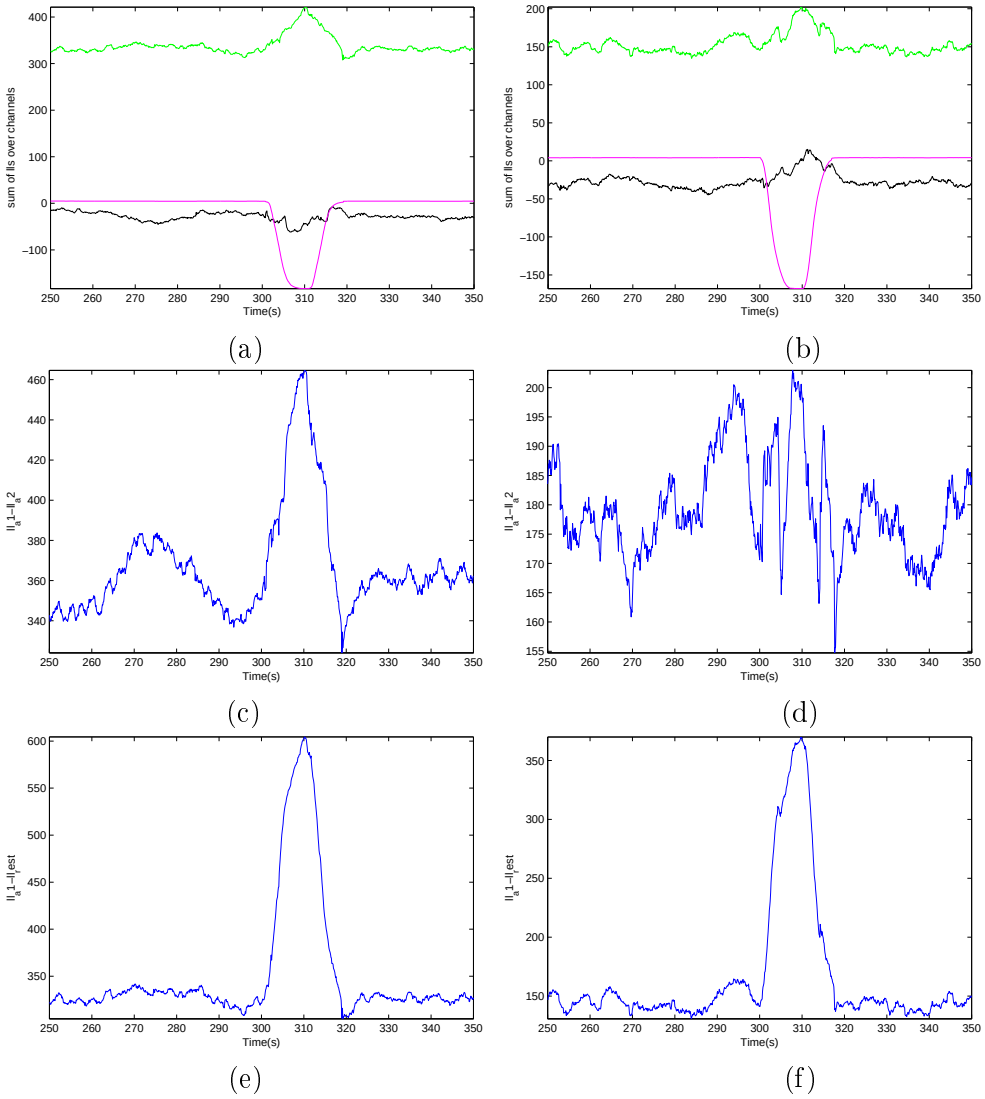

Figure 6.5: Continue of the CHMM processing in Fig 6.4. (a,b) Summation of likelihoods over channels. Green, black and magnesia likelihoods are related to a_1, a_2 and rest models. (c,d): Subtraction of likelihoods of a_1 from a_2. (e,f) Subtraction of likelihoods of a_1 from rest.

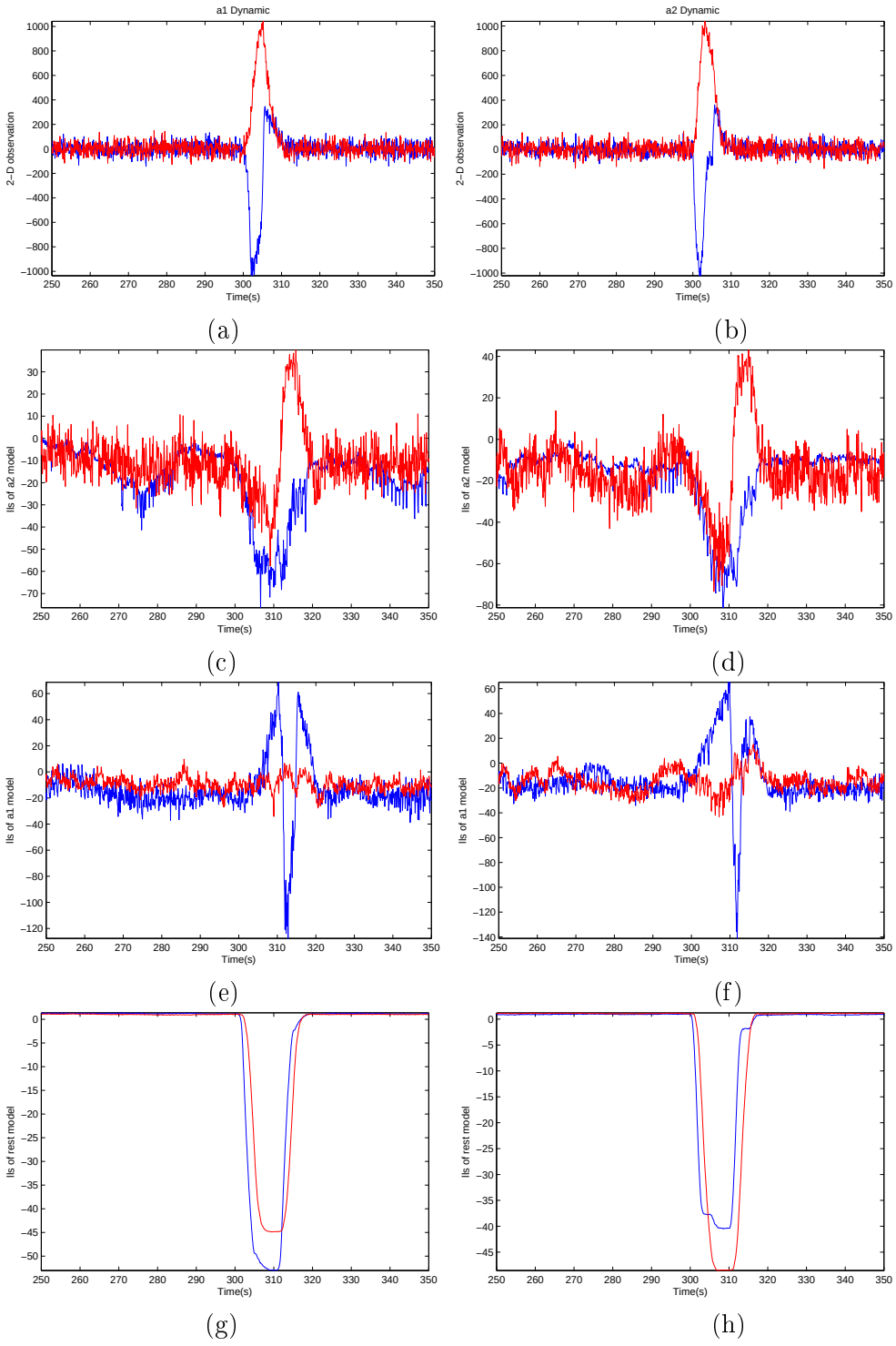

Figure 6.6: The likelihoods obtained after performing CHSMM inference on two 2-D observations generated by different dynamics. ll denotes likelihood. (a) 2-D observation generated by dynamic a_1. (b) 2-D observation generated by dynamic a_1. (c,d) Likelihoods obtained for each channel corresponding to model a_2. (e,f) Likelihoods obtained for each channel corresponding to model a_1. (g,h) Likelihoods obtained for each channel corresponding to rest model. Likelihoods obtained from rest model are multiplied by 0.01 to be displayed appropriately.

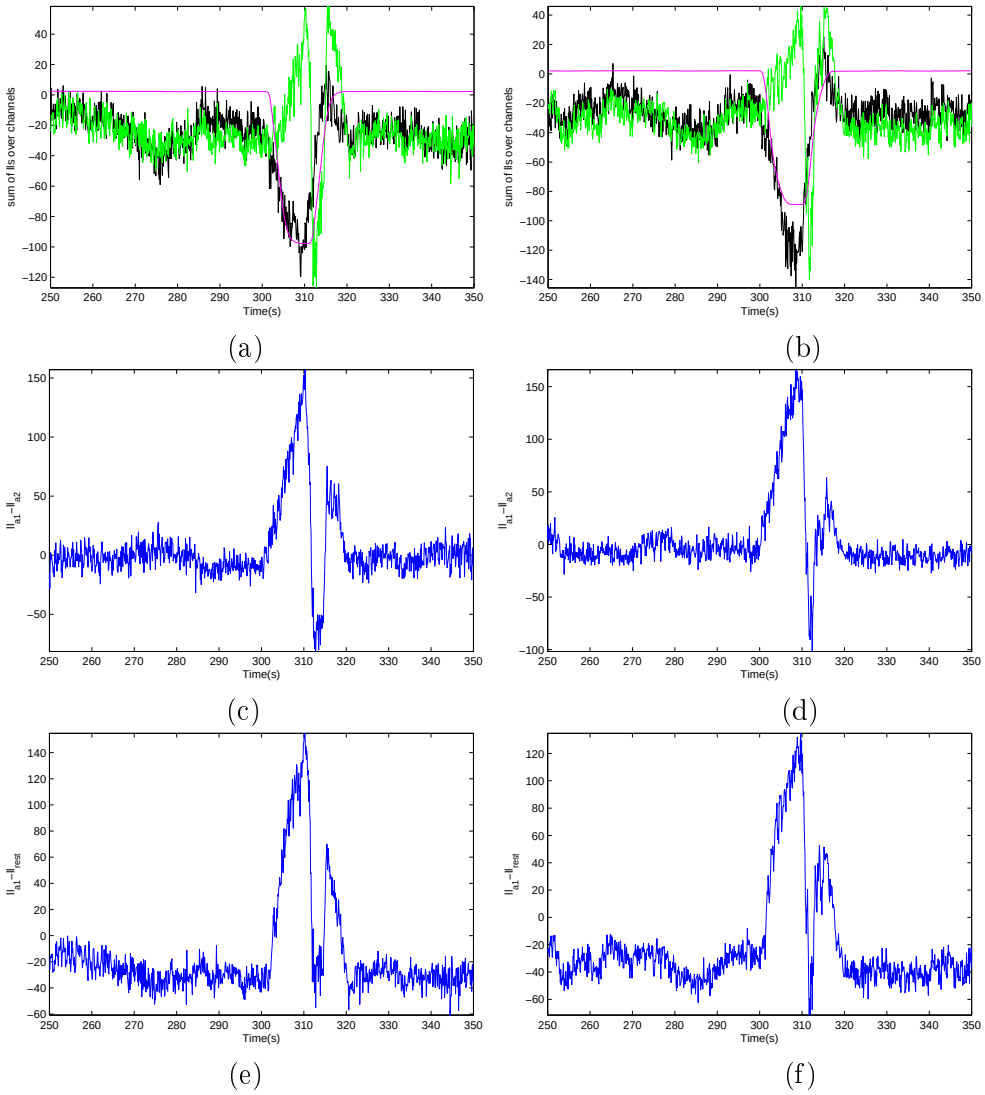

Figure 6.7: Continue of the CHSMM processing in Fig 6.6. (a,b) Summation of likelihoods over channels. Green, black and magnesia likelihoods are related to a_1, a_2 and rest models. (c,d): Subtraction of likelihoods of a_1 from a_2. (e,f) Subtraction of likelihoods of a_1 from rest.

Table 6.2: Comparison of the cross-validation results corresponding to Markovian methods for optimal state number for dynamic a_1 detection

Method	♯ states	SEN(%)	SPC(%)	mean delay(s)	std delay(s)
HMM (v)	2-5-5	61.38±2.65	89.73±2.41	0.40± 0.02	1.01±0.41
HMM (v/r)	2-4-4	86.61±6.56	85.31±3.78	0.24±0.21	1.05±0.19
HSMM (v)	3-5-5	90.78±2.39	98.81±0.36	2.84±0.71	0.11±0.04
HSMM (v/r)	3-5-5	94.99±0.02	99.45±0.01	1.59±0.01	0.0671±0.01
CHMM (v/r)	2-4-4	94.38± 0.09	99.42±0.01	1.78±0.03	0.07±0.04
Our CHMM (v/r)	2-5-5	94.49±0.50	99.34±0.12	0.98±0.21	0.16±0.15
Our CHSMM (v/r)	2-5-5	96.67±0.35	98.98±1.45	1.58±0.18	1.06±0.49

Table 6.3: Summary of the results of proposed CHMM using various synchronization time delays for RAMP.

Synchronization Delay(s)	SEN(%)	SPC(%)	mean delay(s)
3	92.82	91.67	0.95±2.99
3.5	89.34	91.25	1.12±3.55
4	93.45	97.25	-0.15±2.09
4.5	92.38	85.84	-0.59±5.39

40 test data. This process is repeated 5 times for cross-validation performed according to hold-out method. For applying CHMM and CHSMM models in the method described in part 3.4, there exists some parameters that require to be optimized. Furthermore, different combination of feature signals are evaluated to achieve the best results in each tasks.

Among the three time series, RAMP data is modulated by respiration activities [Haskova et al., 2013b]. In other word, the effect of apnea will first appear in the RAMP feature. Hence, there exists an asynchronization between RAMP and the rest of the feature signals which can be parameterized by the synchronization time delay (τ). The delay values are observed between $3 - 4.5$ s [Poets, 2010]; however, it has to be chosen optimally. For CHMM, the performance of the detection in terms of different values of τ are reported in table 6.3. This study is performed using RR and RAMP features while considering 5 states for both models (bradycardia and normal). Similar study is reported in table 6.4 for CHSMM. The results are obtained using RR and RAMP features while considering 2 states for both models (bradycardia and normal). we have used the synchronization time delay (τ) for RAMP signal in all the following results, whose optimal values are observed about 4 s and 4.5 s for CHMM and CHSMM respectively.

The performance for the bradycardia detection and the detection time delays of the

Table 6.4: Summary of the results of proposed CHSMM using various synchronization time delays for RAMP.

Synchronization Delay(s)	SEN(%)	SPC(%)	mean delay(s)
3	83.58	85.03	0.56±4.28
3.5	92.26	91.08	1.24±3.88
4	88.41	89.94	1.03±5.14
4.5	93.03	93.35	0.31±3.22

proposed models comparing to other models are detailed in table 6.5. This comparisons are reported in various combinations of the three features: (RR-QRSd), (RR-RAMP) and (RR-RAMP-QRSd). The results of the proposed models comparing to other HMM based models using different combination of feature signals are illustrated in Fig.s 6.8, 6.9 and 6.10 for CHMM and in Fig.s 6.11, 6.12 and 6.13 for CHSMM. In these figures, trace (a) illustrates the feature signals used as observations. Traces (b,c) and (d,e) are the likelihoods of channels corresponding to models and their summations respectively. (b) and (d) are related to bradycardia and (c) and (e) are for normal model. The total likelihoods are shown in trace (f) in all of these figures.

In the cross-validation approach, the reported metrics of these tables are related to the PD point of the ROC curves obtained by averaging the SENs and SPCs achieved by an identical threshold. Furthermore, three other metrics (distance to PD, AUC and PW) previously defined in section 5.2.3 are also calculated and reported in table 6.5. The optimum number of states per model is also reported in this table.

The best results of our proposed CHMM in terms of AUC and time delay criteria are achieved by (RR-RAMP-QRSd) combination, for which the average SEN and SPC reach 95.74% and 91.88% respectively. This method has an average time delay of -0.59 s with average standard deviation (std) of 2.79 s. Its PW shows that more than the half of the detections occur earlier than the clinically-annotated onset of bradycardia which implies that AB events are better predicted with the proposed CHMM. In Fig. 6.14, the ROC curves traced according to the resulted SEN and SPC of the proposed models and other HMM based models are shown. In addition, the time delay curve in terms of SPC is also depicted for better comparison. As illustrated in this figure, the coupling of all of the three features provides higher detection rates. Moreover, significant improvement can be observed in all metrics except SPC by comparing our proposed three-channel CHMM with its two-channel configuration.

Similarly, the best (except in TD) results of our proposed CHSMM are achieved using (RR-RAMP-QRSd) combination, for which the average value over 5 rounds of cross-validation of SEN and SPC reach 94.87% and 96.52% respectively and the average time delay is reported as 0.73 s with average std of 2.45 s. Furthermore, its superiority is obvious, especially in terms of AUC and the distance to PD. The value PW in CHSMM shows that more than the half of the detections occur earlier than the clinically-annotated onset of AB which implies that AB events are predicted using such

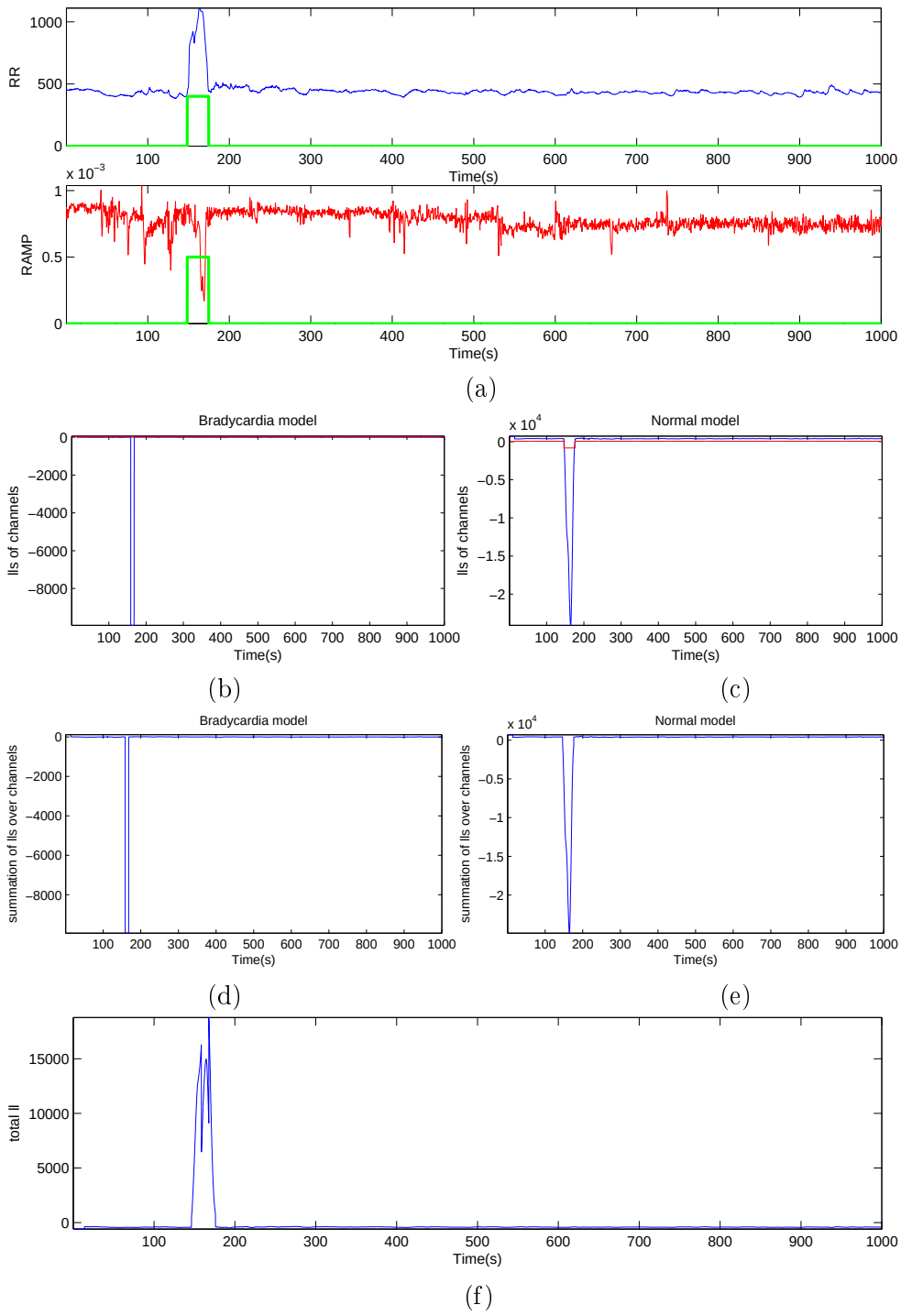

Figure 6.8: The likelihoods obtained after performing CHMM inference on two 2-D observations including RR and RAMP feature signals. (a) 2-D observation. Red line shows the bradycardia incidence. (b,c) Likelihoods obtained for each channel corresponding to bradycardia and normal respectively. (d,e) Summation of likelihoods over channels for each model. (f) Total Likelihood where the threshold is applied.

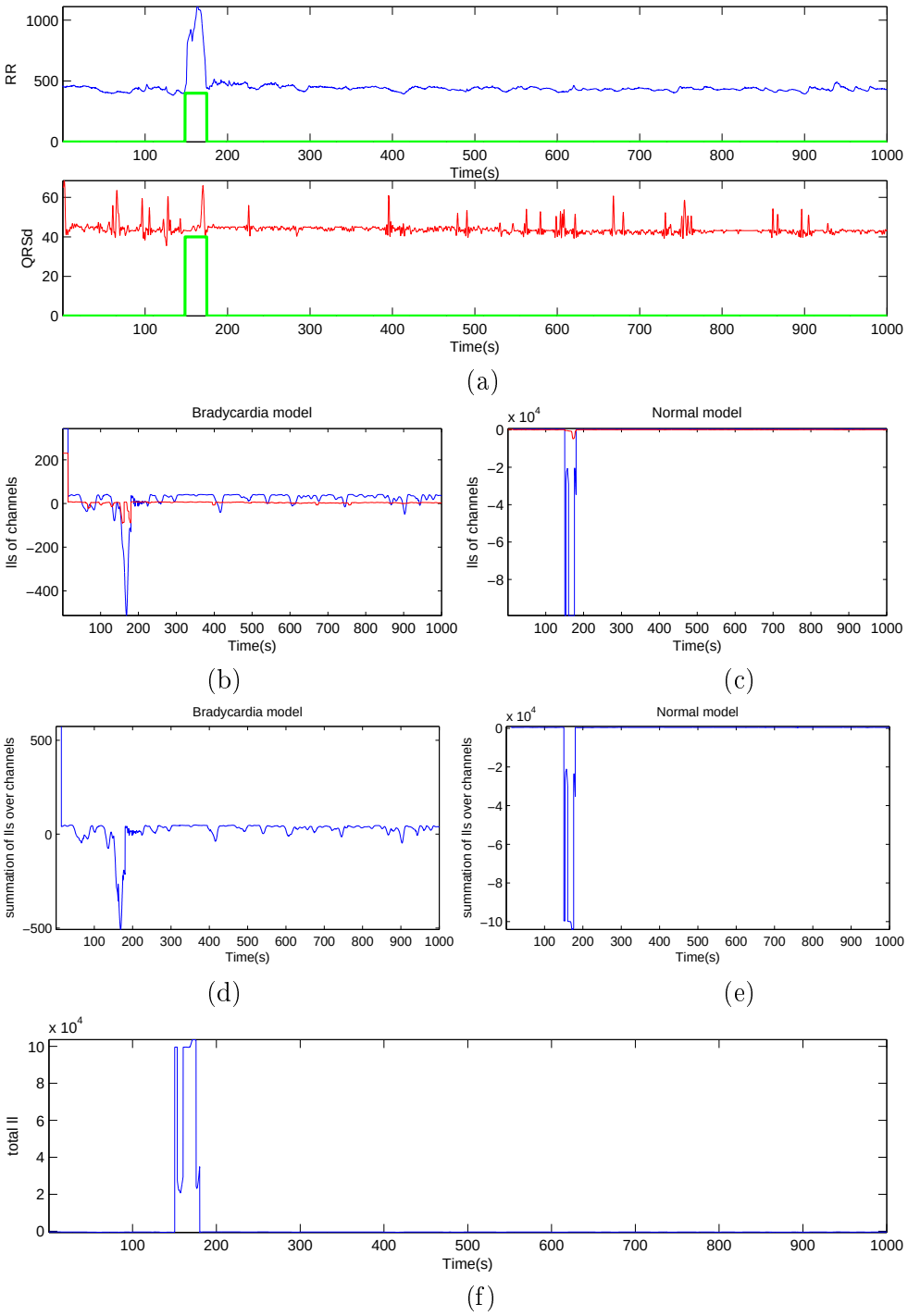

Figure 6.9: The likelihoods obtained after performing CHMM inference on two 2-D observations including RR and QRSd feature signals. (a) 2-D observation. Red line shows the bradycardia incidence. (b,c) Likelihoods obtained for each channel corresponding to bradycardia and normal respectively. (d,e) Summation of likelihoods over channels for each model. (f) Total Likelihood where the threshold is applied.

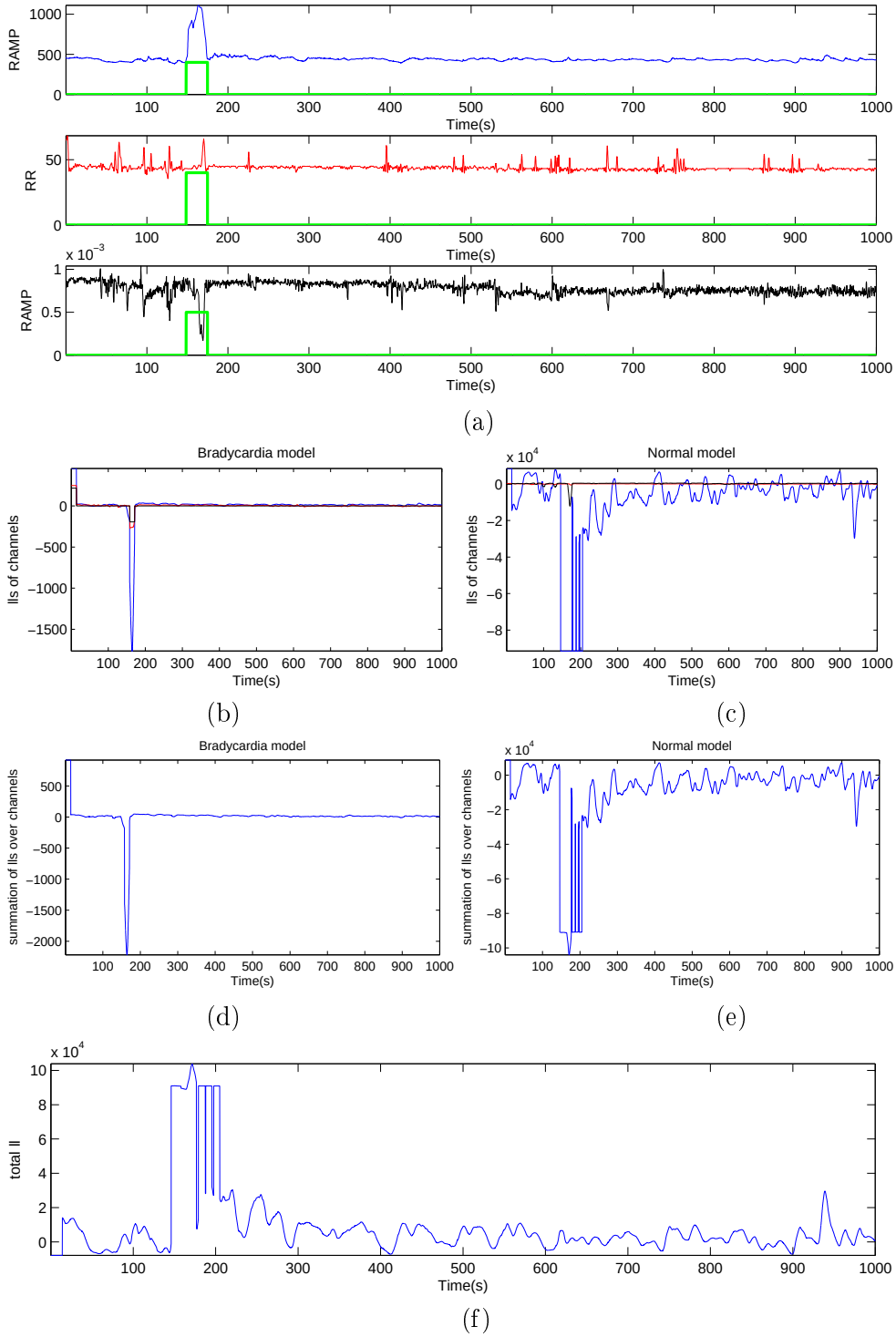

Figure 6.10: The likelihoods obtained after performing CHMM inference on two 3-D observations including RR, QRSd and RAMP feature signals. (a) 3-D observation. Red line shows the bradycardia incidence. (b,c) Likelihoods obtained for each channel corresponding to bradycardia and normal respectively. (d,e) Summation of likelihoods over channels for each model. (f) Total Likelihood where the threshold is applied.

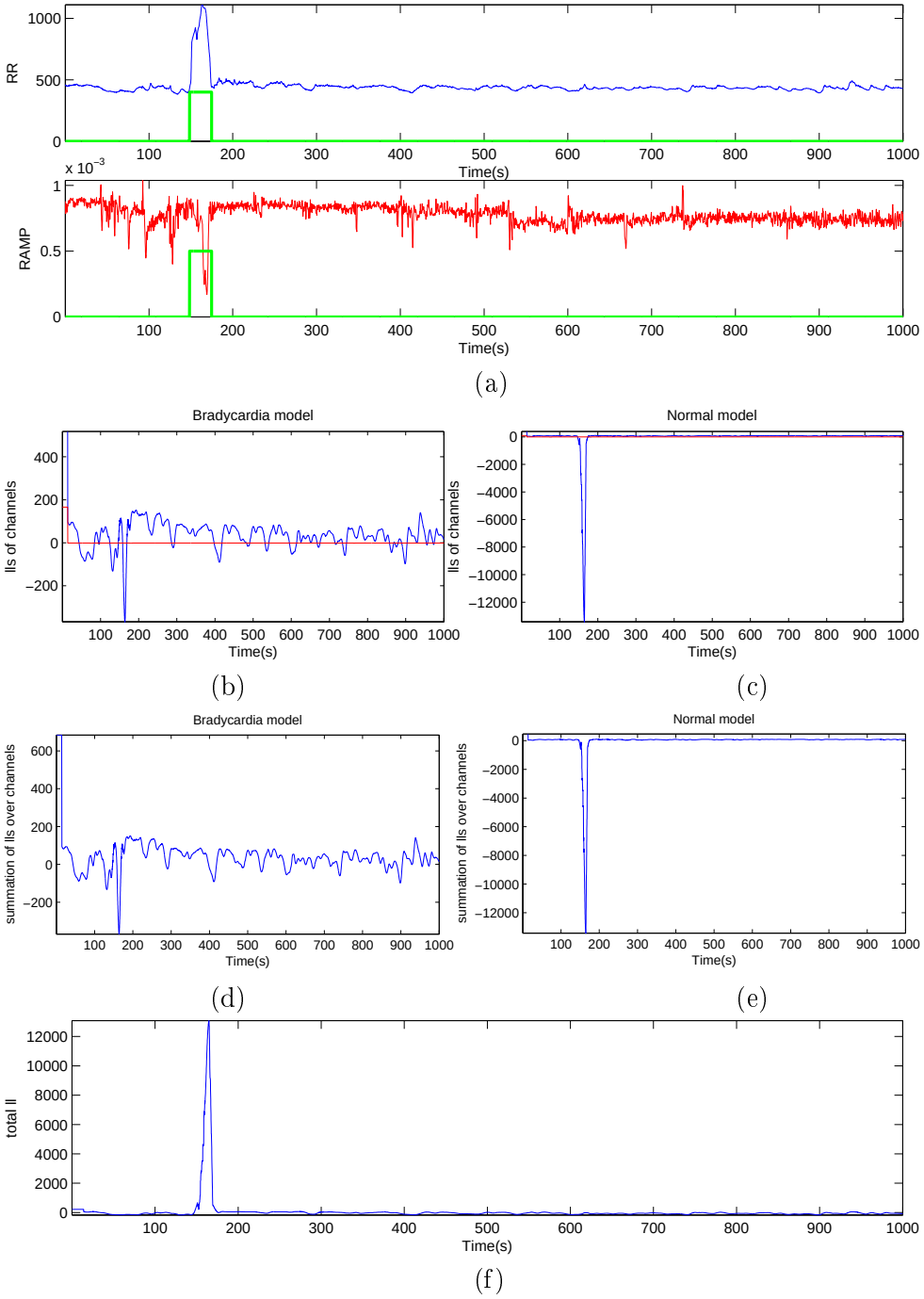

Figure 6.11: The likelihoods obtained after performing CHSMM inference on two 2-D observations including RR and RAMP feature signals. (a) 2-D observation. Red line shows the bradycardia incidence. (b,c) Likelihoods obtained for each channel corresponding to bradycardia and normal respectively. (d,e) Summation of likelihoods over channels for each model. (f) Total Likelihood where the threshold is applied.

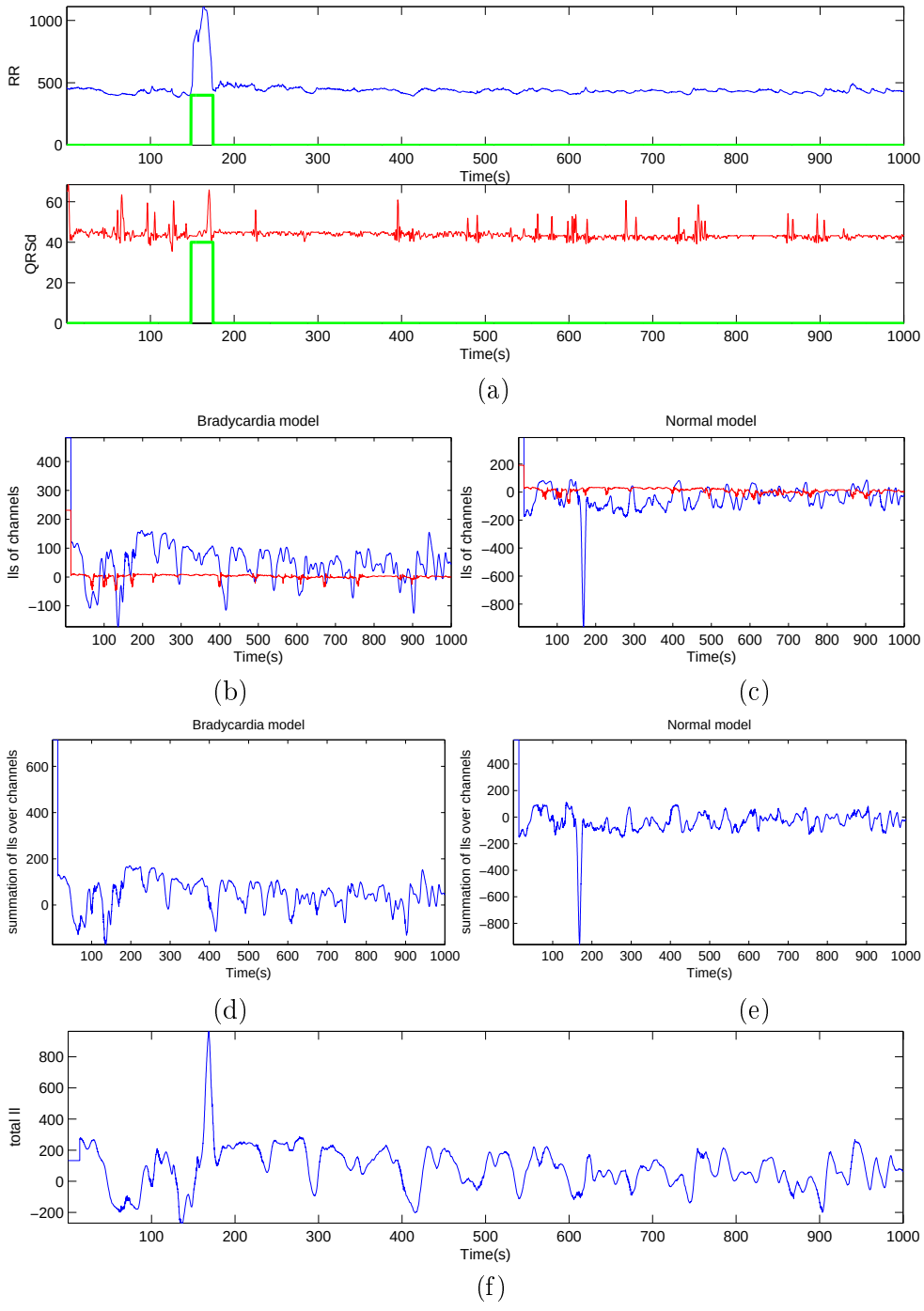

Figure 6.12: The likelihoods obtained after performing CHSMM inference on two 2-D observations including RR and QRSd feature signals. (a) 2-D observation. Red line shows the bradycardia incidence. (b,c) Likelihoods obtained for each channel corresponding to bradycardia and normal respectively. (d,e) Summation of likelihoods over channels for each model. (f) Total Likelihood where the threshold is applied.

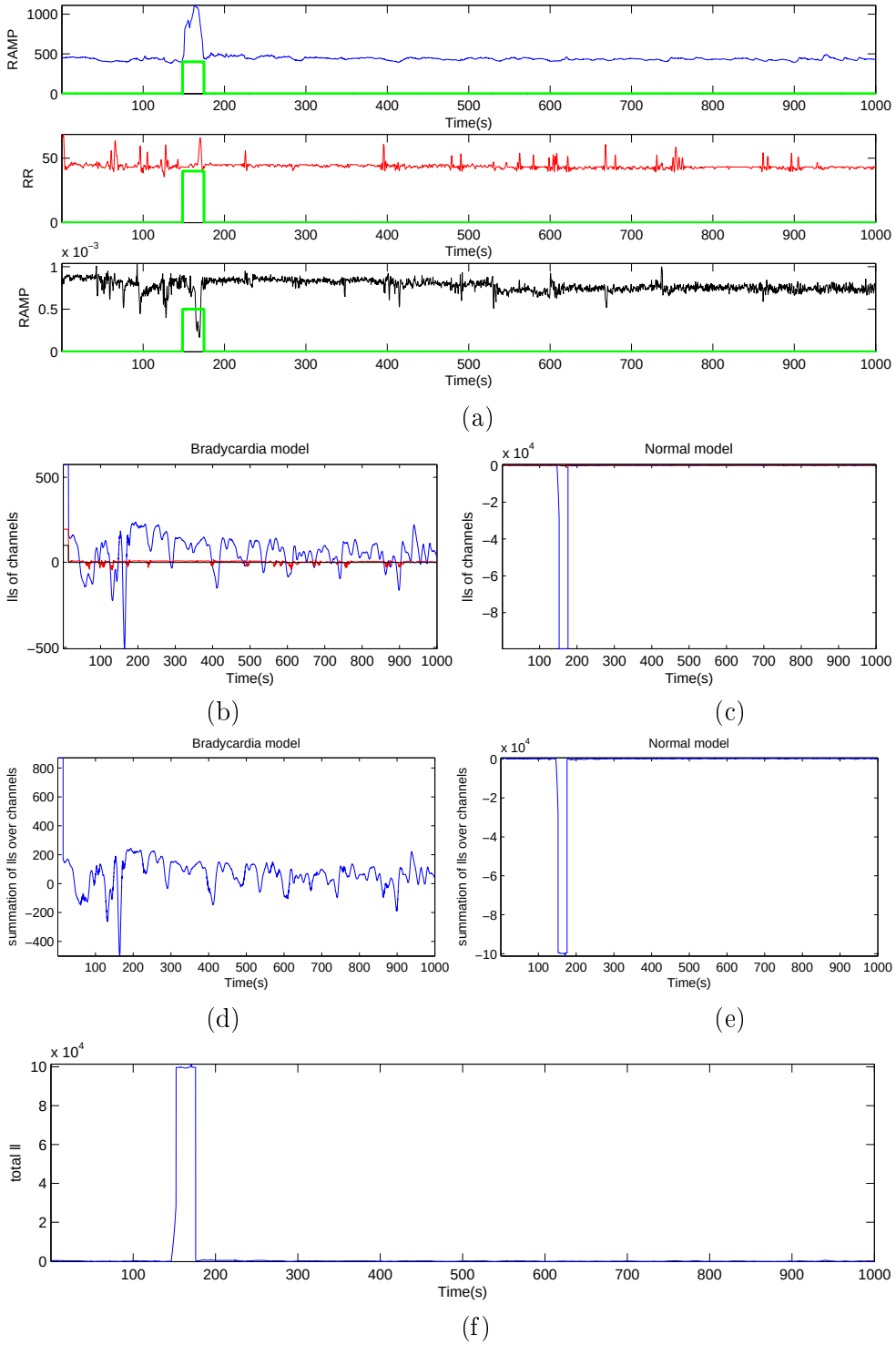

Figure 6.13: The likelihoods obtained after performing CHSMM inference on two 3-D observations including RR, QRSd and RAMP feature signals. (a) 3-D observation. Red line shows the bradycardia incidence. (b,c) Likelihoods obtained for each channel corresponding to bradycardia and normal respectively. (d,e) Summation of likelihoods over channels for each model. (f) Total Likelihood where the threshold is applied.

Table 6.5: Summary of the cross-validation results of proposed CHMM compared with other frameworks using different configurations of features

Features	Method	♯ states	SEN(%)	SPC(%)	mean delay(s)	std delay(s)	PW(%)	AUC	Distance to PD
RR	HMM	6	85.82 ± 6.81	88.34 ± 2.21	0.77 ± 0.97	2.14 ± 0.98	64.30	0.94	0.18
RAMP	HSMM	3	93.74 ± 0.40	91.86 ± 0.23	0.56 ± 0.07	1.49 ± 0.03	60.00	0.95	0.10
	CHMM	4	87.59 ± 0.18	88.05 ± 0.24	1.10 ± 0.03	1.67 ± 0.03	69.23	0.94	0.18
	Our CHMM	5	87.64 ± 0.94	92.13 ± 6.44	0.59 ± 0.01	3.11 ± 1.94	47.29	0.95	0.15
	Our CHSMM	3	95.99 ± 0.31	93.84 ± 0.24	-1.11 ± 0.04	2.56 ± 0.03	35.00	0.97	0.07
RR	HMM	2	82.28 ± 16.87	85.32 ± 3.78	1.81 ± 0.84	4.10 ± 4.09	73.72	0.89	0.23
QRSd	HSMM	5	84.74 ± 0.36	91.33 ±0.18	2.93 ± 0.03	3.08± 0.03	87.50	0.93	0.18
	CHMM	2	84.99 ± 0.26	94.80 ± 0.51	3.43 ± 0.01	2.89 ± 0.03	90.00	0.95	0.17
	Our CHMM	5	87.04 ± 3.78	95.31 ± 1.36	2.51 ± 0.93	3.02 ± 0.18	83.22	0.93	0.14
	Our CHSMM	5	89.97 ± 0.21	92.99 ± 0.31	1.85 ± 0.03	3.14 ± 0.04	78.95	0.95	0.12
RR	HMM	2	91.02 ± 0.64	93.47 ± 0.08	6.55 ± 0.69	2.29 ± 0.79	63.13	0.97	0.11
RAMP	HSMM	7	86.83 ± 25.62	93.77 ± 4.44	1.10 ± 2.74	1.64 ± 1.18	44.10	0.96	0.16
QRSd	CHMM	3	84.42 ± 0.27	92.21 ± 0.32	7.27 ± 0.02	4.26 ± 0.02	60.53	0.94	0.18
	Our CHMM	2	95.74 ± 0.82	91.88 ± 0.31	-0.59 ± 0.21	2.79 ± 0.06	44.23	0.98	0.09
	Our CHSMM	4	94.87 ± 3.18	96.52 ± 0.64	0.73 ± 0.60	2.45 ± 0.69	35.00	0.98	0.06

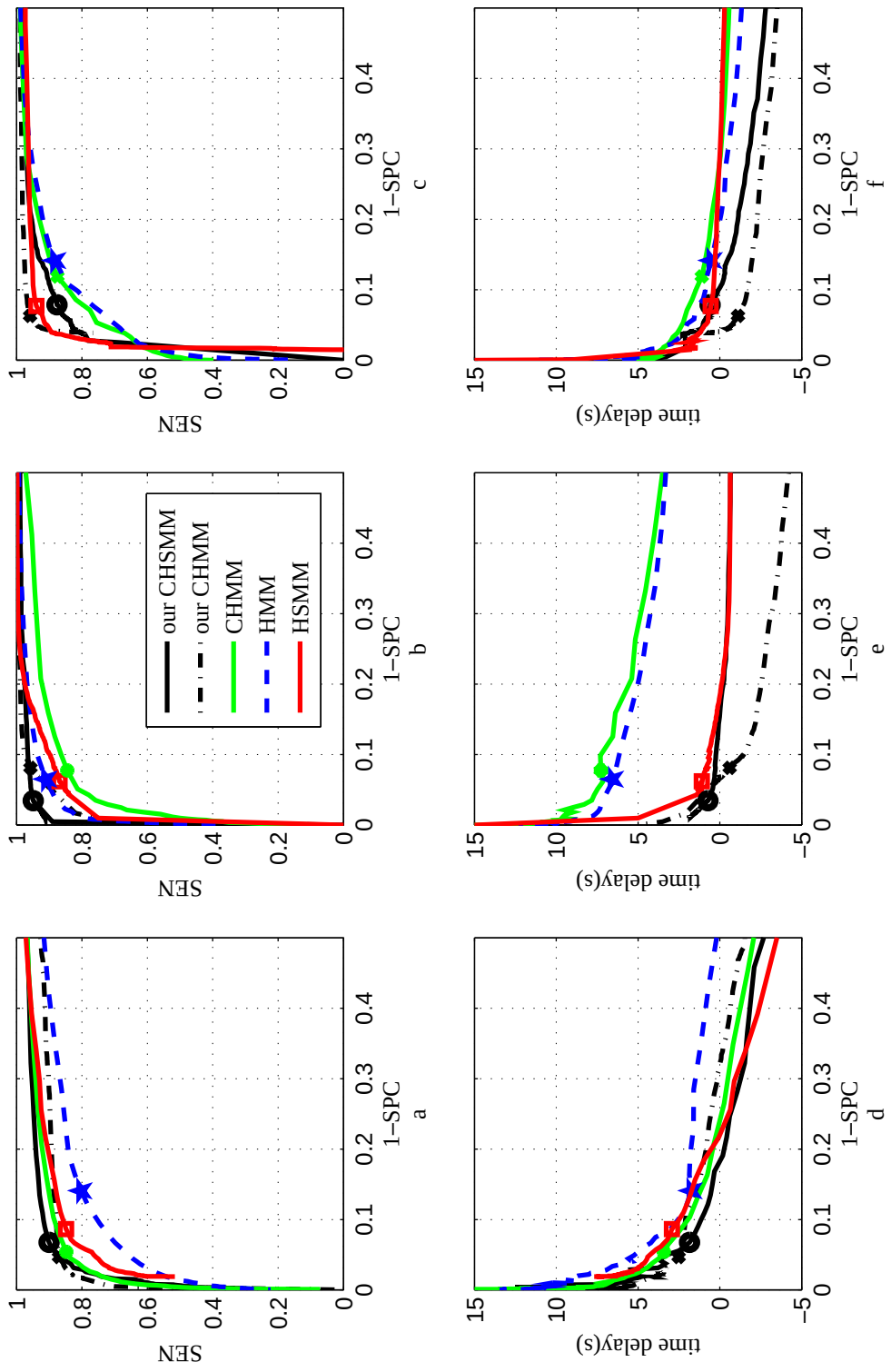

Figure 6.14: Comparing the performance of the proposed methods for AB detection with other frameworks. a-c: ROC curves. d-f: time delay curves versus error. a, d: RR and QRSd. b, e: RR, QRSd and RAMP. c, f: RR and RAMP.

configuration of features. Like CHMM, it can be seen that the (RR-RAMP-QRSd) combination demonstrates the best average performance. (RR-RAMP) combination performs marginally better than the (RR-QRSd), but still underperforms the three-channel combination except in average time delay which is reported as a negative value and shows that the AB episodes are mostly detected sooner than their occurrence in combination of (RR-RAMP).

As illustrated in Fig. 6.14, CHMM performs nearly better comparing to HMM and HSMM frameworks except in (RR-RAMP) combination. It seems that HSMM is able to learn some characteristics of the dynamics which are not considered by HMM and even CHMM. The idea of integrating the HSMM approach in CHSMM arises from comparison of HSMM and CHMM performances. As can be seen in Fig. 6.14, the proposed CHSMM method demonstrates the best average performance in all combination of features except SPC in (RR-QRSd) and time delay in (RR-RAMP-QRSd). It can be totally concluded that better results are reported in the three-channel configuration than the two-channels, since we can observe the improvement in all metrics. The optimum number of states per model is also reported in table 6.5 and $D(c)$ is set equal to 10.

6.3 Evaluation of switching approaches

Implementation of the wave-based and R-based approaches include a training phase primarily for estimation of unknown parameters of the model and a test phase to evaluate the performance of proposed procedure in terms of metrics. Then, quantitative results in terms of SEN, SPC, TD and AC_event of detection are presented. Finally, the best proposed method is compared with existing methods in AB detection context.

6.3.1 Qualitative results of wave-based performance

Fig.6.15 depicts an example of AB detection of wave-based method in a segment. In this figure, a RR signal of the test data is shown in Fig. 6.15(a). Moreover, K_k^i is illustrated in Fig.6.15(b). As depicted in this figure, K_k^i of wave-based method has many fluctuations. Hence, a mode calculation procedure is employed, which is implemented by two methods: MOW and MOB. In MOW, first a mode calculation is employed, hence the local changes in labels which are determined by comparing the K_k^1 with K_k^2, are decreased. Then, the decision on the labels of beat is made using the parameter ξ. The duration of moving window is 360 samples and the value of ξ is optimized using the ROC curve. Each point on ROC curve represents a value for sensitivity and specificity achieved by a specific value for ξ. Fig. 6.16(a) shows ROC curve obtained by altering the value of ξ. The marked point is the perfect detection (PD) point as introduced in Eq. (5.6).

According to Fig. 6.16(a), the parameter ξ is chosen equal to 85% for the best detection of MOW using train-test data. The corresponding time delay plot is also depicted in Fig. 6.16(b). In MOB, the result of mode calculation directly leads to the label determination of the corresponding beat.

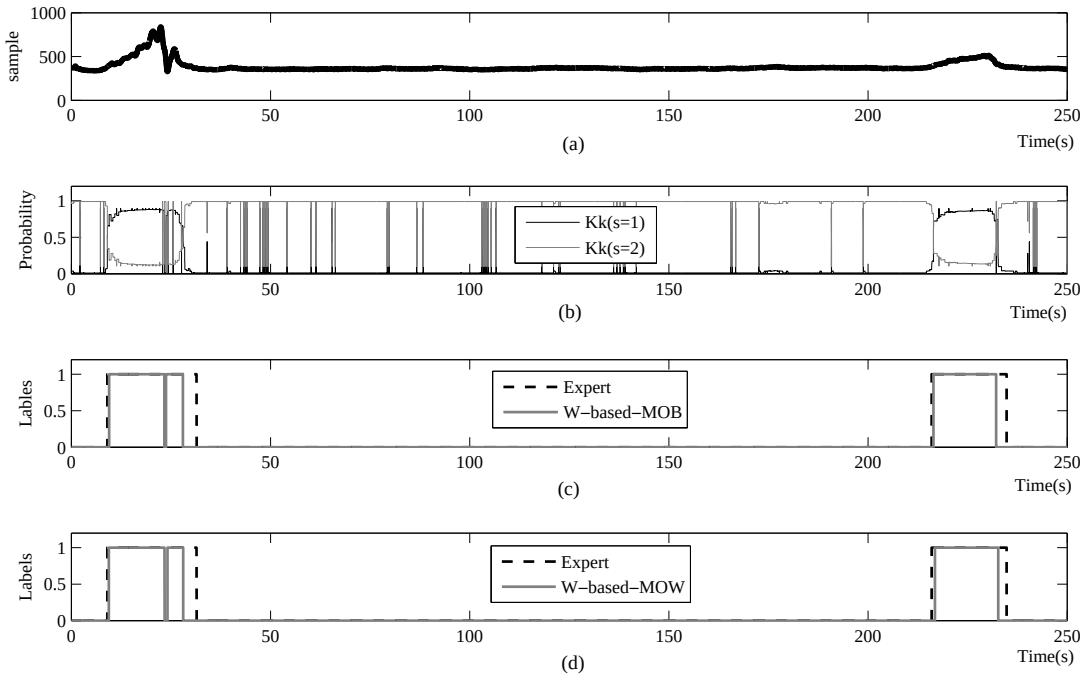

Figure 6.15: The performance of wave-based frameworks. a: RR signal. b: K_k^i for $i = 1, 2$. c and d: The real annotations and labels determined by the algorithm with MOB and MOW processing respectively. Labels 0 and 1 indicate normal and bradycardia respectively.

The annotations determined by an expert and the label decided by the wave-based algorithm with MOB method are shown in Fig. 6.15(c). Similarly, Fig. 6.15(d) illustrates the labels decided by this algorithm with MOW method as well as those determined by an expert.

6.3.2 An example of R-based approach performance

The performance of R-based method on the same example segment is shown in Fig. 6.17. As can be seen, since each sample of the feature RR is extracted from a beat, no fluctuation can be observed in K_k^i (Fig. 6.17(b)), thus unlike wave-based method, mode calculation is not required. The labels obtained by R-based models and the expert annotations are illustrated in Fig. 6.17(c).

6.3.3 Quantitative results

The metrics for the proposed approaches over test data calculated by cross-validation are reported in tabel 6.6. According to table 6.6, MOB comparing with MOW is found to be more reliable to detect AB event. This shows that the assumption used for defining the beginning and ending of the beats in ECG is significant. In addition, calculation of mode over beats mainly reduces the amount of computation comparing to MOW in which the mode is computed sample by sample. The results show the

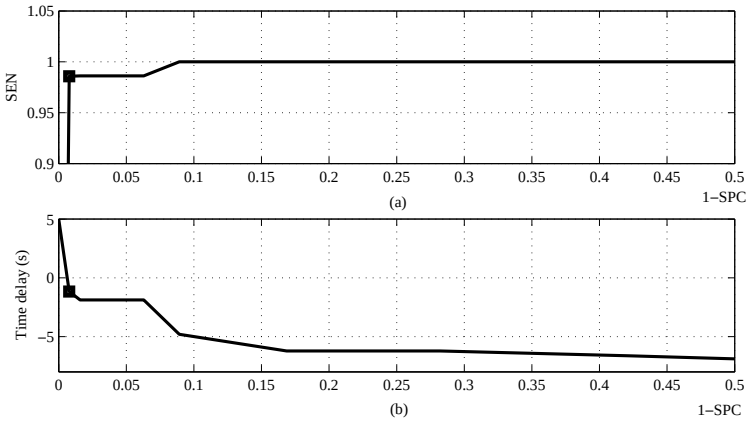

Figure 6.16: Searching for the best value of ξ which is the precentages of the samples in a beat labeled as AB event by our wave-based MOW method. The results are achieved on train-test data. a: ROC curve (SEN vs. 1-SPC). b: Time delay of detection vs. 1-SPC. PD is marked by an square in each plot.

Table 6.6: Proposed Wave-based (MOW and MOB) and R-based methods performance evaluation using cross-validation. Accuracy is computed over the average number of events detected correctly. The rest of the metrics are in terms of samples.

method	AC	SEN	SPC	mean TD (s)	std TD(s)
MOW	99.29±0.40	94.51±0.35	92.68±0.30	1.78±0.52	1.77±0.53
MOB	99.11±0.73	94.74±0.69	94.17±0.87	0.35±0.13	2.12±0.29
R-based	99.13±0.02	90.35±5.46	88.59±11.28	1.80±0.89	1.94±0.43

ability of wave-based and R-based methods to detect the AB occurrence accurately. However, wave-based methods reflect better detection performance, although they are more complicated. Moreover, they are capable of detecting the AB nearly faster than R-based method.

The average values of SEN and SPC using MOB are 94.74% and 94.17% respectively. The low average time delay in this method, 0.35±2.12, is achieved by searching for AB sample by sample in SEKF algorithm with MOB postprocessing and introduces this approach as a fast detector. Whereas, in R-based detector, processing of the RR signal looses the intra-beat information and higher average value for time delay, 1.80±1.94, is reported.

It can be seen that the wave-based MOB demonstrates the best average performance. MOW performs marginally better than the R-based, but still underperforms the MOB approach while the AC_event of all methods are more than 99%.

It should be noted that in wave-based detectors, the average standard deviation (std) of time delay are larger than its mean, which implies that in some cases, these detector predict the occurring of AB since its onset is reported prior to the annotation determined by the expert. Moreover the reported value of time delay is the average of all obtained values consisting of the prediction with negative time delays and the detection cases with positive time delays.

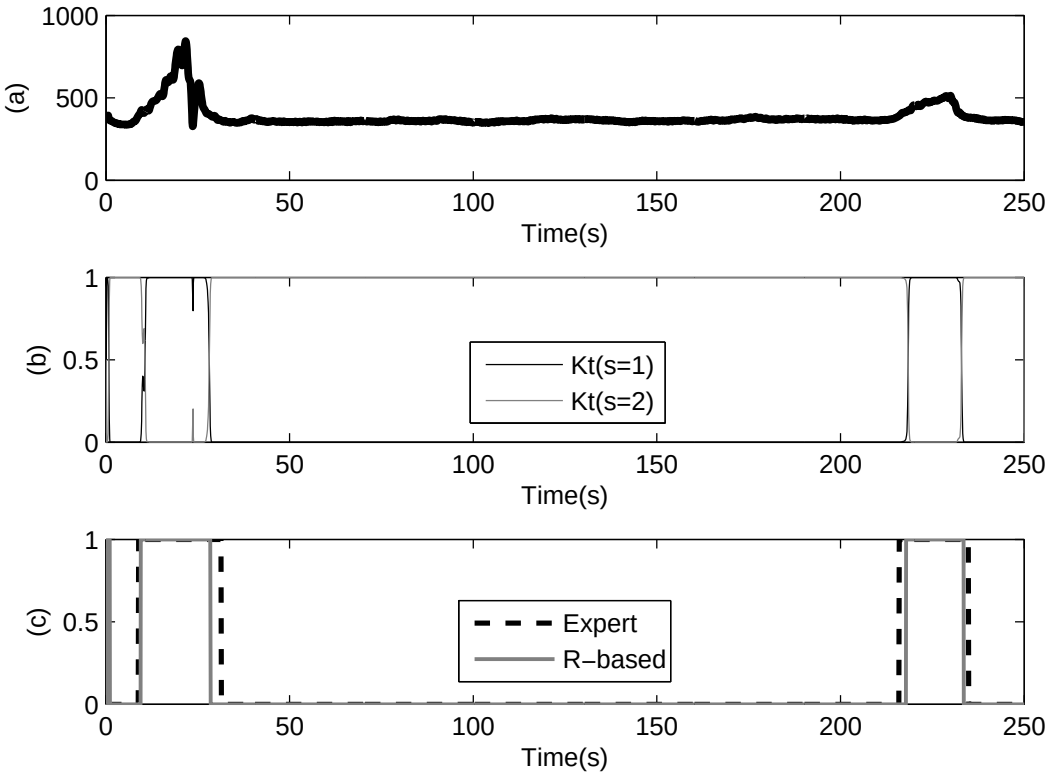

Figure 6.17: The performance of R-based frameworks. a: RR signal. b: K_k^i for $i = 1, 2$. c: The real annotations and labels determined by the algorithm. Labels 0 and 1 indicate normal and bradycardia respectively.

6.3.4 Comparing MOB with other existing methods

In order to compare the proposed methods with other benchmarks, we have studied the previous methods on our database for AB detection in preterm infants. Furthermore, the conventional threhsolding has been tested. In this simple method, the value of threshold is determined on train data and the evaluation is performed over test data. The reported results are based on R peak detection (Pan and Tompkins 1985) and RR signal extraction (the same algorithm used in proposed methods), together with a fixed hard thresholding strategy. For performance evaluation, Altuve et al. (2011) used 148 RR series (series duration = 26.25±11.37 minutes) with 233 bradycardia episodes. They employed 48 series for training and left the rest for test. We have used the same records; however, to ensure the consistency of the results on various subjects and reducing the number of normal temporal samples, the full-length of the records were divided in to 105 segments with 250 s duration with the same total number of AB episodes. The whole procedure was repeated 5 times over the training data consisting of the 20% of selected records, each time using the same initial parameters.

The average SEN and SPC of the AB detection for different methods are depicted in table 6.7. For HSMM and HMM methods, two values for each metric are reported, the first one is the value reported in [Altuve et al., 2015], and the second one (shown by $*$ in the table 6.7) is obtained by our implementation of the corresponding algorithm.

Table 6.7: Comparing the best results with other methods

	method	SEN	SPC	mean TD (s)	std TD (s)
our work	MOB	94.74±0.69	94.17±0.87	0.35±0.13	2.12±0.29
Altuve 2014	HSMM	88.66±1.72	92.87±0.86	1.59±0.24	3.61±0.30
*		88.46±1.16	93.57±1.14	2.10±0.66	2.26±0.81
Altuve 2014	HMM	86.52±3.96	92.27±1.77	1.61±0.43	3.74±0.32
*		89.06±1.08	92.28±0.71	1.59±0.36	1.28±0.17
Masoudi 2013	CHMM	84.92±0.26	94.17±0.51	2.32±0.01	4.82±0.03
-	Thresholding	87.98±0.88	84.46±1.42	2.71±0.12	2.07±0.17

* The results are obtained based on [Altuve et al., 2015] and by our implementation to reconstruct raw results.

The minor differences observed between these two results are due to small differences in implementation (for instance in optimum number of states in the models and the cross-validation procedure). It can be seen that the wave-based MOB demonstrates the best average performance in lower ratio of the number of records in training over those in test procedures. Furthermore, its superiority is obvious, especially in terms of average time delay, where it can detects AB far faster than the others. The thresholding method finds the AB episodes based on the amplitude of observation and according to table 6.7, it has weaknesses in time delay and the precision of detection, while other methods learn the dynamic rather than depending on the amplitude of observation.

Fig. 6.18 is illustrated in order to visually compare the repeatability of the results of table 6.7. It shows five values for each metric resulted from the cross-validation rounds. The values corresponding to HSMM and HMM are obtained from our implementation. Moreover, lower quartile, median and upper quartile values of the results are included. As depicted in Fig. 6.18, the CHMM method proposed by Masoudi et al. demonstrates lower std among the under studied methods. In this method, the QRS complex duration is also employed as the second feature beside RR signal; hence, it might be resulted in low variance in metrics. However, it still takes more time to detect AB. While, MOB shows reasonable std and mostly the best average results.

6.4 Conclusion

In this chapter, the results of the evaluation of the proposed methods in both Markovian and Switching approaches were presented. For Markovian approach, three scenarios based on the proposed models were studied which include: classification for simulated data, detection for simulated data and AB detection for real data. The results in this approach were compared with those of other existing HMM based models. For Switching approach, the methods which are established specifically for AB detection were evaluated. Finally, obtained results were compared with other existing methods presented for AB detection.

It is interesting to compare the results of the two proposed approaches. Table 6.8 illustrates the best results of these methods. As can be observed from this table, CHSMM shown the best precision in AB detection whereas CHMM had the lowest TD

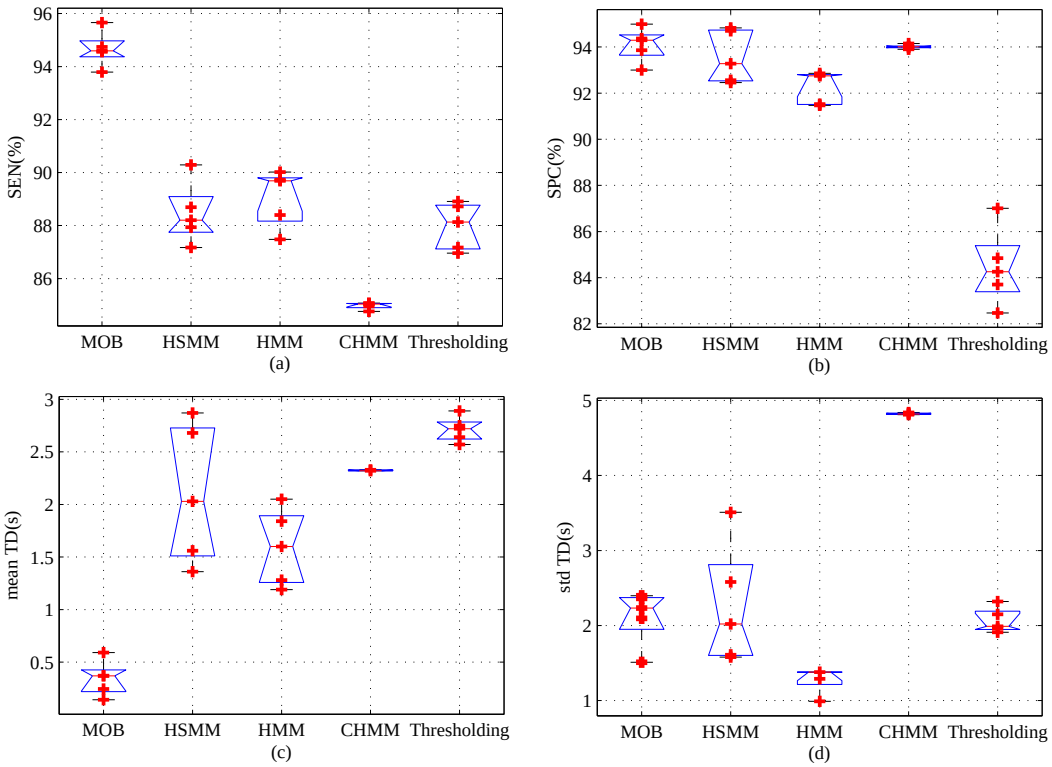

Figure 6.18: The comparison of wave-based MOB performance with existing methods showing the lower quartile, median and upper quartile values. The raw data are illustrated by red crosses. a: sensitivity. b: specificity. c: mean of time delay d: std of time delay.

comparing to others.

Table 6.8: Comparison of the bests results of Markovian and Switching approaches in AB detection.

method	SEN	SPC	mean TD (s)	std TD(s)
CHMM	95.74±0.82	91.88±0.31	-0.59±0.21	2.79±0.06
CHSMM	94.87±3.18	96.52±0.64	0.73±0.60	2.45±0.69
MOB	94.74±0.69	94.17±0.87	0.35±0.13	2.12±0.29
R-based	90.35±5.46	88.59±11.28	1.80±0.89	1.94±0.43

Chapter 7

Discussions, Conclusions and Perspectives

In this thesis, we proposed several methods for on-line detection of event of interest. These methods include two Markovian models (CHSMM and CHMM) evaluated in detection of a specific disturbance in simulated data and AB episodes in real data set consisting of three different feature signals extracted from raw ECG and two Switching algorithms to detect AB episodes from raw ECG and RR feature signal. According to existing methods proposed in our laboratory, [Altuve et al., 2015], HSMM was previously introduced as a powerful model for learning the dynamics of our real data. This thesis suggests to involve the concept of coupling in HSMM framework. However, we could not maintain a meaningful binding between them unless the coupling is firstly studied in HMM context. Moreover, CHMM and CHSMM as two generalizations of HMM were previously proposed in [Brand, 1997, Rezek et al., 2000] and [Natarajan and Nevatia, 2007a] respectively. Hence, due to the lack of simplicity and other problems like higher orders of complexity and overflow issue in these models, we began with CHMM to propose a novel coupling framework based on HMM. Then, it was generalized in CHSMM as a combination of HSMM and coupling. In order to reach our goals, we have used several assumptions to simplify the solutions and achieve FB recursions for each model. Moreover, we demonstrate a re-estimation approach based on ML criterion to train the parameters of each model. In this chapter, we first discuss the results given in chapter 6 and make conclusions. We then present some ideas for future works.

7.1 Discussions and Conclusions on Markovian approaches

7.1.1 Experiments for 3-class classification over simulated database

In part 6.2.1, we first examined the proposed CHMM and CHSMM frameworks in a classification scenario in the context of disturbance detection in simulated database. The disturbances are generated by two distinct dynamics caused by changing the statistical probabilities of a parameter in Fitzhugh-Nagumo model. Since, the patterns of dynamics are not completely different, their separation can be a complex issue. How-

ever, experimental results showed the efficiency of the proposed CHMM in solving the separation of dynamics from each other. Whereas, it fails to detect disturbances from rest condition. In contrast with CHMM framework, the proposed CHSMM is completely successful in detection of disturbances, although it underperforms CHMM in separation of the dynamic separation.

7.1.2 Experiments for on-line detection of a dynamic over simulated database

In the second part of our study, we examined the efficiency of the proposed frameworks for on-line detection of a typical disturbance with specific dynamic in simulated data. To this end, we compared the training procedure of the two proposed framework in terms of their state characteristics illustrated by Gaussian kernels. Accordingly, each state is related to an interval in the range of amplitude of the observations used in training dataset. Totally, in terms of sensitivity and specificity, the proposed CHMM and CHSMM surpass HMM, HSMM and CHMM in disturbance detection. However, comparing to each other, CHMM detects faster than CHSMM. After the proposed CHMM, the methods based on CHSMM, CHMM and multivariate HSMM produce similar results in terms of time delay. Surprisingly, HMM in both univariate and multivariate approaches shows the lowest time delay, whereas does not work precisely in detection issue. Although the results of CHSMM in terms of time delay are not the best, it has acceptable result in terms of specificity.

The first noticeable conclusion from this set of experiments is that coupling approach can enhance the degree of learning the dynamics characteristics. Secondly, they show that the HSMM based methods including CHSMM and univariate and multivariate HSMM frameworks can not detect disturbances as fast as the HMM based frameworks. But, in terms of sensitivity and specificity, they have acceptable results. Even it can be concluded that the characteristics of dynamics can be learn more accurately in semi-Markovian models.

Another remark that should be considered is the levels of total likelihoods obtained by comparing models corresponding to dynamics a_1 and a_2. In a scenario, the alteration of these likelihoods for two disturbances generated by the two different dynamics are compared to each other using proposed CHMM and CHSMM. For both of the frameworks, there exists highlighted contrast between the two total likelihoods. Hence, the disturbance separation can be performed perfectly by applying suitable threshold. It is clear that the total likelihoods computed by comparing models for dynamic a_1 and rest show a significant rises around disturbance incident. Consequently, by using a simple thresholding, the disturbance detection issue can be solved appropriately. It should also be noted that all the above-mentioned results were obtained for optimal number of state which are obtained by BIC method.

7.1.3 Experiments for on-line detection of AB episodes in real data

In the third part of our study, we examined the efficiency of the proposed CHMM and CHSMM frameworks by establishing coupling between RR and RAMP feature signals. However, the results was poor. This was expected since the effect of apnea appears in RAMP signal right after the apnea incident, whereas in RR signal, apnea impacts are observable a few seconds later. We verified this claim by using the manually tuned synchronization time delay to temporally align the alterations induced by apnea in RAMP signal. It is observed that the performance of these methods considerably increases by using this information.

The results obtained for real data were nearly similar to simulated results. The results are prepared for three combination of the three feature signals: 1- RR-RAMP 2- RR-QRSd and 3- RR-QRSd-RAMP. For those combination including RAMP, the suitable values of synchronization time delay are considered. Using combination 1, the CHSMM and HSMM methods surpass the other methods for even lower number of states. However, CHSMM gives better results than even HSMM especially in terms of time delay, specificity and distance to PD. The reported time delay in HMM, HSMM and proposed CHMM are positively low, but not better than the time delay of CHSMM which has a negative value. The negative time delay coming along with PW proves that most of the AB episodes are predicted as the detected onsets are prior to the annotated onsets. Note that the proposed CHMM performs slightly better than the CHMM proposed by Rezek in terms of the metrics, however it underperforms comparing to HSMM. The inspiring idea of CHSMM framework proposal arose from this result, in order to combine the power of HSMM and CHMM to achieve better detection performance.

For the second combination, the coupling frameworks surpass the other methods in terms of sensitivity, specificity and distance to PD. After the coupling frameworks, HSMM gives better results than HMM. Whereas, the time delay reported for HMM is the best which is marginally the same as time delay of CHSMM. The values reported for PW declare that by using such combination, the prediction of AB is likely to be more difficult comparing to the first combination. These experimental results prove that there exists a significant coupling between RR and QRSd features, but the information about the characteristics of AB episodes prior to their annotated onsets is low. Accordingly, all the time delays are higher than those achieved in first combination.

For the third combination including all the feature signals, the proposed coupling frameworks surpass the other methods especially in terms of distance to PD. However, the CHSMM demonstrates superior performance in specificity and sensitivity. Whereas, similar to the results of simulated data, the proposed CHMM is the fastest method with a negative time delay.

In summary, CHSMM has the lowest distance to PD in all the combinations. Among the CHSMM results, the best PD is achieved using all the feature signals as 3-channel observation. Moreover, CHSMM appropriately has the lowest time delay while using first combination. PW and AUC has the lowest and highest values respectively in all combinations using CHSMM framework. The recursive structure of the frameworks based on the nature of standard HMM and sample by sample windowing process makes

the algorithm computationally tractable, and suitable for on-line applications. Moreover, the computational cost of our proposed FB algorithm in CHSMM is $O(TM^C)$ which is comparable with the CHSMM proposed by Natarajan [Natarajan and Nevatia, 2007b] ($O(C^2MT)$). In CHMM, the computational cost is $O(TM^C)$ comparing to the CHMM proposed by Rezek ($O(TM^{2C})$). These findings show that our suggestion to use the conditional definitions for FB parameters and establishment of their corresponding recursions assuming the simplification in Eq. 3.62 for the coupling of observation in HMM context, results in lower complexity and more speed in implementation. Hence, such a complex issue as coupling is simplified logically.

7.2 Discussion and Conclusion on Switching approaches

In this work, the main goal is to use SKF for AB detection. Therefore, we proposed two SKF-based models where in one approach ECG signal is utilized as an observation while, in other approach, RR series is employed. In EKF framework which is previously used in different applications like denoising or compression, McSharry's model is used in order to model ECG signal as a combination of finite number of Gaussian kernels and establish a set of state equations of KF. The performance of this model is eligible if the characteristics of ECG has slight alterations. Hence, when an arrhythmia like AB happens, the parameters of relevant model is totally different from normal situation. Therefore, if we intend to detect AB, it is convenient to integrate two EKF models for normal and AB in the form of a SKF and monitor the probability of discrete state variable corresponding to switch (K_k^i) as a detector.

On the other hand, we suggest a R-based method based on SKF structure proposed in [Wu et al., 2004] which includes two independent linear AR models. The models provide a state equation whose states are the RR signal corresponding to normal and AB. The observation is the calculated RR which can be considered as a noisy version of state variables. The switch indicates to the state variable of one of the models to be used in observation equation.

The SKF algorithm is comprised of several KF, each of which includes the model parameters related to a dynamic. Each filter is able to adapt with different morphologies and temporal nonstationarities since the variance of the observation noise defines the degree of reliability of an observation. This determines a bound of how much tracking the observation interferes in estimation of state variables. However, in case of powerful noises in observation or dynamic alteration, the information that KF obtains from its model is expected to be poor and it decreases the probability of observation to be generated by such model. SKF quantifies this probability and refers to it as the probability of a switch conditions. Compared to the R-based detection scheme, the wave-based model provides a better detection performance, especially in finding bradycardia events without missing them.

The application of HMM-based approaches had been previously reported [Altuve et al., 2011b, Masoudi et al., 2013, Altuve et al., 2015]. The major drawback of such methods is that each beat is represented by finite number of features which results in methods using only the inter-beat information like our R-based method. However,

the wave-based technique does not depend on the specific features and instead uses a dynamic state space representation for adaptive signal tracking sample by sample. Moreover, unlike the EKF that depends on covariance of measurement noise in the case of nonstationarities, the wave-based method evaluates the performance of different EKF and uses the most probable one to match the state variables to the observation. Another point of interest is time delay improvement for wave-based model, that ensures an early detection of bradycardia which is an important factor for the algorithm of a monitoring device of NICU. A simple thresholding approach spends a couple of seconds to detect a rise in RR signal and it mainly uses the amplitude, yet the proposed methods are expected to learn the dynamic and to detect the change in observation earlier with better precisian. The results show that our proposed wave-based method is capable of detecting AB faster than the other methods like [Masoudi et al., 2013, Altuve et al., 2015].

7.3 Discussions and Conclusions on comparing Markovian and Switching approaches

Comparing all the best results achieved by the proposed methods in Markovian and Switching approaches, we can make interesting conclusions. CHSMM approach shows superior performance in precision of detection. After CHSMM, wave-based method comparing to CHMM achieves better results. Whereas, CHMM is able to detect (predict) faster than the others methods. After CHMM, wave-based method has the lowest value for TD. Although, R-based method does not outperform comparing to others, however, the obtained results are considerable according to its simple structure. Totally, Markovian approach is more successful than the Switching approach.

7.4 Future works

As a future work, we can study the clinical usage of the proposed methods on a large number of cases, for AB detection. We can also study more complicated scenarios in subjects with more complications than just apnea. Moreover, since the proposed CHMM and CHSMM approaches are completely generic, we plan to apply them to other clinical applications such as adult apnea event detection, using specific datasets and coupling different cardiorespiratory signals.

As another extension to this work, we can improve properties of resting parameter in CHSMM, for instance defining one of the famous statistical function as its pdf. Moreover, such study can be performed on the pdf of states. Instead of Gaussian kernels, we can employ more than one kernel or use any other pdfs.

In this thesis, the proposed frameworks are evaluated for AB detection. It seems interesting to employ them in any other fields of researches like studying the connectivity or coupling between different features. For example, the proposed frameworks are ready to fit to the dynamic spacial connectivity of the brain signals.

Future works in Switching approaches include incorporating AR parameter estimation methods instead of ML approach for training procedure in both proposed approaches. In wave-based, we can use other dynamical models for ECG waveforms [Ayatollahi et al., 2005, Das and Maharatna, 2013]. Other feature signals (like duration and the amplitude of QRS complex) rather than RR signal can be extracted and then, an SKF model similar to R-based approach can be used for processing and AB detection. Furthermore, we can also use multidimensional observation including RR signal and other features. Higher orders for AR models may result in better detection performance although they make the algorithm more complex.

Appendix A

A.1 An assumption in Markovian approach

For better understanding, consider Fig. 3.2, which depicts a minor loop of the two coupled chains at times $t-1$, t and $t+1$. We simplify the predicted joint probability of states $P(v_t^1(m_1), v_t^2(m_2)|o_{1:t-1})$ by $P(v_t^1(m_1)|o_{1:t-1}) \times P(v_t^2(m_2)|o_{1:t-1})$ i.e. $v_t^1(m_1)$ and $v_t^2(m_2)$ are conditionally independent. We generalize it to more than 2 channels as follows:

$$P(v_t^1(m_1), \cdots, v_t^C(m_C)|o_{1:t-1}) = \prod_{c=1}^{C} P(v_t^c(m_c)|o_{1:t-1}) \tag{A.1}$$

This simplification helps us to prove other statements like Eq. (3.54) and Eq. (3.58). The following proofs are presented in two channels for simplicity.

A.2 Joint probability estimation of channels observations

During the parameters definition and for obtaining more compact and simpler relation for forward and backward parameters, we need to estimate the joint probability of observations $P(o_t^1, o_t^2|o_{1:t-1})$ (Eq. (3.58)). One estimation is to use mean-field assumption of joint probability of the two observations, i.e. the joint probability is calculated by the product of the probability of each variable, which means they are conditionally independent [Ye et al., 2011].

$$P(o_t^1, o_t^2|o_{1:t-1}) = P(o_t^1|o_{1:t-1}) \times P(o_t^2|o_{1:t-1}) \tag{A.2}$$

Other estimations are uniform and weighted. In uniform case, the joint probability $P(o_t^1, o_t^2|o_{1:t-1})$ is calculated by the summation of the conditional probability of each observation, $P(o_t^1|o_{1:t-1}) + P(o_t^2|o_{1:t-1})$, which seems to be incorrect from the weighted point of view as follows.

Weighted case is a more general form. The joint probability of the observations of channels can be expressed as follows [Ye et al., 2011]:

$$P(o_t^1, o_t^2|o_{1:t-1}) = P(o_t^1|o_t^2, o_{1:t-1})P(o_t^1|o_{1:t-1})$$
$$P(o_t^1, o_t^2|o_{1:t-1}) = P(o_t^2|o_t^1, o_{1:t-1})P(o_t^2|o_{1:t-1})$$

so, it is correct to have

$$\begin{aligned} P(o_t^1, o_t^2|o_{1:t-1}) &= [0.5P(o_t^1|o_t^2, o_{1:t-1}]P(o_t^1|o_{1:t-1}) \\ &+[0.5P(o_t^2|o_t^1, o_{1:t-1})]P(o_t^2|o_{1:t-1}) \\ &= w_1 P(o_t^1|o_{1:t-1}) + w_2 P(o_t^2|o_{1:t-1}) \end{aligned}$$

As we know, it is difficult to estimate $P(o_t^1|o_t^2, o_{1:t-1})$ or $P(o_t^2|o_t^1, o_{1:t-1})$. Also, mean-field assumption for the observations seems to be correct, since although they are correlated but they give us no information about each other given the previous observations and it is widely used in other works ([Brand, 1997] and [Zhong and Ghosh, 2001]). This is what we have used in definition of $\tilde{b}$.

Moreover, by Eq. (3.62), we can prove the mean-field assumption as follows:

$$\begin{aligned} P(o_t|o_{1:t-1}) &= \sum_{m_1}\sum_{m_2} P(o_t, v_t^1(m_1), v_t^2(m_2)|o_{1:t-1}) \\ &= \sum_{m_1}\sum_{m_2} P(o_t|v_t^1(m_1), v_t^2(m_2), o_{1:t-1}) \\ &\quad \times P(v_t^1(m_1), v_t^2(m_2)|o_{1:t-1}) \end{aligned} \tag{A.3}$$

then we have:

$$\begin{aligned} P(o_t|o_{1:t-1}) &= \sum_{m_1}\sum_{m_2}\prod_{c=1}^{2} P(o_t^c|v_t^c(m_c), o_{1:t-1}) \\ &\quad \times \prod_{c=1}^{2} P(v_t^c(m_c)|o_{1:t-1}) \\ &= \prod_{c=1}^{2}\sum_{m_c} P(o_t^c|v_t^c(m_c), o_{1:t-1}) \\ &\quad \times P(v_t^c(m_c)|o_{1:t-1}) \\ &= \prod_{c=1}^{2}\sum_{m_c} P(o_t^c, v_t^c(m_c)|o_{1:t-1}) \\ &= \prod_{c=1}^{2} P(o_t^c|o_{1:t-1}) \end{aligned} \tag{A.4}$$

where it can be generalized for $C > 2$.

A.3 The filtered joint probability of forward parameters

In this part a prove for Eq. (3.54), the filtered joint probability of states given the observations, is presented by using Eq. (A.1) and Eq. (A.2). Therefore, we start with

adding the observations at time t to the states:

$$\begin{aligned} P(v_t^1, v_t^2|o_{1:t}) &= \frac{P(v_t^1, v_t^2, o_t|o_{1:t-1})}{P(o_t|o_{1:t-1})} \\ &= \frac{P(o_t^1|v_t^1)P(o_t^2|v_t^2)P(v_t^1, v_t^2|o_{1:t-1})}{P(o_t|o_{1:t-1})} \end{aligned} \tag{A.5}$$

On the other hand, we can write:

$$\begin{aligned} P(v_t^1|o_{1:t})P(v_t^2|o_{1:t}) &= \frac{P(v_t^1, o_t|o_{1:t-1})P(v_t^2, o_t|o_{1:t-1})}{P(o_t|o_{1:t-1})^2} \\ &= \frac{P(o_t^1|v_t^1)P(v_t^1|o_{1:t-1})\sum_{v_t^2} P(o_t^2|v_t^2)P(v_t^2|o_{1:t-1})}{P(o_t|o_{1:t-1})} \\ &\times \frac{P(o_t^2|v_t^2)P(v_t^2|o_{1:t-1})\sum_{v_t^1} P(o_t^1|v_t^1)P(v_t^1|o_{1:t-1})}{P(o_t|o_{1:t-1})} \end{aligned} \tag{A.6}$$

with some simple manipulation, we can derive the term $P(v_t^1, v_t^2|o_{1:t})$ and simplify the statement above as follows:

$$\begin{aligned} P(v_t^1|o_{1:t})P(v_t^2|o_{1:t}) &= P(v_t^1, v_t^2|o_{1:t}) \\ &\times \frac{\sum_{v_t^1}\sum_{v_t^2} P(o_t^1, v_t^1|o_{1:t-1})P(o_t^2, v_t^2|o_{1:t-1})}{P(o_t|o_{1:t-1})} \\ &= P(v_t^1, v_t^2|o_{1:t}) \times \frac{P(o_t^1|o_{1:t-1})P(o_t^2|o_{1:t-1})}{P(o_t|o_{1:t-1})} \end{aligned} \tag{A.7}$$

where the second term is equal to 1 based on Eq. (A.2). Thus, we have:

$$P(v_t^1|o_{1:t})P(v_t^2|o_{1:t}) = P(v_t^1, v_t^2|o_{1:t}) \tag{A.8}$$

Appendix B

B.1 Functions of Switching appraoches

FilteringKF: evolution of the KF from $k-1$ to k.

Require: x^j_{k-1}, P^j_{k-1},M_i,R_i

Ensure: x^{ji}_k, P^{ji}_k,

$i, j = 1 : N$

1: $x^{-ji}_k = A_i x^j_{k-1}$

2: $P^{-ji}_k = A_i P^j_{k-1} A_i^\intercal + \Gamma_i$

3: $e_k = y_k - M x^{-ji}_k$

4: $L^{ji}_k = \mathcal{N}(e_k; 0, M P^{-ji}_k M^\intercal + R)$

5: $B = (P^{-ji}_k M^\intercal)(M P^{-ji}_k M^\intercal + R)^{(-1)}$

6: $P^{ji}_k = (I - BM) P^{-ji}_k$

7: $x^{ji}_k = x^{-ji}_k + B(y_k - M(x^{-ji}_k))$

---

where $(...)^- \equiv p(...|y_{1:k-1})$. L^{ji}_k has a Gaussian pdf as $\mathcal{N}(M x^{-ji}_k, M P^{-ji}_k M^\intercal + R)$ which is usually substituted by its zero-mean version, $\mathcal{N}(y_k - M x^{-ji}_k; 0, M P^{-ji}_k M^\intercal + R)$.

FilteringSEKF: evolution of the EKF from $k-1$ to k.

Require: x^j_{k-1}, P^j_{k-1},M,R,G_k,F_k

Ensure: x^{ji}_k, P^{ji}_k,

$i, j = 1 : N$

1: $\omega^{-ji}_k = A_i \omega^j_{k-1}$

$\varphi^{-ji}_k = (\varphi^j_{k-1} + \delta \omega^j_{k-1}) mod 2\pi$

$z^{-ji}_k = -\sum_n \delta \frac{\alpha_n \omega^j_{k-1}}{(b_n)^2} \Delta\theta_n \exp(-\frac{\Delta(\theta_n)^2}{2(b_n)}) + z^j_{k-1}$

$x^{-ji}_k = [\omega^{-ji}_k \ \varphi^{-ji}_k \ z^{-ji}_k]$

2: $P^{-ji}_k = G_k P^j_{k-1} G_k^\intercal + F_k \Gamma_i F_k^\intercal$

3: $e_k = y_k - M x^{-ji}_k$

4: $L^{ji}_k = \mathcal{N}(e_k; 0, M P^{-ji}_k M^\intercal + R)$

5: $B = (P^{-ji}_k M^\intercal)(M P^{-ji}_k M^\intercal + R)^{(-1)}$

6: $P^{ji}_k = (I - BM) P^{-ji}_k$

7: $x^{ji}_k = x^{-ji}_k + B(y_k - M x^{-ji}_k)$

---

where $(...)^- \equiv p(...|y_{1:k-1})$.

StatesProbability

Require: L_k^{ji}, c_{ji}, K_{k-1}^{j},

Ensure: K_k^{ji}, K_k^{i}, $g_k^{j|i}$,

$i, j = 1 : N$

1: $K_k^{ji} = \frac{L_k^{ji} c_{ji} K_{k-1}^{j}}{\sum_i \sum_j L_k^{ji} c_{ji} K_{k-1}^{j}}$

2: $K_k^{i} = \sum_j K_k^{ji}$

3: $g_k^{j|i} = \frac{K_k^{ji}}{K_k^{i}}$

During the last stage, moment matching is performed for reduction of the mixture components to N Gaussians by the following relations

Collapsing: reducing the number of Gaussian kernels.

Require: x_k^{ji}, P_k^{ji},

Ensure: x_k^{i}, P_k^{i},

1: $x_k^{i} = \sum_j x_k^{ji} g_k^{j|i}$

2: $P_k^{i} = g_k^{j|i}(P_k^{ji} + (x_k^{ji} - x_k^{i})(x_k^{ji} - x_k^{i})^\intercal$

Appendix C

1. N. Montazeri, M. B. Shamsollahi, G. Carrault, and A. I. HernÃ¡ndez, "Paroxysmal atrial fibrillation prediction using Kalman Filter," in Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies, 2011, p. 89.

2. S. Masoudi, N. Montazeri, M. Shamsollahi, D. Ge, A. Beuchee, P. Pladys, and A. Hernandez, "Early detection of apnea-bradycardia episodes in preterm infants based on coupled hidden Markov model," in Signal Processing and Information Technology (ISSPIT), 2013 IEEE International Symposium on, 2013, pp. 000243-000248.

3. N. Montazeri, S. Masoudi, M. B. Shamsollahi, D. Ge, and A. I. Hernandez, ?Coupled hidden Markov model based me|thod for apnea bradycardia detection", IEEE Journal of Biomedical and Health Informatics (J-BHI), - accepted and published in early print DOI : 10.1109/JBHI.2015.2405075.

4. N. Montazeri, M. B. Shamsollahi, D. Ge, and A. I. Hernandez, "Switching Kalman filter based methods for apnea bradycardia detection from ECG signals", Physiological Measurement, accepted, in press 2015.

Bibliography

[Abu-Shaweesh and Martin, 2008] Abu-Shaweesh, J. M. and Martin, R. J. (2008). Neonatal apnea: what's new? *Pediatric pulmonology*, 43(10):937 944.

[Aguirre et al., 1999] Aguirre, L. A., Barros, V. C., and Souza, Á. V. (1999). Nonlinear multivariable modeling and analysis of sleep apnea time series. *Computers in Biology and Medicine*, 29(3):207 228.

[Akhbari et al., 2012] Akhbari, M., Shamsollahi, M. B., Jutten, C., and Coppa, B. (2012). Ecg denoising using angular velocity as a state and an observation in an extended kalman filter framework. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 2897 2900. IEEE.

[Altuve, 2011] Altuve, M. (2011). *Détection multivariée des épisodes d'apnée-bradycardie chez le prématuré par modéles semi-Markovien cachés.* PhD thesis, Université de Rennes 1.

[Altuve et al.,] Altuve, M., Carrault, G., Beuchée, A., Flamand, C., Pladys, P., and Hernández, A. I. Comparing hidden markov model and hidden semi-markov model based detectors of apnea-bradycardia episodes in preterm infants.

[Altuve et al., 2011a] Altuve, M., Carrault, G., BeuchÂ´ee, A., Pladys, P., and HernÂ´andez, A. I. (2011a). On-line apnea-bradycardia detection using hidden semi-markov models. In *33rd Annual International Conference of the IEEE EMBS*, Boston, Massachusetts USA.

[Altuve et al., 2015] Altuve, M., Carrault, G., Beuchée, A., Pladys, P., and Hernández, A. I. (2015). Online apnea bradycardia detection based on hidden semi-markov models. *Medical & biological engineering & computing*, 53(1):1 13.

[Altuve et al., 2009] Altuve, M., Carrault, G., Cruz, J., Beuchée, A., Pladys, P., and Hernandez, A. (2009). Analysis of the qrs complex for apnea-bradycardia characterization in preterm infants. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 946 949. IEEE.

[Altuve et al., 2011b] Altuve, M., Carrault, G., Cruz, J., Beuchée, A., Pladys, P., and Hernandez, A. (2011b). Multivariate ecg analysis for apnoea? bradycardia detection and characterisation in preterm infants. *International Journal of Biomedical Engineering and Technology*, 5(2):247 265.

[Ayatollahi et al., 2005] Ayatollahi, A., Dabanloo, N. J., McLernon, D., Majd, V. J., and Zhang, H. (2005). A comprehensive model using modified zeeman model for generating ecg signals. *Iranian Journal of Electrical & Electronic Engineering*, 1(2):88.

[Baum et al., 1970] Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, pages 164 171.

[Belal et al., 2011] Belal, S. Y., Emmerson, A. J., and Beatty, P. C. W. (2011). Automatic detection of apnoea of prematurity. *Physiological measurement*, 32(5):523.

[Brand, 1997] Brand, M. (1997). Coupled hidden markov models for modeling interacting processes.

[Brand et al., 1997a] Brand, M., Oliver, N., and Pentland, A. (1997a). Coupled hidden markov models for complex action recognition. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 994 999. IEEE.

[Brand et al., 1997b] Brand, M., Oliver, N., and Pentland, A. (1997b). Coupled hidden markov models for complex action recognition. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 994 999. IEEE.

[Brewer et al., 2006] Brewer, N., Liu, N., De Vel, O., and Caelli, T. (2006). Using coupled hidden markov models to model suspect interactions in digital forensic analysis. In *Integrating AI and Data Mining, 2006. AIDM'06. International Workshop on*, pages 58 64. IEEE.

[Camargo et al., 2014] Camargo, V. C., Honorato da Silva, S., de Amorim, M. F., and Nohama, P. (2014). Instrumentation for the detection and interruption of apnea episodes for premature newborn. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 2127 2130. IEEE.

[Cetin et al., 2007] Cetin, O., Ostendorf, M., and Bernard, G. D. (2007). Multirate coupled hidden markov models and their application to machining tool-wear classification. *Signal Processing, IEEE Transactions on*, 55(6):2885 2896.

[Chair and Varshney, 1986] Chair, Z. and Varshney, P. (1986). Optimal data fusion in multiple sensor detection systems. *Aerospace and Electronic Systems, IEEE Transactions on*, 1:98 101.

[Chen, 2013] Chen, Y.-C. (2013). The effectiveness of sleeping time and episodes of apnea on preterm infants in the prone position. In *Sigma Theta Tau International's 24th International Nursing Research Congress*. STTI.

[Coast, 1993] Coast, D. A. (1993). Segmentation of high-resolution ecgs using hidden markov models. In *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, volume 1, pages 67 70. IEEE.

[Cruz et al., 2006] Cruz, J., Hernández, A., Wong, S., Carrault, G., and Beuchee, A. (2006). Algorithm fusion for the early detection of apnea-bradycardia in preterm infants. In *Computers in Cardiology, 2006*, pages 473 476. IEEE.

[Dageville, 2011] Dageville, C. (2011). Le debut de la vie d'un grand premature.

[Das and Maharatna, 2013] Das, S. and Maharatna, K. (2013). Fractional dynamical model for the generation of ecg like signals from filtered coupled van-der pol oscillators. *Computer methods and programs in biomedicine*, 112(3):490 507.

[de Chazal et al., 2004] de Chazal, P., Penzel, T., and Heneghan, C. (2004). Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram. *Physiological measurement*, 25(4):967.

[Di Fiore et al., 2001] Di Fiore, J. M., Arko, M. K., Miller, M. J., Krauss, A., Betkerur, A., Zadell, A., Kenney, S. R., and Martin, R. J. (2001). Cardiorespiratory events in preterm infants referred for apnea monitoring studies. *Pediatrics*, 108(6):1304 1308.

[Dorostkar et al., 2005] Dorostkar, P. C., Arko, M. K., Baird, T. M., Rodriguez, S., and Martin, R. J. (2005). Asystole and severe bradycardia in preterm infants. *Neonatology*, 88(4):299 305.

[Dumont, 2008] Dumont, J. (2008). *Fouille de dynamiques multivariées, application ï¿½ des données temporelles en cardiologie*. PhD thesis, Université de Rennes 1.

[Dumont et al., 2008] Dumont, J., Hernández, A. I., Fleureau, J., and Carrault, G. (2008). Modelling temporal evolution of cardiac electrophysiological features using hidden semi-markov models. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 165 168. IEEE.

[Fauci et al., 2008] Fauci, A., Braunwald, E., Kasper, D., Hauser, S., Longo, D., Jameson, J., and Loscalzo, J. (2008). *Harrisons Principles of Internal Medicine Self-Assessment and Board Review 17th Edition*. McGraw Hill Professional.

[Felblinger et al., 1999] Felblinger, J., Jung, B., Slotboom, J., Boesch, C., and Kreis, R. (1999). Methods and reproducibility of cardiac/respiratory double-triggered 1h-mr spectroscopy of the human heart. *Magnetic resonance in medicine*, 42(5):903 910.

[Ferguson, 1980] Ferguson, J. D. (1980). Variable duration models for speech. In *Proc. Symposium on the application of hidden Markov models to text and speech*, pages 143 179.

[Finer et al., 2006] Finer, N. N., Higgins, R., Kattwinkel, J., and Martin, R. J. (2006). Summary proceedings from the apnea-of-prematurity group. *Pediatrics*, 117(Supplement 1):S47 S51.

[FitzHugh, 1961] FitzHugh, R. (1961). Impulses and physiological states in theoretical models of nerve membrane. *Biophysical journal*, 1(6):445.

[Fu et al., 2003] Fu, T., Liu, X. X., Liang, L. H., Pi, X., and Nefian, A. V. (2003). Audio-visual speaker identification using coupled hidden markov models. In *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, volume 3, pages III 29. IEEE.

[Gai et al., 2007] Gai, J., Li, Y., and Stevenson, R. L. (2007). Coupled hidden markov models for robust eo/ir target tracking. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 1, pages I 41. IEEE.

[Ge et al., 2013] Ge, D., Beuchée, A., Carrault, G., Pladys, P., and Hernández, A. (2013). Online apnea-bradycardia detection using recursive order estimation for auto-regressive models. In *Computing in Cardiology Conference (CinC), 2013*, pages 1247 1250. IEEE.

[Ge et al., 2014] Ge, D., Carrault, G., and Hernández, A. (2014). Online bayesian apnea-bradycardia detection using auto-regressive models. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 4428 4432. IEEE.

[Gerhardt and Bancalari, 1984] Gerhardt, T. and Bancalari, E. (1984). Apnea of prematurity: I. lung function and regulation of breathing. *Pediatrics*, 74(1):58 62.

[Ghahramani, 1998] Ghahramani, Z. (1998). Learning dynamic bayesian networks. In *Adaptive processing of sequences and data structures*, pages 168 197. Springer.

[Guo et al., 2012] Guo, P., Miao, Z., Zhang, X.-P., Shen, Y., and Wang, S. (2012). Coupled observation decomposed hidden markov model for multiperson activity recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 22(9):1306 1320.

[Hall, 2010] Hall, J. E. (2010). *Guyton and Hall Textbook of Medical Physiology: Enhanced E-book*. Elsevier Health Sciences.

[Haskova et al., 2013a] Haskova, K., Javorka, K., Javorka, M., Matasova, K., and Zibolen, M. (2013a). Apnea in preterm newborns: determinants, pathophysiology, effects on cardiovascular parameters and treatment. *Acta Medica Martiniana*, 13(3):5 17.

[Haskova et al., 2013b] Haskova, K., Javorka, K., Javorka, M., Matasova, K., and Zibolen, M. (2013b). Apnea in preterm newborns: determinants, pathophysiology, effects on cardiovascular parameters and treatment. *Acta Medica Martiniana*, 13(3):5 17.

[Henderson-Smart and Steer, 2004] Henderson-Smart, D. J. and Steer, P. (2004). Doxapram treatment for apnea in preterm infants. *Cochrane Database Syst Rev*, 4.

[Hernández et al., 2012] Hernández, A., Dumont, J., Altuve, M., Beuchée, A., and Carrault, G. (2012). Evolutionary optimization of ecg feature extraction methods: Applications to the monitoring of adult myocardial ischemia and neonatal apnea bradycardia events. In *ECG Signal Processing, Classification and Interpretation*, pages 237 273. Springer.

[Janvier et al., 2004] Janvier, A., Khairy, M., Kokkotis, A., Cormier, C., Messmer, D., and Barrington, K. J. (2004). Apnea is associated with neurodevelopmental impairment in very low birth weight infants. *Journal of perinatology*, 24(12):763 768.

[Johansson et al., 1999] Johansson, A., Öberg, P. Å., and Sedin, G. (1999). Monitoring of heart and respiratory rates in newborn infants using a new photoplethysmographic technique. *Journal of clinical monitoring and computing*, 15(7-8):461 467.

[Klabunde, 2008] Klabunde, R. E. (2008). Cardiovascular physiology concepts. It is available on http://www.cvphysiology.com/Arrhythmias/A013c.htm.

[Kodali et al., 2010] Kodali, A., Pattipati, K., and Singh, S. (2010). A coupled factorial hidden markov model (cfhmm) for diagnosing coupled faults. In *Aerospace Conference, 2010 IEEE*, pages 1 11. IEEE.

[Kristjansson et al., 2000] Kristjansson, T. T., Frey, B. J., and Huang, T. S. (2000). Event-coupled hidden markov models. In *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on*, volume 1, pages 385 388. IEEE.

[Kurth et al., 1987] Kurth, C., Spitzer, A., Broennle, A., and Downes, J. (1987). Postoperative apnea in preterm infants. *Anesthesiology*, 66(4):483 488.

[Laugel et al., 2000] Laugel, V., Belladdale, J., Escande, B., Messer, J., and Simeoni, U. (2000). Apnées du nouveau-né prématuré. *Journal de pediatrie et de puericulture*, 13(2):67 71.

[Levinson, 1986] Levinson, S. E. (1986). Continuously variable duration hidden markov models for automatic speech recognition. *Computer Speech & Language*, 1(1):29 45.

[Lin et al., 2011] Lin, C., Bugallo, M., Mailhes, C., and Tourneret, J.-Y. (2011). Ecg denoising using a dynamical model and a marginalized particle filter. In *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*, pages 1679 1683. IEEE.

[Lin et al., 2012] Lin, J.-C., Wu, C.-H., and Wei, W.-L. (2012). Error weighted semi-coupled hidden markov model for audio-visual emotion recognition. *Multimedia, IEEE Transactions on*, 14(1):142 156.

[Lindberg et al., 1992] Lindberg, L.-G., Ugnell, H., and Öberg, P. (1992). Monitoring of respiratory and heart rates using a fibre-optic sensor. *Medical and Biological Engineering and Computing*, 30(5):533 537.

[Liporace, 1982] Liporace, L. (1982). Maximum likelihood estimation for multivariate observations of markov sources. *Information Theory, IEEE Transactions on*, 28(5):729 734.

[Manfredi et al., 2005] Manfredi, V., Mahadevan, S., and Kurose, J. F. (2005). Switching kalman filters for prediction and tracking in an adaptive meteorological sensing network. In *SECON*, pages 197 206.

[Manzionna and Di Mauro, 2014] Manzionna, M. and Di Mauro, A. (2014). Rooming-in organization to prevent neonatal mortality and morbidity in late preterm infants. *Italian Journal of Pediatrics*, 40(Suppl 2):A4.

[Marculescu et al., 1998] Marculescu, R., Marculescu, D., and Pedram, M. (1998). Probabilistic modeling of dependencies during switching activity analysis. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 17(2):73 83.

[Martin and Fabes, 2008] Martin, C. L. and Fabes, R. (2008). *Discovering child development*. Cengage Learning.

[Martin et al., 2010] Martin, J. A., Hamilton, B. E., Sutton, P. D., Ventura, S. J., Mathews, T., Kirmeyer, S., and Osterman, M. J. (2010). Births: final data for 2007. *National vital statistics reports*, 58(24):1 125.

[Martin et al., 2013] Martin, J. A., Hamilton, B. E., Ventura, S. J., Osterman, M. J., and Mathews, T. (2013). Births: final data for 2011. *National Vital Statistics Report*, 62(1):1 90.

[Masoudi et al., 2013] Masoudi, S., Montazeri, N., Shamsollahi, M., Ge, D., Beuchee, A., Pladys, P., and Hernandez, A. (2013). Early detection of apnea-bradycardia episodes in preterm infants based on coupled hidden markov model. In *Signal Processing and Information Technology (ISSPIT), 2013 IEEE International Symposium on*, pages 000243 000248. IEEE.

[Mathews and MacDorman, 2013] Mathews, T. and MacDorman, M. F. (2013). Infant mortality statistics from the 2009 period linked birth/infant death data set. *National vital statistics reports*, 61(8).

[McNames and Fraser, 2000] McNames, J. and Fraser, A. (2000). Obstructive sleep apnea classification based on spectrogram patterns in the electrocardiogram. In *Computers in Cardiology 2000*, pages 749 752. IEEE.

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

[Mendez et al., 2007] Mendez, M. O., Ruini, D. D., Villantieri, O. P., Matteucci, M., Penzel, T., Cerutti, S., and Bianchi, A. M. (2007). Detection of sleep apnea from surface ecg based on features extracted by an autoregressive model. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 6105 6108. IEEE.

[Millard, 1999] Millard, R. K. (1999). Inductive plethysmography components analysis and improved non-invasive postoperative apnoea monitoring. *Physiological measurement*, 20(2):175.

[Moody et al., 1985] Moody, G. B., Mark, R. G., Zoccola, A., and Mantero, S. (1985). Derivation of respiratory signals from multi-lead ecgs. *Computers in cardiology*, 12:113 116.

[Murphy, 1998] Murphy, K. P. (1998). Switching kalman filters. Technical report, Citeseer.

[Nakajima et al., 1996] Nakajima, K., Tamura, T., and Miike, H. (1996). Monitoring of heart and respiratory rates by photoplethysmography using a digital filtering technique. *Medical engineering & physics*, 18(5):365 372.

[Natarajan and Nevatia, 2007a] Natarajan, P. and Nevatia, R. (2007a). Coupled hidden semi markov models for activity recognition. In *Motion and Video Computing, 2007. WMVC'07. IEEE Workshop on*, pages 10 10. IEEE.

[Natarajan and Nevatia, 2007b] Natarajan, P. and Nevatia, R. (2007b). Coupled hidden semi markov models for activity recognition. In *Motion and Video Computing, 2007. WMVC'07. IEEE Workshop on*, pages 10 10. IEEE.

[Nefian et al., 2002] Nefian, A. V., Liang, L., Pi, X., Xiaoxiang, L., Mao, C., and Murphy, K. (2002). A coupled hmm for audio-visual speech recognition. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, volume 2, pages II 2013. IEEE.

[Nottingham, 2014] Nottingham, T. U. (2014). A beginners guide to normal heart function, sinus rhythm & common cardiac arrhythmias. It is available on http://www.nottingham.ac.uk.

[Pallas-Areny et al., 1989] Pallas-Areny, R., Colominas-Balague, J., and Rosell, F. J. (1989). The effect of respiration-induced heart movements on the ecg. *Biomedical Engineering, IEEE Transactions on*, 36(6):585 590.

[Pan and Tompkins, 1985] Pan, J. and Tompkins, W. J. (1985). A real-time qrs detection algorithm. *Biomedical Engineering, IEEE Transactions on*, 3:230 236.

[Pavlovic et al., 1999] Pavlovic, V., Rehg, J. M., Cham, T.-J., and Murphy, K. P. (1999). A dynamic bayesian network approach to figure tracking using learned dynamic models. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, pages 94 101. IEEE.

[Penzel et al., 2002] Penzel, T., McNames, J., De Chazal, P., Raymond, B., Murray, A., and Moody, G. (2002). Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings. *Medical and Biological Engineering and Computing*, 40(4):402 407.

[Pichler et al., 2003] Pichler, G., Urlesberger, B., and Müller, W. (2003). Impact of bradycardia on cerebral oxygenation and cerebral blood volume during apnoea in preterm infants. *Physiological measurement*, 24(3):671.

[Poets, 2010] Poets, C. F. (2010). Apnea of prematurity: What can observational studies tell us about pathophysiology? *Sleep medicine*, 11(7):701 707.

[Poets et al., 1993] Poets, C. F., Stebbens, V. A., Samuels, M. P., and Southall, D. P. (1993). The relationship between bradycardia, apnea, and hypoxemia in preterm infants. *Pediatric research*, 34(2):144 147.

[Porges, 1995] Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of our evolutionary heritage. a polyvagal theory. *Psychophysiology*, 32(4):301 318.

[Porges, 1996] Porges, S. W. (1996). Physiological regulation in high-risk infants: A model for assessment and potential intervention. *Development and Psychopathology*, 8(01):43 58.

[Portet et al., 2007] Portet, F., Gao, F., Hunter, J., and Sripada, S. (2007). Evaluation of on-line bradycardia boundary detectors from neonatal clinical data. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 3288 3291. IEEE.

[Pravisani et al., 2003] Pravisani, G., Beuchée, A., Mainardi, L., and Carrault, G. (2003). Short term prediction of severe bradycardia in premature newborns. In *Computers in Cardiology, 2003*, pages 725 728. IEEE.

[Rabiner, 1989a] Rabiner, L. (1989a). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257 286.

[Rabiner, 1989b] Rabiner, L. (1989b). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257 286.

[Rezek et al., 2000] Rezek, I., Sykacek, P., and Roberts, S. J. (2000). Learning interaction dynamics with coupled hidden markov models. *IEE Proceedings-Science, Measurement and Technology*, 147(6):345 350.

[Rigatto and Brady, 1972] Rigatto, H. and Brady, J. P. (1972). Periodic breathing and apnea in preterm infants. ii. hypoxia as a primary event. *Pediatrics*, 50(2):219 228.

[Sameni et al., 2007] Sameni, R., Shamsollahi, M. B., Jutten, C., and Clifford, G. D. (2007). A nonlinear bayesian filtering framework for ECG denoising. *IEEE Transaction on Biomedical Engineering*, 54(12):2172 2185.

[Sayadi et al., 2010] Sayadi, O., Shamsollahi, M., and GD, C. (2010). Synthetic ecg generation and bayesian filtering using a gaussian wave-based dynamical model. *Physiological Measurement*, 31:1309 1329.

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

[Schwartz et al., 2002] Schwartz, P., Garson, A., Paul, T., Stramba-Badiale, M., Vetter, V., Villain, E., and Wren, C. (2002). Guidelines for the interpretation of the neonatal electrocardiogram. *European heart journal*, 23(17):1329 1344.

[Taha et al., 1997] Taha, B., Dempsey, J., Weber, S., Badr, M., Skatrud, J., Young, T., Jacques, A., and Seow, K. (1997). Automated detection and classification of sleep-disordered breathing from conventional polysomnography data. *Sleep*, 20(11):991 1001.

[THI, 2014] THI (2014). Heart anatomy. It is available on http://www.texasheart.org/HIC/Anatomy/anatomy2.cfm.

[Travaglini et al., 1998] Travaglini, A., Lamberti, C., DeBie, J., and Ferri, M. (1998). Respiratory signal derived from eight-lead ecg. In *Computers in Cardiology 1998*, pages 65 68. IEEE.

[Urlesberger et al., 1999] Urlesberger, B., Kaspirek, A., Pichler, C., and Müller, W. (1999). Apnoea of prematurity and changes in cerebral oxygenation and cerebral blood volume. *Neuropediatrics*, 30(01):29 33.

[Várady et al., 2000] Várady, P., Nagy, L., and Szilagyi, L. (2000). On-line detection of sleep apnea during critical care monitoring. In *Engineering in Medicine and Biology Society, 2000. Proceedings of the 22nd Annual International Conference of the IEEE*, volume 2, pages 1299 1301. IEEE.

[Veeraraghavan et al., 2006] Veeraraghavan, H., Papanikolopoulos, N., and Schrater, P. (2006). Deterministic sampling-based switching kalman filtering for vehicle tracking. In *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*, pages 1340 1345. IEEE.

[Veeraraghavan et al., 2005] Veeraraghavan, H., Schrater, P., and Papanikolopoulos, N. (2005). Switching kalman filter-based approach for tracking and event detection at traffic intersections. In *Intelligent Control, 2005. Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation*, pages 1167 1172. IEEE.

[Wikipedia, 2011] Wikipedia (2011). Heart, ecg. It is available on http://en.wikipedia.org/wiki/Heart/.

[Wu et al., 2003] Wu, W., Black, M. J., Mumford, D., Gao, Y., Bienenstock, E., and Donoghue, J. P. (2003). A switching kalman filter model for the motor cortical coding of hand motion. In *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, volume 3, pages 2083 2086. IEEE.

[Wu et al., 2004] Wu, W., Black, M. J., Mumford, D., Gao, Y., Bienenstock, E., and Donoghue, J. P. (2004). Modeling and decoding motor cortical activity using a switching kalman filter. *Biomedical Engineering, IEEE Transactions on*, 51(6):933 942.

[Xie and Liu, 2006] Xie, L. and Liu, Z.-Q. (2006). Speech animation using coupled hidden markov models. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 1, pages 1128 1131. IEEE.

[Ye et al., 2011] Ye, F., Yi, N., and Wang, Y. (2011). Em algorithm for training high-order hidden markov model with multiple observation sequences. *Journal of Information & Computational Science*, 8(10):1761 1777.

[Yu, 2010] Yu, S.-Z. (2010). Hidden semi-markov models. *Artificial Intelligence*, 174(2):215 243.

[Yu and Kobayashi, 2006] Yu, S.-Z. and Kobayashi, H. (2006). Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden markov model. *Signal Processing, IEEE Transactions on*, 54(5):1947 1951.

[Zeitlin, 2009] Zeitlin, J. (2009). Surveillance et évaluation de la santé périnatale en île-de-france á partir des certificats de santé. *Bulletin épidémiologique hebdomadaire*, pages 44 45.

[Zeitlin et al., 2013] Zeitlin, J., Szamotulska, K., Drewniak, N., Mohangoo, A., Chalmers, J., Sakkeus, L., Irgens, L., Gatt, M., Gissler, M., and Blondel, B. (2013). Preterm birth time trends in europe: a study of 19 countries. *BJOG: An International Journal of Obstetrics & Gynaecology*, 120(11):1356 1365.

[Zhao et al., 2011] Zhao, J., Gonzalez, F., and Mu, D. (2011). Apnea of prematurity: from cause to treatment. *European journal of pediatrics*, 170(9):1097 1105.

[Zhao et al., 1994] Zhao, L., Reisman, S., and Findley, T. (1994). *Derivation of respiration from electrocardiogram during heart rate variability studies*. IEEE.

[Zheng and Hasegawa-Johnson, 2003] Zheng, Y. and Hasegawa-Johnson, M. (2003). Acoustic segmentation using switching state kalman filter. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 1, pages I 752. IEEE.

[Zhong and Ghosh, 2001] Zhong, S. and Ghosh, J. (2001). A new formulation of coupled hidden markov models. *Department of Electrical and Computer Engineering, The University of Texas at Austin*.

[Zhou and Wong, 2007] Zhou, Q. and Wong, W. H. (2007). Coupling hidden markov models for the discovery of cis-regulatory modules in multiple species. *The Annals of Applied Statistics*, pages 36 65.

VU: VU:

Le Directeur de Thèse
ALFREDO I HERNÁNDEZ

Le Responsable de l'école doctorale

VU pour autorisation de soutenance
Rennes, le

Le Président de l'Université de Rennes 1

Guy CATHELINEAU

VU après soutenance pour autorisation de publication:

Le Président de Jury