BIOMEDICAL SIGNAL PROCESSING AND INFERENCE IN WEARABLE SENSING APPLICATIONS

By

SANDEEP GUTTA

Bachelor of Engineering in Electronics and
Communication Engineering
Andhra University
Visakhapatnam, Andhra Pradesh, India
2008

Master of Science in Electrical Engineering
Oklahoma State University
Stillwater, Oklahoma
2011

Submitted to the Faculty of the
Graduate College of
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2016

BIOMEDICAL SIGNAL PROCESSING AND INFERENCE IN WEARABLE SENSING APPLICATIONS

Dissertation Approved:

Dr. Qi Cheng

Dissertation Advisor

Dr. Weihua Sheng

Dr. Guoliang Fan

Dr. Bruce A. Benjamin

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Dr. Qi Cheng, for her excellent support, guidance, encouragement, and patience. It has been a true privilege to have had the chance to learn under her mentorship. Her dedication and enthusiasm for research have always been an inspiration to me.

I would also like to thank my committee members, Dr. Bruce Benjamin, Dr. Weihua Sheng, and Dr. Guoliang Fan, for their valuable teaching, advice, and support.

I am very grateful to the members of the Statistical Signal Processing Lab (SSPL) for their friendship and cooperation.

Last but not least, I thank my parents for their caring, love, encouragement and support, without which I could not have come this far. I also thank my family and friends, who were always there for me during this endeavor.

This research has been supported in part by the National Science Foundation (NSF) Grant CISE/IIS 1231671.

Acknowledgments reflect the views of the author and are not endorsed by committee members or Oklahoma State University.

Name: Sandeep Gutta

Date of Degree: December, 2016

Title of Study: BIOMEDICAL SIGNAL PROCESSING AND INFERENCE IN WEARABLE SENSING APPLICATIONS

Major Field: Electrical Engineering

Abstract:

With the increase in health care costs, there is a great need to develop a low-cost and efficient health care system with better accessibility. Recent advances in the fields of embedded sensing, mobile computing, and wireless communication have led to the development of several low-cost wearable health sensors that can noninvasively collect different physiological signals such as electrocardiogram (ECG), peripheral oxygen saturation (SpO_2), blood pressure, airflow, etc. With the availability of these wearable sensors and powerful smartphones, it is now possible to build a smart health monitoring system that can continuously monitor and track a person's health without the need of a hospital visit. In this research, we primarily focus on the signal processing and statistical inference aspects of the smart health system. First, we develop a biometric recognition system using the ECG signal, which is easily measured by several wearable devices. A multitask learning framework, in which the feature selection and classifier design are combined, is proposed to improve the overall learning efficiency. Experimental results on real ECG data show the effectiveness of the proposed method over other approaches. In the next part, we focus on the mathematical modeling of the local control mechanism of the cardiorespiratory system. We use a nonlinear model of the cardiorespiratory system with heart rate and ventilation rate as the control signals. An iterative algorithm is proposed to calculate the optimal control signals. In the final part, we focus on the detection of a chronic respiratory sleep disorder, sleep apnea, using measurement signals from wearable sensors. A new framework combining multiple sensor measurement data with the cardiorespiratory system model information is proposed. Experimental results on both synthetic and real data show the effectiveness of the proposed framework. Comparisons with purely data-driven apnea detection methods demonstrate the advantage of combining the sensor measurement data with cardiorespiratory model information.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Rising health care cost is an important economic issue in many countries around the world. For example, the United States health care spending reached $3 trillion in 2014, which is 17.5% of the nation's Gross Domestic Product [14]. A significant portion of the health care costs does not lead to better health care [15]. There is a great need to develop a better health care system at lower cost that is accessible to more people worldwide.

According to the United States National Academy of Medicine, a health care system should be safe, effective, patient-centered, timely, efficient, and equitable [16]. The National Academy of Medicine proposed the following ten general principles to help redesign the health care system [16]:

- Health care is available to patients whenever they need it and in many forms such as by telephone, over the Internet, etc. in addition to in-person visits.
- Health care is customized according to the patient needs and preferences.
- Patients should have all the necessary information and opportunity to make their health care decisions. The system should encourage shared decision making.
- Patients should have unrestricted access to their medical information and to the clinical knowledge.
- The system should have evidence-based decision making. Health care based

on the best available scientific knowledge should be provided and it should not vary from one clinician to another.

- The system should be safe and error-free.
- The system should be transparent and should enable patients or their families to make informed decisions regarding the treatments.
- The system should be able to anticipate patient needs.
- The system should not waste patient's time and resources.
- Clinicians and medical institutions should have proper cooperation, collaboration, and coordination to provide the best health care.

The United States President's Council of Advisors on Science and Technology (PCAST) recommends the use of data science and systems engineering approaches to provide an affordable and high-quality health care to the people [15]. Systems engineering reduces waste and increases safety, reliability, and efficiency of the health care system.

1.1 Wearable Sensing and Smart Health Monitoring

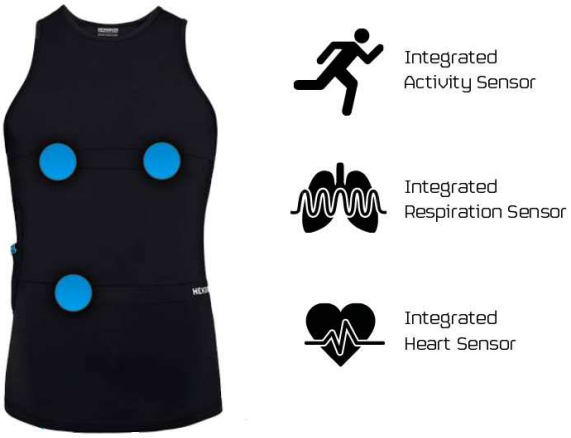
Recent advances in the fields of sensing and wireless communication have led to the development of nano- and micro-electromechanical systems (NEMS/MEMS) based wearable smart sensors. Advanced very-large-scale integration (VLSI) and ultra-large-scale integration (ULSI) technologies have led to the development of powerful system-on-a-chip (SoC) and system-in-a-package (SiP) integrated circuits, which are used in smartwatches, smartphones, and tablet computers. For example, the 64-bit Apple A10 Fusion SoC in iPhone 7 contains around 3.3 billion transistors in a die area of 125 mm^2 [17]. The success and rapid increase in the adoption of wearable sensors and smart mobile devices can be illustrated by the fact that over 86 billion ARM-

based processors have been shipped to date.[1] A vast majority of these processors are used in the SoCs and SiPs of various wearable sensors, embedded systems, and mobile devices [18].

Due to these technological advances, several low-cost wearable health devices are now available that can collect various physiological signals like electrocardiogram (ECG), heart rate, respiratory rate, peripheral oxygen saturation, etc. noninvasively. For example, some popular wearable health devices are shown in Figure 1.1, which typically cost few hundred dollars. The Hexoskin biometric smart shirt has a single-lead ECG sensor, thoracic and abdominal respiratory bands, three-axis accelerometer, and can measure motion activity, ECG, heart rate, breathing and ventilation rates. The Beddit sleep tracker has a piezoelectric force sensor, humidity sensor, temperature sensor, and measures heart rate and breathing signals using ballistocardiography (BCG) technique. The AliveCor Kardia Band for Apple Watch and Nymi Band measure the ECG signal. Moreover, Apple Inc. recently announced ResearchKit [19] and CareKit [20] software frameworks to help develop iPhone mobile applications capable of monitoring and tracking the user's health.

These wearable health devices have wireless connectivity and can record the corresponding physiological signals for several hours at a time. Rapid increase in the adoption of smartphones, smartwatches, and the availability of several low-cost wearable sensors has led to movements like mHealth, quantified self [21], and personal informatics [22], where a person is able to monitor different health metrics and self-track themselves throughout the day.

[1] https://www.arm.com/company



(a) Hexoskin smartshirt [23]

(b) AliveCor Kardia Band for Apple Watch [24]

(c) Beddit sleep tracker [25]

(d) Nymi Band [26]

(e) Omron Project Zero Wrist Blood Pressure Monitor (BP6000) [27]

Figure 1.1: Wearable health devices.

With the availability of these low-cost wearable health devices that can provide real-time physiological information, it is now possible to develop a noninvasive smart health monitoring system capable of providing accurate diagnostic information. Such a system can continuously monitor the user's health and provide timely vital medical alerts. These timely health alerts will greatly help physicians in providing an accurate diagnosis and prognosis. The monitoring system helps in providing the best possible proactive personal health care. Thus, the smart health monitoring system plays a significant role in providing better health care to more people at lower cost. The basic architecture of such a smart health monitoring system is shown in Figure 1.2.

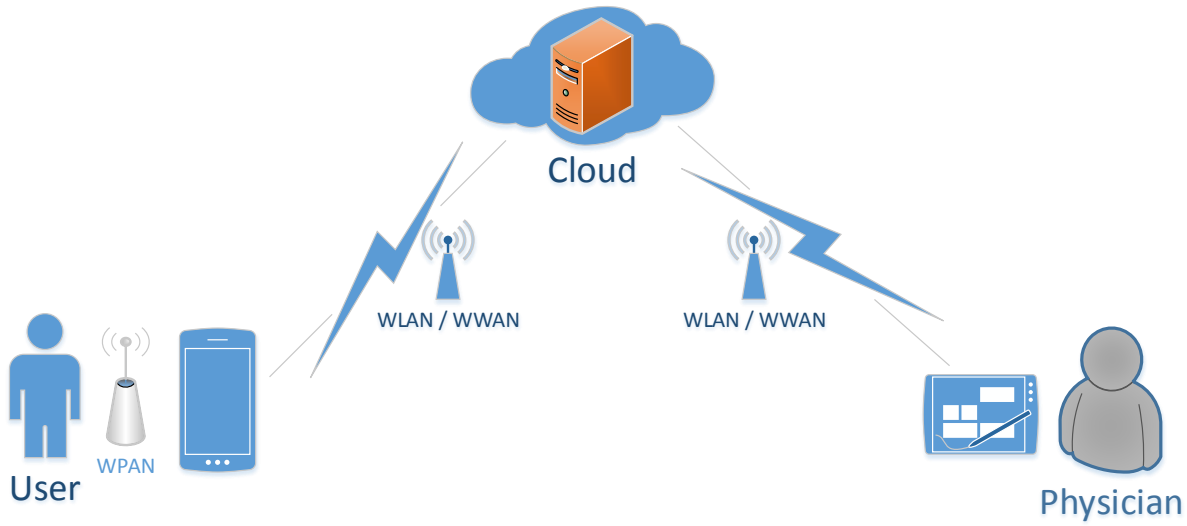


Figure 1.2: Smart health monitoring system.

1.1.1 Ambient Intelligence

Ambient Intelligence (AmI) is a new intelligent computing paradigm where smart environments (embedded with smart sensors) empower people inhabiting them. An AmI system is context aware, personalized, anticipatory, adaptive, ubiquitous, and transparent [28].

Wearable sensor-based intelligent computing was first used in activity recognition. In [29], Lee and Mase propose a method to determine a person's location and recognize activities such as walking, standing, and sitting using wearable accelerometers and gyroscopes. In [30], a network of three-axis accelerometers distributed on a person's body are used to calculate the orientation and movement of the corresponding body part. Lukowicz *et al.* [31] use body-worn microphones and accelerometers to recognize workshop activities. Rashidi *et al.* [32] propose an unsupervised method of discovering and tracking activities in a smart home environment. A sequential mining and clustering method is used to group the activities, and a hidden Markov model (HMM) is used to recognize them.

In [33], Gouaux *et al.* develop a portable and intelligent personal ECG monitor (PEM) for the early detection of cardiac ischemia and arrhythmia. Sung *et al.* [34] use a Gaussian mixture model (GMM) to detect hypothermia by classifying shivering through wearable accelerometers. In [35], Heilman and Porges studied the accuracy and precision of heart rate detection by the noninvasive physiological monitoring system LifeShirt (developed by VivoMetrics Inc.). Wearable sensor-based systems are also used to develop efficient assisted living technologies. In [36], Mubashir *et al.* survey different fall detection algorithms based on wearable, ambient, and vision sensors. Debes *et al.* [37] review different sensor technologies, signal processing and machine learning methods for monitoring the activities of daily living (ADLs). The monitoring of ADLs can be used to develop ambient assisted living (AAL) technologies to help elderly people.

Multiple wearable sensors measure different physiological and hence provide different viewpoints of the person's health. The main challenge here is the fusion of the multimodal sensor data to extract useful information. The information from multiple sensors needs to be efficiently processed and fused to provide an accurate representation of the person's health status. In some cases, the information from multiple

sensors can be redundant. It is also important to identify the redundancy between different sensors. Another main challenge is the fusion of the multisensor data with other health information of the person such as health records, previous diagnostic reports, medical lab test results, etc. to detect a certain disease or health disorder.

1.2 Research Objectives and Contributions

In this research, we mainly focus on the signal processing and statistical inference aspects of the smart health monitoring system. Since ECG signal is widely popular and is collected by many wearable health devices, we first find an interesting application of the ECG signal for biometric recognition. Biometric recognition systems based on traditional traits such as fingerprint, face, etc. can be easily attacked by duplicating the corresponding traits. However, it is very hard to attack an ECG-based biometric recognition system. So instead of relying on the traditional biometric recognition methods, an efficient biometric recognition system using the ECG signal is developed. The main challenge here is extracting the relevant features from the ECG signal. We propose a multitask learning framework, in which the feature selection and classifier design are combined, to improve the overall learning efficiency. The proposed recognition framework can easily deal with high-dimensional and nonlinearly separable data.

Mathematical modeling is a powerful way to study the underlying physiological systems when direct interaction with the system is not possible. Mathematical models try to capture and quantify the complex interactions between different physiological systems processes. We focus on modeling the control mechanism of the cardiorespiratory system. Control mechanisms play a vital role in effectively regulating the cardiorespiratory system. We specifically develop a model of the local control mechanism in which the optimal control signals of heart rate and ventilation rate are calculated. We adopt a nonlinear model of the cardiorespiratory system and propose

an iterative algorithm to compute the optimal control signals of the system.

Finally, we focus on detection of a serious chronic respiratory sleep disorder, sleep apnea. Individuals with sleep apnea are rarely aware of the condition and are often left untreated. If left untreated, sleep apnea leads to some serious adverse physiological conditions such as hypertension, cardiac arrhythmia, heart attack, and stroke. We develop a new framework for sleep apnea detection, in which multiple sensor measurement data are combined with the cardiorespiratory model information to achieve better detection accuracy.

1.3 Organization

This chapter has introduced the background and motivation for the development of a smart health monitoring system using low-cost wearable sensors. The remainder of this dissertation is organized as follows.

In Chapter 2, we consider the biometric recognition problem. A biometric recognition system based on a single-lead ECG signal is developed. A multitask learning approach is proposed to efficiently perform feature selection and classifier design simultaneously.

In Chapter 3, we review several existing mathematical models of cardiovascular and respiratory systems.

In Chapter 4, we describe the local control mechanism modeling of the cardiorespiratory system. The optimal control problem is formulated as a convex optimization problem, which can be solved efficiently using interior-point methods.

In Chapter 5, the problem of sleep apnea detection is considered. A new fusion framework combining multiple sensor measurement data and the cardiorespiratory model information is proposed.

Finally, conclusions and suggestions for future work are provided in Chapter 6.

CHAPTER 2

ECG-BASED BIOMETRIC RECOGNITION

Biometric recognition is the science of recognizing the identity of a person based on the person's physical or behavioral attributes [38]. Biometric recognition systems operate on the fact that certain physical or behavioral attributes of humans are unique to an individual. The attributes used in a biometric recognition system are referred to as biometric traits. In general, to determine the suitability of a biometric trait, the following seven factors are considered [38]:

- *Universality:* Every individual should possess the trait.
- *Uniqueness:* The trait should be sufficiently different across individuals in the population.
- *Permanence:* The trait of an individual must be sufficiently invariant with respect to the biometric matching algorithm over a period of time.
- *Measurability:* The trait should be easy to acquire using suitable devices without causing any inconvenience to the individual.
- *Performance:* Recognition accuracy and the computational resources required to achieve that accuracy and throughput should meet the constraints of the biometric application.
- *Acceptability:* Individuals utilizing the application should be willing to present their trait to the biometric system.

- *Circumvention:* It should be hard to mimic the trait and fool the biometric system.

In practice, no single biometric trait effectively meets all the above requirements imposed by all applications all the time. However, based on nature and type of application, certain traits are acceptable. The commonly used biometric traits can be broadly classified into the following two categories:

- *Physical:* Traits based on the direct measurement of a part of the human body such as fingerprint, palmprint, retina, iris, face, ear, etc.

- *Behavioral:* Traits based on a certain behavioral action of a person such as signature, gait, voice, keystroke, etc.

Behavioral traits are generally less secure compared to the physical traits as it is relatively easy to mimic a behavioral trait of an individual. Current biometric recognition systems mostly use physical biometric traits such as fingerprint, face, iris, etc. These systems can also be fooled as one may impersonate others by copying their fingerprints, face images, iris scans, etc [39]. For example, in December 2014, a German hacker reproduced a fingerprint of the German Defense Minister using some high-resolution photos of her hand [40]. In [41], Sharif *et al.* developed a new class of attacks against facial biometric recognition systems, which can be implemented by printing a pair of eyeglass frames.

Nowadays, it is easy to acquire physiological signals such as electrocardiogram (ECG) signal with the available low-cost wearable sensors. So it is now possible to develop biometric recognition systems based on the traits from such physiological signals. These physiological biometric traits are hard to copy, and hence can be used in high security biometric applications.

An ECG signal records the electrical activity of the heart through electrodes placed on a subject's body. The electrodes measure changes in the electric potential

on the skin which arise due to the depolarization and repolarization of the heart muscles during each heartbeat. The ECG signal during a single heart beat is shown in Figure 2.1. The ECG consists of three waves per heart beat: a P wave, a QRS wave, and a T wave. The P wave is caused by the electric potential changes during the depolarization of the atria. The QRS wave is caused by the electric potential changes during the depolarization of the ventricles. The T wave is caused by the electric potential changes during the repolarization of the ventricles. Since ventricles have a larger muscle mass compared to the atria, the QRS wave has larger amplitude than the P wave. ECG is one of the most widely used noninvasive measurement signal for several clinical diagnosis purposes.

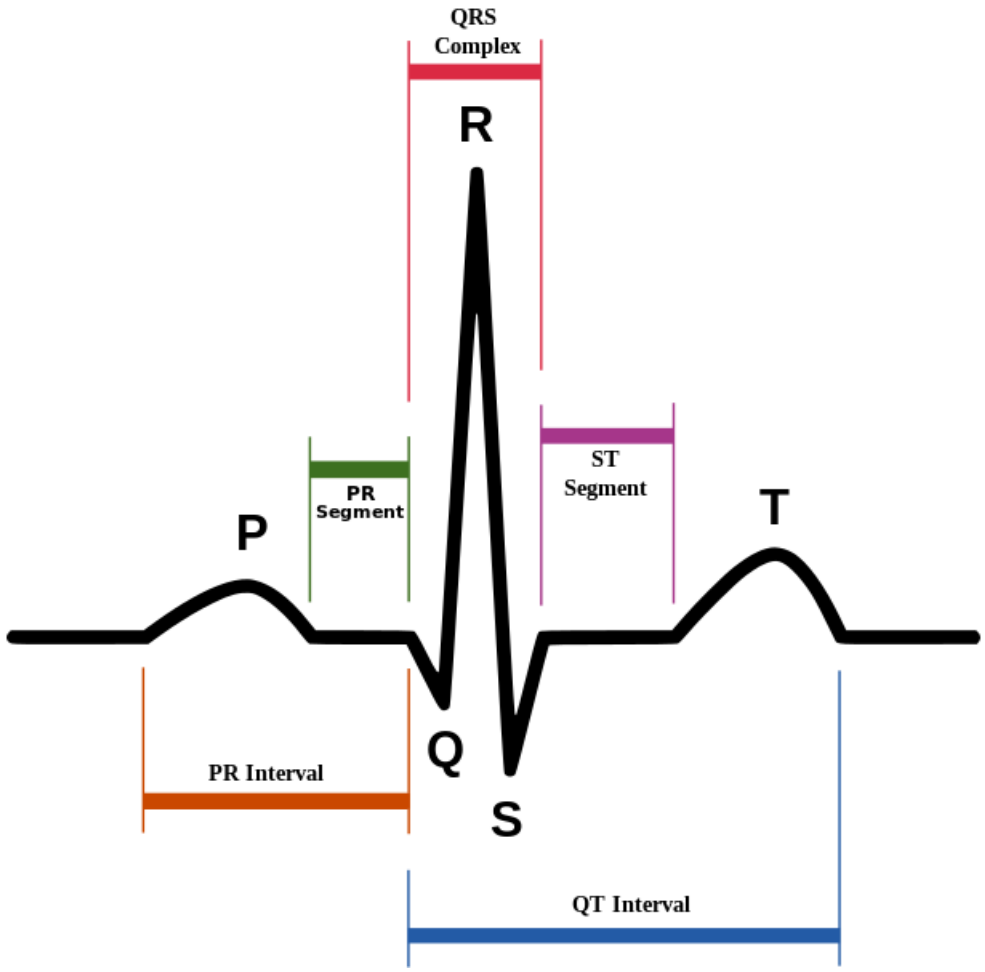


Figure 2.1: ECG signal during a single heart beat.

In this chapter, we develop a biometric recognition system based on the ECG signal. Recent years have seen a growing interest in ECG based biometric recognition techniques, especially in clinical medicine [42]. Each person has a unique ECG pattern due to the unique physical and geometrical structure of his/her heart and body [43], which makes ECG useful for biometric recognition.

Generally in ECG based biometric recognition systems, features are defined in two ways: fiducial and non-fiducial [44]. In fiducial methods, the ECG features are defined using certain fixed fiducial points such as the peaks of P, QRS, and T waves. Using the time intervals between and amplitudes at these fiducial points, we can define several features for an ECG signal. Non-fiducial methods do not require detecting fiducial points, and usually consist of defining features using the ECG record of a longer duration. For example, a discrete cosine transform can be used to define the non-fiducial features of an ECG signal, where the transform coefficients are used as features [44]. After defining the ECG features, standard classification algorithms are used to identify the individuals.

In [45], Biel *et al.* defined features using the fiducial points from a standard 12-lead ECG, and used the soft independent modeling by class analogy (SIMCA) method for classification. Israel *et al.* also defined features using the fiducial points, and used linear discriminant analysis (LDA) for classification [46]. Wübbeler *et al.* used a non-fiducial method for defining features [47]. A two-dimensional heart vector signal is constructed using three Einthoven leads. Wang *et al.* defined non-fiducial features using autocorrelation and discrete cosine transform (DCT) methods [44]. Irvine *et al.* also used a non-fiducial method based on principal component analysis (PCA) to define features [48]. The nearest neighbor method was used for classification in [44, 47, 48]. Tawfik *et al.* defined non-fiducial features using DCT of the QRS complex, and used neural networks for classification [49].

Table 2.1: Related work on ECG-based biometric recognition

	Features	Classifier
Biel *et al.* '01 [45]	Fiducial	Soft independent modeling by class analogy (SIMCA)
Israel *et al.* '05 [46]	Fiducial	Linear discriminant analysis (LDA)
Wübbeler *et al.* '07 [47]	Non-fiducial (from 2D heart vector signal)	Nearest neighbor
Wang *et al.* '08 [44]	Non-fiducial (autocorrelation and DCT)	Nearest neighbor
Irvine *et al.* '08 [48]	Non-fiducial (PCA)	Nearest neighbor
Tawfik *et al.* '11 [49]	Non-fiducial (DCT of QRS complex)	Neural network

Most of existing work uses simple classification algorithms such as linear classifiers, nearest neighbor methods, etc., which may not be robust in practice. Since the ECG feature vectors need not always be linearly separable, the linear classifiers may not perform well. ECG feature vectors may also be noisy in practice. The performance of nearest neighbor methods degrades as evidenced in [50] and our experimental results. Moreover, existing methods have not applied feature selection/extraction to identify only relevant ECG features for classification. The redundant or irrelevant features may not only lead to the problem of overfitting, but also increase the overall computational complexity [50].

In this chapter, we develop an efficient biometric recognition system using a single-lead ECG signal. A single-lead ECG signal can be easily acquired in many situations [51]. A novel recognition framework is proposed capable of dealing with nonlinearly separable data of a high dimension. We convert the classification problem of multiple subjects into a set of binary classification problems (tasks). Each binary classification task corresponds to identifying one subject. For each binary classification task, a robust nonlinear kernel classifier is designed. In each classifier, the features are scaled according to their relevance so that more relevant features have more effect during classification. This step not only removes redundancy/irrelevance in features, but also greatly enhances the generalization of the classification algorithm. The classifier parameters and feature scaling parameters for all the tasks are jointly estimated using the available training data. More specifically, we combine feature selection and classifier design into a single learning problem. Furthermore, we combine the learning of all the tasks, as the tasks are not completely independent. The learning of one classification task may be useful for the learning of other (related) tasks. This greatly improves the learning efficiency.

2.1 Related Work on Joint Feature Selection and Classifier Design

Recent years have seen growing research interest in simultaneous (or joint) feature selection and classifier design due to the fact that selecting features independently of the classifier is not always optimal. The relative importance of features may sometimes depend on the type of classifier used. For linear classifiers, joint feature selection and classifier design can be easily done as the number of classifier parameters is determined by the number of features. In such cases, we can directly use the classifier parameters to model the relevance of individual features to classification. However, for nonlinear classifiers, joint feature selection and classifier design is not straightforward. In [52], Weston *et al.* proposed a feature selection approach for support vector machines (SVMs) by using scaling factors for individual features. The scaling factors were found by minimizing the SVM generalization bound through the gradient descent approach. The similar problem was also studied in [53], where the scaling factors and the SVM parameters were jointly obtained by minimizing the standard SVM empirical risk function. In [54], Krishnapuram *et al.* proposed a Bayesian approach for this type of problem. Sparsity promoting zero-mean Gaussian priors were placed on both the feature scaling factors and the classifier parameters. With these priors, the final estimates of the feature scaling factors and classifier parameters were found by maximizing the posterior probability.

In general, each binary classifier in an M-ary classification problem is trained separately using the available training data. This is referred to as single task learning (STL). Often, these learning tasks are not completely independent, as they may share some common information (or structure). Hence, learning of one task may be useful for learning of other tasks. Multitask learning (MTL) is an approach in which multiple related tasks are learned in parallel with some shared representation [55]. This greatly improves the efficiency of learning, as each task learns from the other related tasks as well. For example, in face recognition, learning one particular human's

face will be useful for learning others' as all humans share the same facial structure and biology. The MTL approach has been applied to several other prediction and classification problems. In [55], Caruana proposed a multitask learning approach based on backpropagation neural networks. The hidden layer of the backpropagation neural network was shared between all the tasks. In [56], the authors proposed a Bayesian approach for a set of similar classification and regression problems, where some model parameters (input-hidden layer weights) were shared among all the tasks and a joint prior distribution was placed on the other parameters.

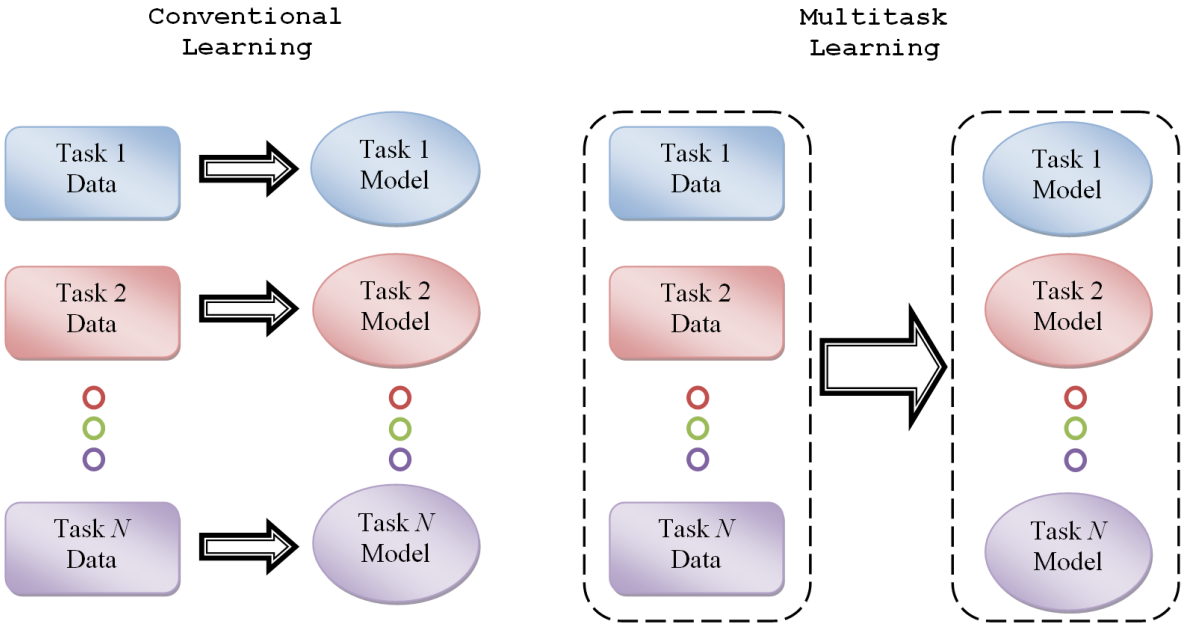


Figure 2.2: Multitask learning.

The MTL approach has also been applied to feature selection. In [57], Jebara proposed a common feature selection method for multiple SVMs based on maximum entropy discrimination. In [58], Argyriou *et al.* proposed an approach for learning common sparse features across multiple related supervised learning tasks. They formulated this multitask feature learning problem as a convex optimization problem and used the alternating minimizing algorithm to solve it. In [59], Chen *et al.* considered the problem of learning sparse and low-rank patterns from multiple tasks. They

considered linear classifiers for each task, the feature scaling parameters of which were decomposed into sparse and low-rank components. The sparse component highlights the discriminative features for each classification task, and the low-rank component captures the common discriminative feature subspace between different tasks.

Most of the existing work still treats feature selection and classifier design for multiple tasks as separate learning problems. Even though in [59], joint feature selection and classifier design for multiple tasks was proposed, the authors focused on linear classifiers only. Linear classifiers have limited performance for many real-world problems, and cannot be extended to nonlinear ones straightforwardly. As mentioned in Section I, we formulate ECG based biometric recognition as a multitask learning problem. The original classification problem with more than two classes is first converted into a set of kernel based binary classification tasks. A matrix is formed which consists of the feature scaling factors for all the tasks. It is then decomposed into a sparse component and a low-rank component. The sparse component determines the features that are relevant only to each individual task, and the low-rank component determines the common feature subspace that is relevant for all the tasks. Thus, the relatedness between different tasks is modeled through the feature subspace sharing between tasks.

2.2 ECG Signal Processing and Feature Extraction

We use a non-fiducial method to define the features of a single-lead ECG signal. Non-fiducial features are generally preferred over fiducial features, as using a small number of fiducial features may not generalize well to a large number of subjects [48]. Moreover, reliably detecting fiducial points in ECG signals for all the subjects automatically is not always possible in practice [48]. Our data preprocessing has two main steps:

1. filtering,

2. non-fiducial feature value calculation.

To remove the noise in raw ECG signals, a fourth-order Butterworth bandpass filter with cutoff frequencies 1 and 40 Hz is first applied. Then, the entire ECG record of a subject is divided into non-overlapping windows of 5 second duration. The non-fiducial features for each window are computed using the autocorrelation and DCT technique in [44]. Using the autocorrelation reduces the effect of heart rate variability on the recognition task [60]. Specifically, the normalized autocorrelation function is given by,

$$R_{xx}[m] = \frac{\sum_{i=0}^{N-m-1} x[i]x[i+m]}{R_{xx}[0]}, \tag{2.1}$$

where $x[i]$ is the ith ECG sample in one window, N is the window length, and $x[i+m]$ is the time-shifted version of $x[i]$ with time lag $m = 0, 1, ..., M-1$, $M \ll N$. The DCT of the obtained autocorrelation is as follows,

$$Z[k] = w[k] \sum_{m=0}^{M-1} R_{xx}[m] \cos\left(\frac{\pi(2m+1)k}{2M}\right), \tag{2.2}$$

where $k = 0, 1, ..., M-1$, and $w[k]$ is given by

$$w[k] = \begin{cases} \sqrt{\frac{1}{M}}, & k = 0, \\ \sqrt{\frac{2}{M}}, & 1 \leq k \leq M-1. \end{cases} \tag{2.3}$$

In general, the first d DCT coefficients, $\{Z[k], k = 0, ..., d-1\}$, may contain significant information, $d < M$. Therefore, these d DCT coefficients are defined as the non-fiducial features of one ECG window.

2.3 Problem Formulation

We consider an L-class classification problem (recognition of L subjects), $L > 2$. We convert this into L one-versus-rest binary classification problems (tasks). That is, we decide whether the input ECG record belongs to subject l or not, $l = 1, 2, ..., L$. For each subject, we initially divide the entire ECG record into non-overlapping

windows of 5 second duration. The non-fiducial ECG features are then computed for each window. Let $\mathbf{T}_l = \{(\mathbf{x}_i^l, y_i^l); i = 1, ..., N\}$ be the available training data for the lth binary task, where N is the number of training ECG windows for each task. Let d be the total number of non-fiducial features defined for the ith window, i.e., $\mathbf{x}_i^l \in \mathbb{R}^d$. The label of the sample $\mathbf{x}_i^l$ is $y_i^l \in \{0, 1\}$ corresponding to "not subject l" and "subject l", respectively. Let $f_l(\cdot)$ be the classifier function for the lth binary task, i.e., $y^l = f_l(\mathbf{x}^l; \mathbf{w}^l)$, where $\mathbf{w}^l$ are the classifier parameters. In general, these L different tasks are related to each other. Our goal here is to learn the relevant subset of non-fiducial features for each task and the L binary classifier functions simultaneously using the available data $\mathbf{D} = \{\mathbf{T}_1, \mathbf{T}_2, ..., \mathbf{T}_L\}$.

2.3.1 Binary classifier

For each binary classification task, we consider the following probabilistic kernel classifier based on the generalized linear model [61],

$$P(y^l = 1|\mathbf{x}^l) = \sigma\left(w_0^l + \sum_{i=1}^{N} w_i^l K(\mathbf{x}^l, \mathbf{x}_i^l)\right), \tag{2.4}$$

where $\mathbf{w}^l = [w_0^l, w_1^l, ..., w_N^l]^T$ are the classifier parameters, N is the size of the training dataset, $K(\cdot, \cdot)$ is a valid kernel function, and $\sigma(a) = \frac{1}{1+e^{-a}}$. The classifier function $f_l(\cdot)$ assigns the class label based on thresholding the class probability, $P(y^l = 1|\mathbf{x}^l)$. Here, the Gaussian kernel function is applied as it corresponds to an infinite dimensional feature mapping,

$$K(\mathbf{x}, \mathbf{z}) = \exp\left(-\sum_{j=1}^{d} (x_j - z_j)^2\right). \tag{2.5}$$

The main advantage of the probabilistic kernel classifier is that it directly models the class probabilities without the need of modeling the underlying class generative distributions. Moreover, the data is projected to a very high dimensional space (infinite in our case) where the data is well separated. The kernel function eliminates the need of defining this feature mapping explicitly.

2.3.2 Feature scaling

For each task, a nonnegative scaling factor is introduced corresponding to each feature. This factor measures the relevance of a particular feature for the given task. Higher its value, the more relevant the corresponding feature is to the learning of the task. Let θ_j^l be the scaling factor corresponding to feature j for task l. Accordingly, $\boldsymbol{\theta}^l = [\theta_1^l, \theta_2^l, ..., \theta_d^l]^T$ is a vector of all the scaling factors for task l. For all the tasks, the scaling factors can be represented in a matrix form as follows,

$$\boldsymbol{\Theta} = \left[\boldsymbol{\theta}^1 | \boldsymbol{\theta}^2 | \dots | \boldsymbol{\theta}^L\right]_{d \times L}. \tag{2.6}$$

Hence, feature relevance is modeled through matrix $\boldsymbol{\Theta}$. We decompose $\boldsymbol{\Theta}$ into two components, one sparse and one low rank, as follows:

$$\boldsymbol{\Theta} = \mathbf{S} + \mathbf{R}, \tag{2.7}$$

where $\mathbf{S}$ is the sparse matrix, which means that only a few elements of the matrix are nonzero. In other words, there will be only a few features selected corresponding to each task. $\mathbf{R}$ is the low-rank matrix. That is, a common low dimensional subspace is extracted for all tasks. Note that all the matrices here are of dimension $d \times L$, and each column corresponds to a task. Let the rank of matrix $\mathbf{R}$ be $r \leq \min(d, L)$. We can decompose $\mathbf{R}$ further as follows,

$$\begin{aligned} \mathbf{R}_{d \times L} &= \mathbf{B}_{d \times r} \mathbf{C}_{r \times L} \\ &= \begin{bmatrix} | & | & & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_r \\ | & | & & | \end{bmatrix} \begin{bmatrix} c_1^1 & c_1^2 & \dots & c_1^L \\ c_2^1 & c_2^2 & \dots & c_2^L \\ \vdots & \vdots & \ddots & \vdots \\ c_r^1 & c_r^2 & \dots & c_r^L \end{bmatrix}, \end{aligned} \tag{2.8}$$

where $\mathbf{B}$ is a matrix with r basis column vectors each having a unit norm, and $\mathbf{C}$ is the coefficient matrix. Therefore, for the l-th task, the feature scaling vector (or

feature weight vector) is given by

$$\begin{aligned}\boldsymbol{\theta}^l &= \mathbf{s}^l + \mathbf{r}^l \\ &= \mathbf{s}^l + c_1^l \mathbf{b}_1 + c_2^l \mathbf{b}_2 + \ldots + c_r^l \mathbf{b}_r \\ &= \mathbf{s}^l + \sum_{i=1}^{r} c_i^l \mathbf{b}_i,\end{aligned} \tag{2.9}$$

where $\mathbf{s}^l$ is the l-th column of the sparse matrix $\mathbf{S}$ and is different for each task. The basis vectors $\{\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_r\}$ span a linear subspace. So the low-rank matrix $\mathbf{R}$ relates all tasks by sharing this common feature subspace.

By integrating the feature scaling factors into the Gaussian kernel, we have[1]

$$K_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}) = \exp\left(-\sum_{j=1}^{d} \theta_j \left(x_j - z_j\right)^2\right). \tag{2.10}$$

This scaled kernel function, $K_{\boldsymbol{\theta}}(\cdot, \cdot)$, is used in the classifier (2.4) instead. By putting all classifier parameters together, we have a more compact representation,

$$\mathbf{W} = \left[\mathbf{w}^1 | \mathbf{w}^2 | \ldots | \mathbf{w}^L\right]_{(N+1)\times L}. \tag{2.11}$$

Our goal is to learn both matrices $\mathbf{W}$ and $\boldsymbol{\Theta}$ based on the available training data $\mathbf{D}$ efficiently.

2.4 Multitask Learning Algorithm

With the generalized probabilistic kernel classifier (as in (2.4)) for each task, the multitask likelihood function of all training data is given by

$$\mathcal{L}(\mathbf{W}, \boldsymbol{\Theta}) = P(\mathbf{Y} | \{\mathbf{X}^l\}, \mathbf{W}, \boldsymbol{\Theta}) = \prod_{l=1}^{L} P(\mathbf{y}^l | \mathbf{X}^l, \mathbf{w}^l, \boldsymbol{\theta}^l), \tag{2.12}$$

where the likelihood function of task l is given by

$$\begin{aligned}&P(\mathbf{y}^l | \mathbf{X}^l, \mathbf{w}^l, \boldsymbol{\theta}^l) \\ &= \prod_{i=1}^{N} \left[\sigma\left(\mathbf{w}^{lT} \mathbf{k}_{\boldsymbol{\theta}^l}(\mathbf{x}_i^l)\right)\right]^{y_i^l} \left[\sigma\left(-\mathbf{w}^{lT} \mathbf{k}_{\boldsymbol{\theta}^l}(\mathbf{x}_i^l)\right)\right]^{(1-y_i^l)},\end{aligned} \tag{2.13}$$

[1] Task index l is omitted here because the kernel form is the same for all tasks.

where $\mathbf{k}_{\boldsymbol{\theta}^l}(\mathbf{x}) = [1, e^{-\boldsymbol{\theta}^{lT}\mathbf{z}_1}, e^{-\boldsymbol{\theta}^{lT}\mathbf{z}_2}, ..., e^{-\boldsymbol{\theta}^{lT}\mathbf{z}_N}]^T$, $\mathbf{z}_i = [(x_1 - x_{1i})^2, (x_2 - x_{2i})^2, ..., (x_d - x_{di})^2]^T$, and $\mathbf{x} = [x_1, x_2, ..., x_d]^T$ is a sample.

In this likelihood function, we have the following unknown variables: $\mathbf{W}$, the $(N+1) \times L$ classifier parameter matrix, $\boldsymbol{\Theta} = \mathbf{S} + \mathbf{R}$, the scaling factor matrix with $\mathbf{S}$ being a $d \times L$ sparse matrix and $\mathbf{R}$ a $d \times L$ low-rank matrix. We intend to estimate these parameters in the sense that the multitask likelihood in (2.12) is maximized. Therefore, our optimization problem can be formulated as follows,

$$\begin{aligned} \min_{\mathbf{W},\mathbf{S},\mathbf{R}} \quad & -\log \mathcal{L}(\mathbf{W}, \mathbf{S}, \mathbf{R}) + \lambda_1 \|\mathbf{W}\|_0 + \lambda_2 \|\mathbf{S}\|_0 + \alpha \operatorname{rank}(\mathbf{R}) \\ \text{s.t.} \quad & \mathbf{S} + \mathbf{R} \succeq \mathbf{0}, \end{aligned} \tag{2.14}$$

where $\{\lambda_1, \lambda_2, \alpha\}$ are the regularization parameters, $\|\cdot\|_0$ is the ℓ_0 norm (the number of nonzero elements), and $\operatorname{rank}(\cdot)$ denotes the rank of the matrix. The regularization term of $\mathbf{W}$ penalizes the number of nonzero elements in $\mathbf{W}$, which reduces overfitting. Similarly, the regularization term of $\mathbf{S}$ penalizes the number of nonzero elements in $\mathbf{S}$, which results in a sparse feature scaling component. The regularization term of $\mathbf{R}$ penalizes the rank of matrix $\mathbf{R}$, which results in a low dimensional common feature subspace. The constraint $\mathbf{S} + \mathbf{R} \succeq \mathbf{0}$ ensures that the feature scaling factors are nonnegative.

The rank function and the ℓ_0 norm are nonconvex functions. Nonconvex optimization is in general difficult, as nonconvex functions have many local optima. So we replace the ℓ_0 norm by its convex envelope (tightest convex approximation), ℓ_1 norm [62], and the rank function by its convex envelope, nuclear (or trace) norm [63]. Therefore, the new optimization problem becomes

$$\begin{aligned} \min_{\mathbf{W},\mathbf{S},\mathbf{R}} \quad & -\log \mathcal{L}(\mathbf{W}, \mathbf{S}, \mathbf{R}) + \lambda_1 \|\mathbf{W}\|_1 + \lambda_2 \|\mathbf{S}\|_1 + \alpha \|\mathbf{R}\|_* \\ \text{s.t.} \quad & \mathbf{S} + \mathbf{R} \succeq \mathbf{0}, \end{aligned} \tag{2.15}$$

where $\|\mathbf{A}\|_1 = \sum_{i,j} |a_{ij}|$ is the ℓ_1 norm and $\|\mathbf{A}\|_* = \sum_i \sigma_i$ is the trace norm (or nuclear norm), σ_i are the singular values of the matrix $\mathbf{A}$. Under the availability of

large datasets, convex relaxation achieves a good trade-off between decreasing the computational complexity and minimizing the risk objective function [64].

2.4.1 Optimization algorithm

It is still difficult to solve the optimization problem (2.15) directly as the objective function has some non-differentiable components (ℓ_1 norm and nuclear norm). Besides, the objective function is not strictly convex due to the quasiconvexity of the negative log-likelihood term. Note the non-differentiable components in the objective function are separable in the variables, which can be exploited. Accordingly, we divide the optimization problem into three optimization problems in the three variables $\mathbf{W}$, $\mathbf{S}$, and $\mathbf{R}$. At each step, we perform optimization with respect to one variable while fixing the other two. Specifically, at the kth iteration, we solve the following subproblems:

- Optimizing with respect to $\mathbf{W}$ by fixing $\mathbf{S}$ and $\mathbf{R}$:

$$\mathbf{W}^{(k)} = \arg\min_{\mathbf{W}} -\log \mathcal{L}\left(\mathbf{W}, \mathbf{S}^{(k-1)}, \mathbf{R}^{(k-1)}\right) + \lambda_1 \|\mathbf{W}\|_1. \tag{2.16}$$

- Optimizing with respect to $\mathbf{S}$ by fixing $\mathbf{W}$ and $\mathbf{R}$:

$$\begin{aligned} \mathbf{S}^{(k)} = \arg\min_{\mathbf{S}} \quad & -\log \mathcal{L}\left(\mathbf{W}^{(k)}, \mathbf{S}, \mathbf{R}^{(k-1)}\right) + \lambda_2 \|\mathbf{S}\|_1 \\ \text{s.t.} \quad & \mathbf{S} + \mathbf{R}^{(k-1)} \succeq \mathbf{0}. \end{aligned} \tag{2.17}$$

- Optimizing with respect to $\mathbf{R}$ by fixing $\mathbf{W}$ and $\mathbf{S}$:

$$\begin{aligned} \mathbf{R}^{(k)} = \arg\min_{\mathbf{R}} \quad & -\log \mathcal{L}\left(\mathbf{W}^{(k)}, \mathbf{S}^{(k)}, \mathbf{R}\right) + \alpha \|\mathbf{R}\|_* \\ \text{s.t.} \quad & \mathbf{R} + \mathbf{S}^{(k)} \succeq \mathbf{0}. \end{aligned} \tag{2.18}$$

The optimization with respect to $\mathbf{W}$ is unconstrained. We convert the above second and third subproblems into unconstrained problems by using the logarithmic barrier

function. So we have

$$\begin{aligned}\mathbf{S}^{(k)} = \ & \arg\min_{\mathbf{S}} \quad -\log \mathcal{L}\left(\mathbf{W}^{(k)}, \mathbf{S}, \mathbf{R}^{(k-1)}\right) + \lambda_2 \left\|\mathbf{S}\right\|_1 \\ & \qquad -\frac{1}{t_1}\sum_{i=1}^{d\cdot L} \log\left(s_i + r_i^{(k-1)}\right), \\ \mathbf{R}^{(k)} = \ & \arg\min_{\mathbf{R}} \quad -\log \mathcal{L}\left(\mathbf{W}^{(k)}, \mathbf{S}^{(k)}, \mathbf{R}\right) + \alpha \left\|\mathbf{R}\right\|_* \\ & \qquad -\frac{1}{t_2}\sum_{i=1}^{d\cdot L} \log\left(s_i^{(k)} + r_i\right).\end{aligned} \tag{2.19}$$

Note that all the above subproblems are in the following form:

$$\min_{\mathbf{y}} \quad h(\mathbf{y}) = f(\mathbf{y}) + \lambda g(\mathbf{y}), \tag{2.20}$$

where $\mathbf{y}$ is the optimization variable, $f(\cdot)$ is a smooth (differentiable) quasiconvex function, and $g(\cdot)$ is a nonsmooth convex function. We adopt the proximal gradient method [65] to solve these three problems. At each iteration, we linearize the function $f(\cdot)$ at the estimate from the previous iteration and then minimize it. That is, at step k, we have

$$\mathbf{y}^{(k)} = \arg\min_{\mathbf{y}} \quad f_l(\mathbf{y}, \mathbf{y}^{(k-1)}) + \lambda g(\mathbf{y}), \tag{2.21}$$

where $f_l(\mathbf{y}, \mathbf{y}^{(k-1)})$ is given by

$$\begin{aligned}f_l(\mathbf{y}, \mathbf{y}^{(k-1)}) = \ & f(\mathbf{y}^{(k-1)}) + \left\langle \mathbf{y} - \mathbf{y}^{(k-1)}, \nabla f(\mathbf{y}^{(k-1)}) \right\rangle \\ & + \gamma \left\|\mathbf{y} - \mathbf{y}^{(k-1)}\right\|_2^2.\end{aligned} \tag{2.22}$$

Here, ∇ denotes the gradient of the function and $\langle \cdot, \cdot \rangle$ denotes the inner product (dot product). We initially start with a feasible point, and stop when $\left\|h\left(\mathbf{y}^{(k)}\right) - h\left(\mathbf{y}^{(k-1)}\right)\right\|_2^2 < \epsilon$, where $h(\cdot)$ is the objective function and ϵ is some preset tolerance level. More explicitly, at the kth iteration, we have[2]

[2]We vectorize the matrices $\mathbf{W}$ and $\mathbf{S}$ for simplicity in calculations. That is, we convert the matrices $\mathbf{W}$ and $\mathbf{S}$ into column vectors $\mathbf{w}$ and $\mathbf{s}$, respectively, by stacking their columns on top of one another.

- Optimizing with respect to $\mathbf{W}$ by fixing $\mathbf{S}$ and $\mathbf{R}$:

$$\begin{aligned}\mathbf{w}^{(k)} = \ \arg\min_{\mathbf{w}} \ & \left\langle \mathbf{w} - \mathbf{w}^{(k-1)}, \nabla f_1\left(\mathbf{w}^{(k-1)}\right)\right\rangle \\ & + \gamma_1 \left\|\mathbf{w} - \mathbf{w}^{(k-1)}\right\|_2^2 + \lambda_1 \left\|\mathbf{w}\right\|_1,\end{aligned} \tag{2.23}$$

  where $f_1(\mathbf{w}) = -\log \mathcal{L}\left(\mathbf{w}, \mathbf{s}^{(k-1)}, \mathbf{r}^{(k-1)}\right)$.

- Optimizing with respect to $\mathbf{S}$ by fixing $\mathbf{W}$ and $\mathbf{R}$:

$$\begin{aligned}\mathbf{s}^{(k)} = \ \arg\min_{\mathbf{s}} \ & \left\langle \mathbf{s} - \mathbf{s}^{(k-1)}, \nabla f_2\left(\mathbf{s}^{(k-1)}\right)\right\rangle \\ & + \gamma_2 \left\|\mathbf{s} - \mathbf{s}^{(k-1)}\right\|_2^2 + \lambda_2 \left\|\mathbf{s}\right\|_1,\end{aligned} \tag{2.24}$$

  where $f_2(\mathbf{s}) = -\log \mathcal{L}\left(\mathbf{W}^{(k)}, \mathbf{S}, \mathbf{R}^{(k-1)}\right) - \frac{1}{t_1}\sum_{i=1}^{d\cdot L} \log\left(s_i + r_i^{(k-1)}\right)$.

- Optimizing with respect to $\mathbf{R}$ by fixing $\mathbf{W}$ and $\mathbf{S}$:

$$\begin{aligned}\mathbf{R}^{(k)} = \ \arg\min_{\mathbf{R}} \ & \left\langle \mathbf{R} - \mathbf{R}^{(k-1)}, \nabla f_3\left(\mathbf{R}^{(k-1)}\right)\right\rangle \\ & + \gamma_3 \left\|\mathbf{R} - \mathbf{R}^{(k-1)}\right\|_F^2 + \alpha \left\|\mathbf{R}\right\|_*,\end{aligned} \tag{2.25}$$

  where $f_3(\mathbf{R}) = -\log \mathcal{L}\left(\mathbf{W}^{(k)}, \mathbf{S}^{(k)}, \mathbf{R}\right) - \frac{1}{t_2}\sum_{i=1}^{d\cdot L} \log\left(s_i^{(k)} + r_i\right)$.

The optimization problems (2.23) and (2.24) can be solved in closed forms. Specifically, the solution of (2.23) is given by

$$\mathbf{w}^{(k)} = \text{soft}\left(\mathbf{w}^{(k-1)} - \frac{1}{2\gamma_1}\nabla f_1\left(\mathbf{w}^{(k-1)}\right), \frac{\lambda_1}{2\gamma_1}\right), \tag{2.26}$$

where $\text{soft}(\cdot,\cdot)$ is the soft thresholding operator defined as [65]

$$\left(\text{soft}(\mathbf{y}, \alpha)\right)_i = \text{sgn}(y_i) \cdot \max(|y_i| - \alpha, 0), \quad i = 1, 2, ..., n, \tag{2.27}$$

where $\text{sgn}(\cdot)$ is the sign function, $\mathbf{y} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}_+$. Similarly, the solution of (2.24) is given by

$$\mathbf{s}^{(k)} = \text{soft}\left(\mathbf{s}^{(k-1)} - \frac{1}{2\gamma_2}\nabla f_2\left(\mathbf{s}^{(k-1)}\right), \frac{\lambda_2}{2\gamma_2}\right). \tag{2.28}$$

The problem (2.25) can be simplified into the following form:

$$\begin{aligned}\mathbf{R}^{(k)} = \ \arg\min_{\mathbf{R}} \ \ & \gamma_3 \left\| \mathbf{R} - \left(\mathbf{R}^{(k-1)} - \frac{1}{2\gamma_3} \nabla f_3 \left(\mathbf{R}^{(k-1)} \right) \right) \right\|_F^2 \\ & + \alpha \left\| \mathbf{R} \right\|_* ,\end{aligned} \tag{2.29}$$

where $\left\| \cdot \right\|_F$ is the Frobenius norm of the matrix and is defined as $\left\| \mathbf{A} \right\|_F = \sqrt{\sum_i \sum_j |a_{ij}|^2}$. Let $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ be the SVD of the matrix $\mathbf{R}^{(k-1)} - \frac{1}{2\gamma_3} \nabla f_3 \left(\mathbf{R}^{(k-1)} \right)$. Then the solution of the problem (2.29) is given by

$$\mathbf{R}^{(k)} = \mathbf{U}\mathbf{\Sigma}_\alpha \mathbf{V}^T, \tag{2.30}$$

where $\left(\mathbf{\Sigma}_\alpha \right)_{ii} = \max\{0, \mathbf{\Sigma}_{ii} - \alpha\}$.

2.4.2 Convergence and computational complexity

The proximal gradient method is known to have a sublinear rate of convergence [65],

$$h\left(\mathbf{x}^{(k)} \right) - h\left(\mathbf{x}^\star \right) \simeq O\left(\frac{1}{k} \right), \tag{2.31}$$

where $\mathbf{x}^\star$ is the global minimum. The proximal gradient method converges to the global minimum when the smooth (differentiable) function $f(\cdot)$ is convex. Here in our case, the smooth functions are not strictly convex, but are quasiconvex. For quasiconvex functions, the first order condition of convexity,

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}), \ \ \forall \ \ \mathbf{x}, \mathbf{y} \in \text{dom } f, \tag{2.32}$$

does not hold strictly. This means that the gradient $\nabla f(\mathbf{x})$ can be zero even when $\mathbf{x}$ is not a global minimum. This may sometimes lead us to a local minimum. To mitigate this issue, we run the algorithm using different initial starting points and finally pick the solution with minimum objective value. This will not have major effect in practice, as we often use parallel computation techniques.

Based on (2.26)-(2.29), the number of computations in each iteration is proportional to the number of parameters in the matrices $\mathbf{W}$, $\mathbf{S}$, and $\mathbf{R}$. The computational

complexity of the algorithm in each iteration is $O((N + 2d)L)$, where N is the size of the training data, d is the number of non-fiducial features, and L is the number of subjects.

2.4.3 Final classification decision rule

After obtaining $\mathbf{W}$, $\mathbf{S}$, and $\mathbf{R}$, we calculate the class probabilities of all subjects, $P(y^l = 1|\mathbf{x})$, $l = 1, 2, ..., L$, given a data sample $\mathbf{x}$. The probability $P(y^l = 1|\mathbf{x})$ represents the total evidence available for telling whether the data sample $\mathbf{x}$ belongs to subject l or not. Using the Dempster-Shafer theory of evidence, we assign the new data sample $\mathbf{x}$ to the class (subject) with the highest belief [66]. Since all the available belief of class l is given by its class probability $P(y^l = 1|\mathbf{x})$, the data sample $\mathbf{x}$ is assigned to a class with the highest class probability, i.e.,

$$l^\star = \arg\max_l \ P(y^l = 1|\mathbf{x}). \tag{2.33}$$

2.5 Experimental Results

For performance evaluation, the MIT-BIH Normal Sinus Rhythm ECG database from PhysioNet [13] is adopted. The database contains long-term (about 24 hours) two-lead ECG recordings of 18 subjects referred to the Arrhythmia Laboratory at Boston's Beth Israel Hospital. All the ECG records are sampled at 128 Hz. Subjects include 5 men, aged 26 to 45, and 13 women, aged 20 to 50. These 18 subjects were found to have had no significant arrhythmias. Lead I ECG data is used in our experiments, as it is most commonly used and easy to acquire in many situations [51]. For each person, ECG signal from hour 1 to hour 2 is used as training data, and the ECG signal from hour 13 to hour 14 is used as test data.

The ECG records are preprocessed as described above. The non-fiducial features are defined as the first d DCT coefficients of the autocorrelation function with 70 time lags, for different values of d. Since the QRS wave is more stable than the P and

T waves, we choose the number of lags of the autocorrelation function corresponding to the length of the QRS wave portion in the ECG signal. With the sampling rate of 128 Hz and each window of 5 seconds, 70 time lags are considered sufficient for the autocorrelation function. Based on the size of the available data, 20% of the training data is used as validation data to set the regularization parameters of the algorithm. Choosing 10% results in a smaller validation set. We can also use 30% instead of 20%. 20% is chosen since it is adequate in our experiments. The estimated sparse components for all the tasks are shown in Figure 2.3, when $d = 40$. From the figure, we can see the features that are relevant to each individual task. Note that the least relevant features have the scaling factors close to zero. The rank of the estimated low-rank matrix $\hat{\mathbf{R}}$ is 12. To demonstrate the advantage of computing the underlying low-rank subspace, we use the scaled linear kernel, $K_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}) = \mathbf{x}^T \boldsymbol{\Theta} \mathbf{z}$, where $\boldsymbol{\Theta}$ is a diagonal matrix with $\boldsymbol{\theta}$ as its main diagonal. Since the scaling parameters are all nonnegative, matrix $\boldsymbol{\Theta}$ is symmetric and positive semidefinite. Hence, the above scaled linear kernel is a valid kernel. For illustration, we project the training data of arbitrary subjects into arbitrary two-dimensional subspaces of the space spanned by the estimated low-rank matrix (using its left singular vectors). The projections of the data with respect to different left singular vectors (basis vectors) are shown in Figure 2.4. From the figure, we can see that the computed low-rank space is useful for the recognition of these subjects.

Support vector machine (SVM) is one of the best supervised classifiers, since it is flexible in handling nonlinearly separable classes by projecting the data into a very high dimensional space, and constructing maximum margin decision hyperplanes in that space [67]. We use SVM with a Gaussian kernel for comparison. We extend the binary SVM classifiers to multi-class classification using posterior class probabilities. Platt's approach is used to compute the posterior class probabilities [68]. The basic idea is to fit a logistic sigmoid function to the outputs of the trained SVM. The

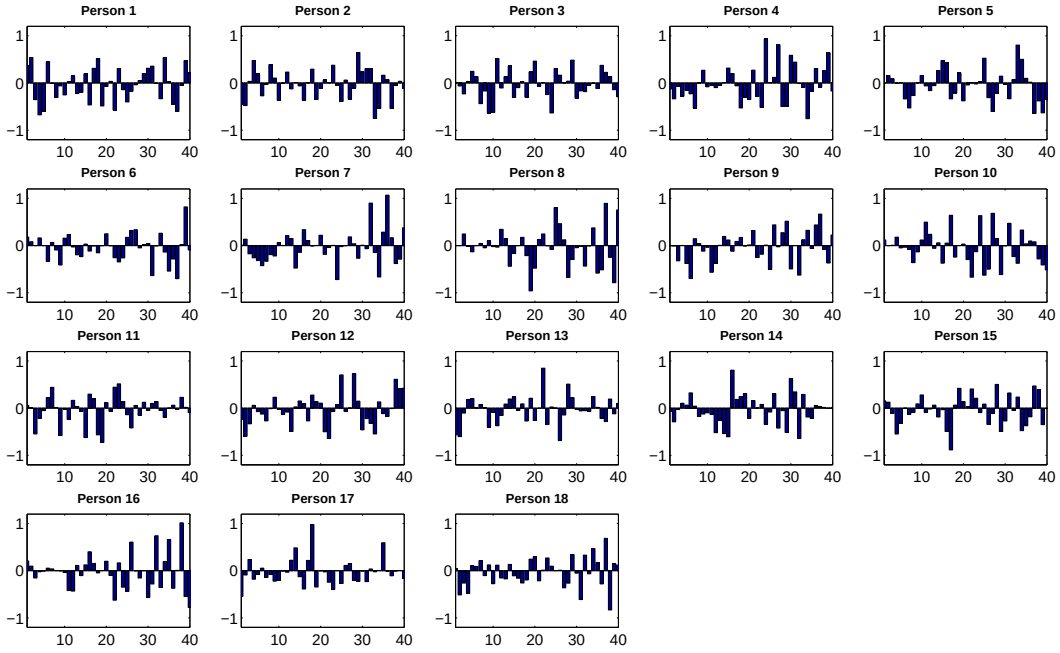


Figure 2.3: Estimated sparse components showing the feature relevance for each individual task.

proposed sigmoid model is as follows,

$$P(y = 1|f(\mathbf{x})) = \frac{1}{1 + e^{\mu f(\mathbf{x}) + \gamma}}, \tag{2.34}$$

where μ and γ are the parameters of the model and they can be found by minimizing the negative log-likelihood of the training data using the standard Newton's method with backtracking line search. The final classification decision rule is to assign a test sample to the class with the highest posterior probability.

We also compare the classification performance of the proposed method with the nearest neighbor method in [44], and the neural network method in [49]. For all the methods, we use the same set of non-fiducial features. We use a normalized Euclidean distance measure in the nearest neighbor classifier. For the neural network method in [49], we use a network of one hidden layer with d nodes. The correct detection rates, training and testing time of all the methods are shown in Figure 2.5.

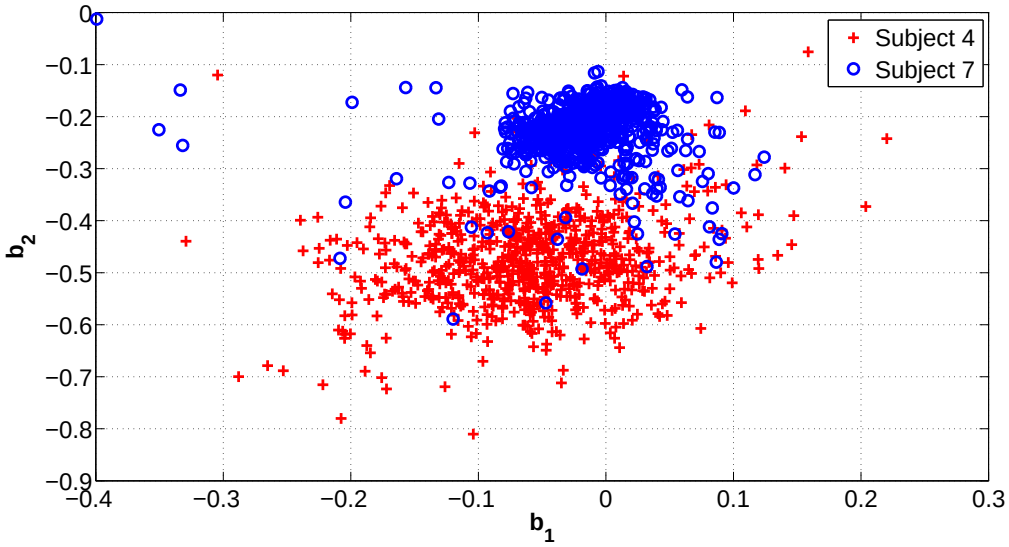
From the results, we can see that the proposed method outperforms the SVM-based method and the neural network and nearest neighbor methods in [49] and [44], especially for higher numbers of non-fiducial features. This is because in the proposed method, the features are scaled according to their relevance. Besides, the multitask learning approach is effective in estimating the classifier and feature scaling parameters. The computational complexity of the proposed method, SVM, and neural network during training is $O((N + 2d)L)$, $O(LN^2)$ [69], and $O(N(d^2 + dL)^2)$ [50], respectively. The total training time of the proposed method in the experiments is a bit higher because multiple initializations are required. In practice, training is generally conducted offline. Therefore, the practical applicability of the method is not affected. Note that the testing time is low and comparable to other methods. Unlike other methods, the nearest neighbor method does not require a training step. But it consumes more time in testing. The computational complexity of the nearest neighbor method during testing is $O(dN)$. Also, it has a high space complexity as it needs to store the entire training data over time.

In real situations, there may exist measurement noise in the acquired ECG signal, due to poor electrode contact, inherent sensor thermal noise, power line interference, etc. The noise due to poor electrode contact can be modeled as a randomly occurring step signal with exponential decay [70]. Majority of the high frequency noise is often removed by bandpass filtering in the preprocessing step. Our method is also tested in the presence of additive white Gaussian noise. The performance degrades smoothly with the increase of noise variance. In [71] [72], the SVM and neural network algorithms are shown insensitive to the additive measurement noise like our method, while the nearest neighbor classifier is sensitive [73]. The presence of artifacts in the ECG signal due to arrhythmia or ischemia will have little effect on the performance of the proposed method as long as they are present in both training and test data. If the ECG pattern significantly changes due to the development of heart diseases, the

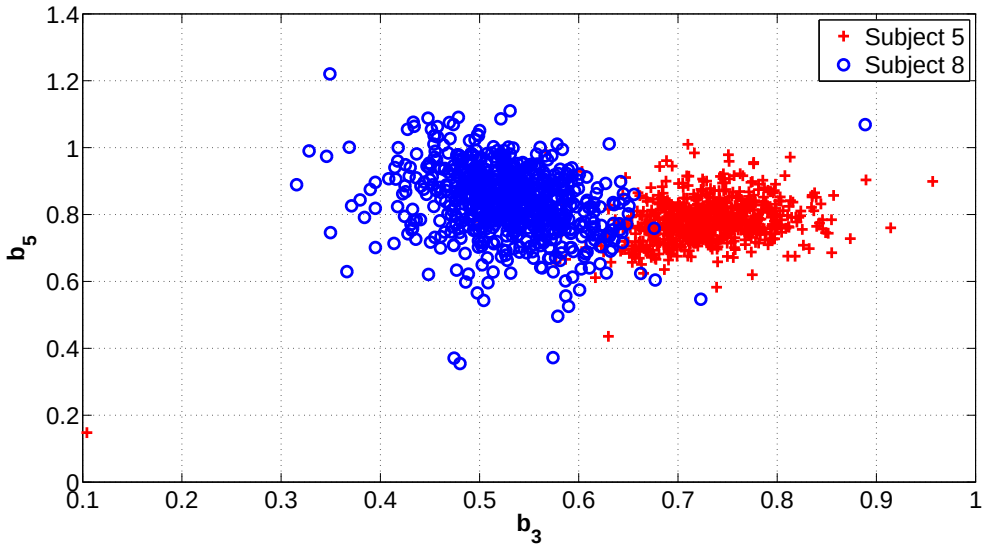
classification performance of the proposed method will be affected and new training samples should be collected for those subjects.

2.6 Summary

In this chapter, a biometric recognition system based on single-lead ECG signal is developed. A new framework based on multitask learning is proposed. The proposed framework combines feature selection and classifier design into a single learning problem. A fast first-order optimization algorithm is used to estimate the model parameters. The performance of the proposed method is evaluated using a real ECG database.

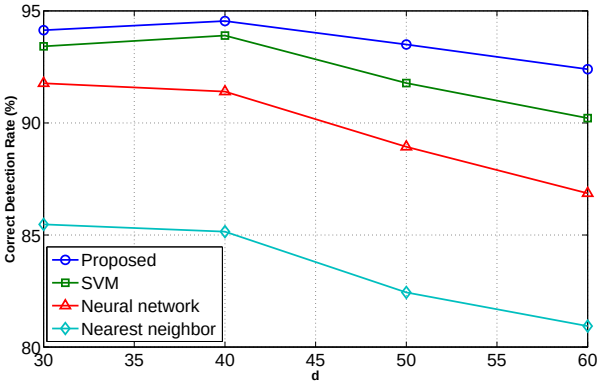


(a)

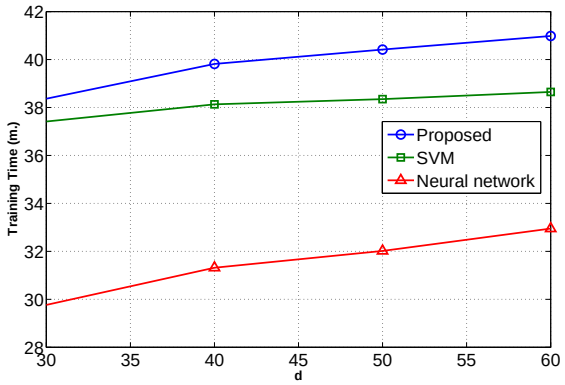


(b)

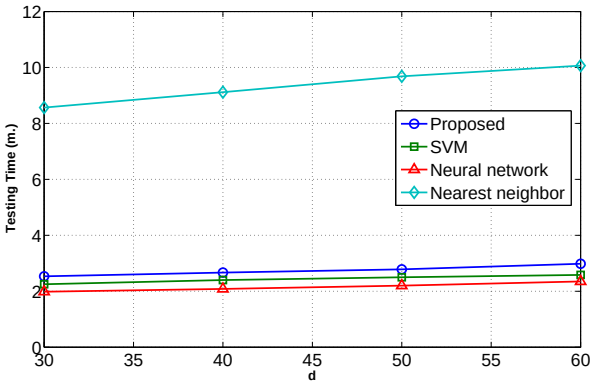
Figure 2.4: Projection of the training data of arbitrary subjects into arbitrary two-dimensional subspaces of the estimated low-rank space.



(a) Correct detection rate



(b) Training time



(c) Testing time

Figure 2.5: Comparison results for different number of non-fiducial features.

CHAPTER 3

REVIEW: MATHEMATICAL MODELING OF CARDIORESPIRATORY SYSTEM

Noncommunicable diseases (NCDs) are the major leading causes of death in the world. There are four main types of noncommunicable diseases: cardiovascular diseases, chronic respiratory diseases, cancers, and diabetes. According to the World Health Organization (WHO) estimates [74]:

- NCDs kill 38 million people each year.
- Around 75% of NCD deaths (about 28 million) occur in low-income and middle-income countries.
- Around 16 million NCD deaths are premature (occurring before the age of 70), and 82% of these deaths occur in low-income and middle-income countries.
- Cardiovascular and respiratory diseases account for most NCD deaths. Around 17.5 million annually due to cardiovascular diseases and 4 million annually due to respiratory diseases.

According to the Centers for Disease Control and Prevention (CDC), around 610,000 people die of heart disease in the United States every year (1 in every 4 deaths) and around 47% of cardiac deaths occur outside a hospital [75]. NCDs also have far-reaching socioeconomic consequences. For example, health care costs associated with the NCDs are forcing millions of people into poverty, especially in low-income countries. The WHO recommends developing a comprehensive approach that requires

different sectors such as health, finance, agriculture, education, planning, etc. to work in conjunction with each other to reduce and control the risk factors associated with the NCDs. Moreover, according to the WHO [74]:

- "NCDs can be reduced through high impact essential NCD interventions that can be delivered through a primary health-care approach to strengthen early detection and timely treatment."

- "Evidence shows that such interventions are excellent economic investments because, if applied to patients early, can reduce the need for more expensive treatment."

- "The greatest impact can be achieved by creating healthy public policies that promote NCD prevention and control and reorienting health systems to address the needs of people with such diseases."

Hence, a smart health monitoring system that can track and monitor the health of the cardiorespiratory system plays a significant role in reducing the number of cardiorespiratory deaths.

Mathematical modeling is a powerful way to study the underlying systems when direct interaction with the system is not possible. Mathematical modeling allows us to accurately capture and quantify the complex interactions between different subsystems, physiological processes, and control mechanisms. It can help to detect certain diseases which alter the normal system function. Mathematical models of the cardiorespiratory system can greatly help to provide more information for a physician to diagnose a cardiorespiratory disease. In this chapter, we will review several mathematical models of the cardiovascular and respiratory systems that can be useful to our studies. The blood circulation in the cardiorespiratory system is shown in Figure 3.1.[1] Mathematical models of the cardiorespiratory system mainly focus on

[1]Source: http://anatomy-bodychart.us/

modeling the various physiological mechanisms, subsystems, and complex dynamics in the cardiorespiratory circulation system.

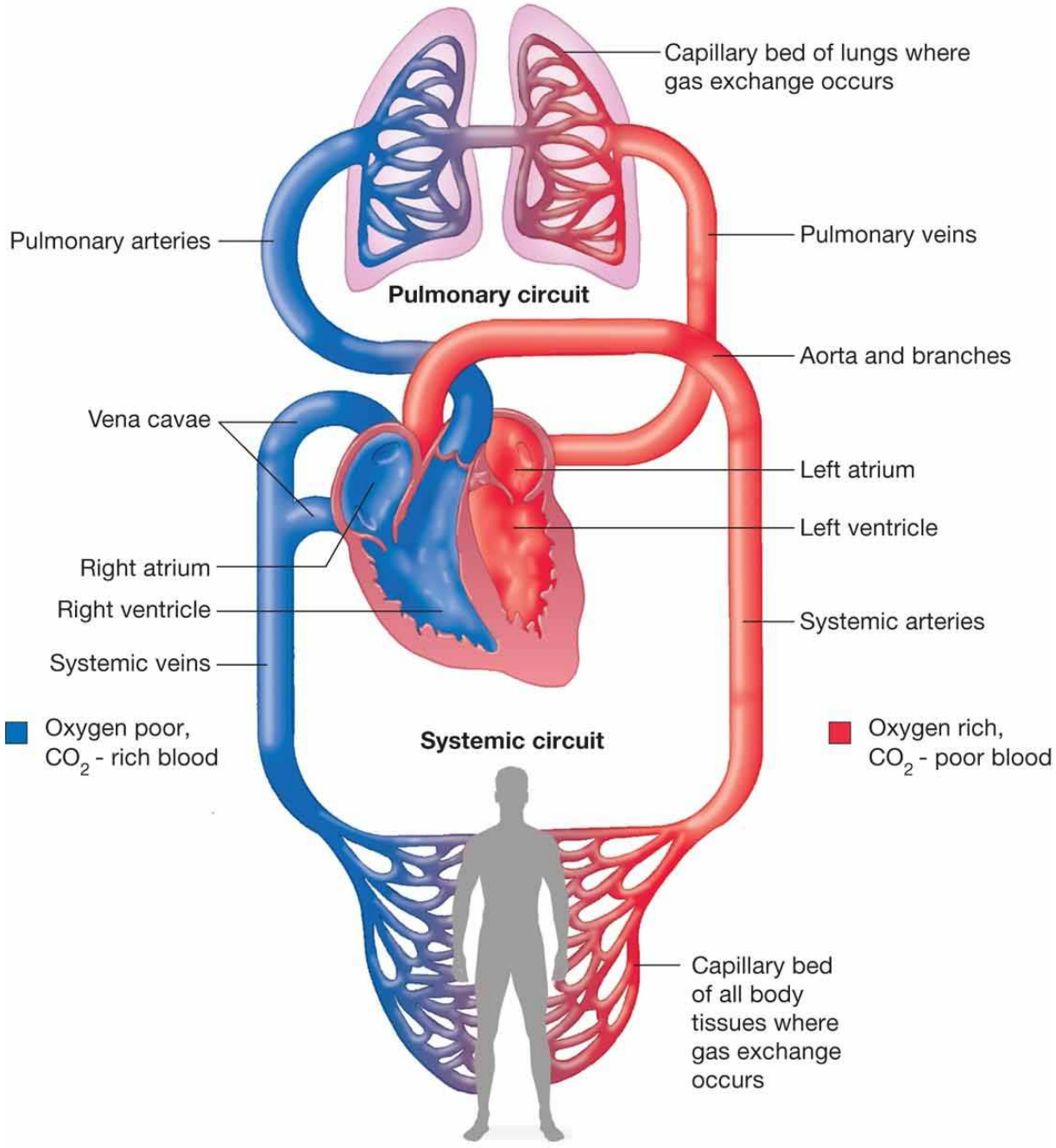
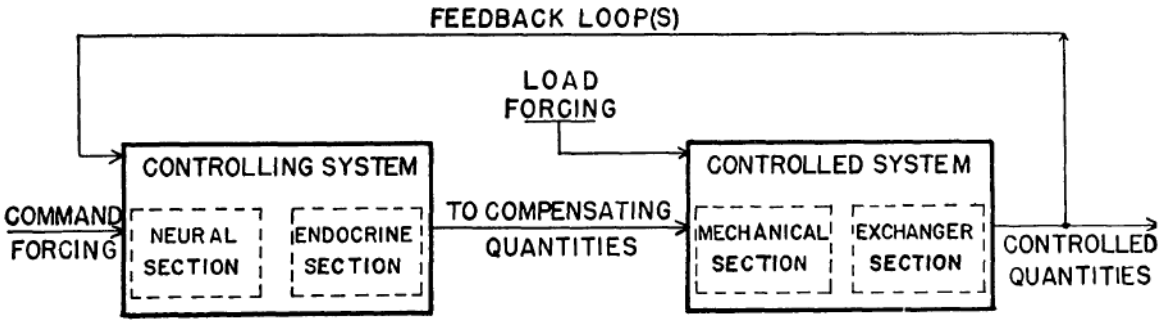


Figure 3.1: Cardiorespiratory circulation system.

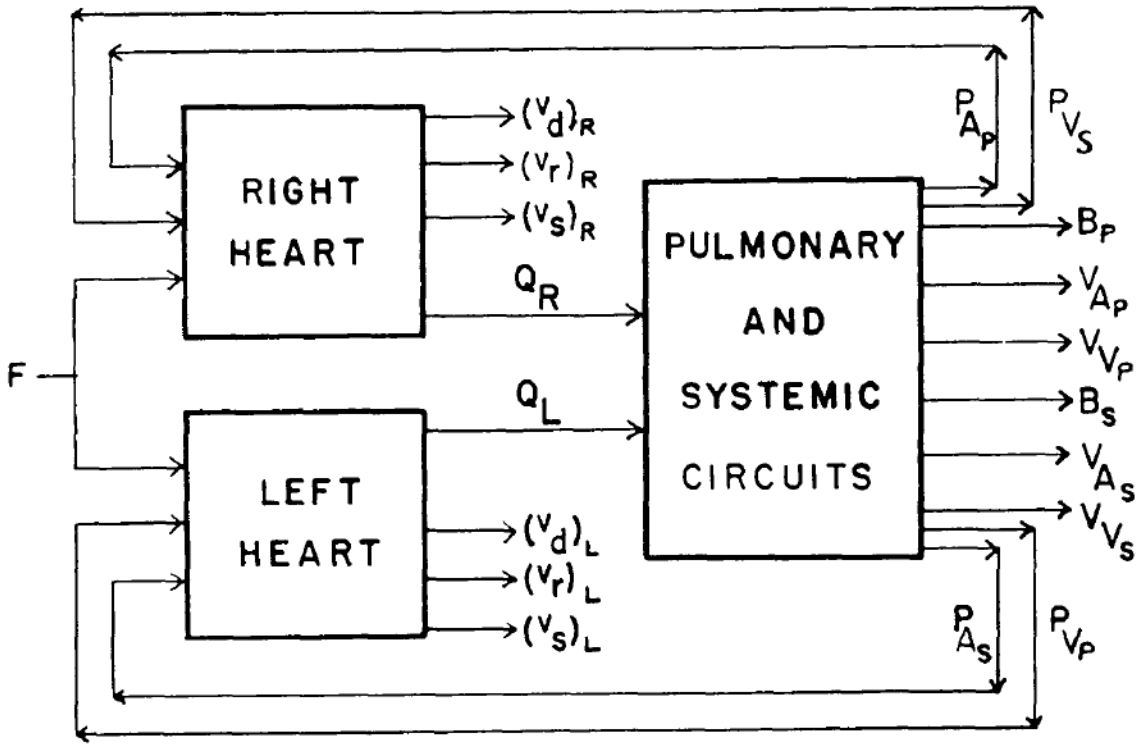
3.1 Cardiovascular System Models

Mathematical modeling of the cardiovascular physiology can be traced back to mid-twentieth century works of Grodins [2] and Guyton *et al.* [3]. In [2], the cardiovascular system is modeled and analyzed using a feedback regulator. The cardiovascular system is divided into two subsystems: a controlling system and a controlled system. The controlling system consists of medullary cardiac and vasomotor centers, and endocrine glands which operate on the heart and blood vessels. The controlled system consists of mechanical and gas exchange elements. Each block of this cardiovascular regulator system is defined using the corresponding input-output relationships (using certain physical laws). Each block is represented using an electrical system consisting of resistances, compliances, etc. The system model is shown in Figure 3.2.

In [3], system analysis of the circulatory regulation is developed. The circulatory system is divided into 18 major subsystems like circulatory dynamics, capillary membrane dynamics, pulmonary dynamics, vascular stress relaxation, etc. Each subsystem is further divided into several small blocks, each representing a particular function like blood flow in aorta, venous resistance, sympathetic stimulation, etc. There are about 354 blocks in the developed system model. Each block is described by a set of mathematical equations derived using the physical laws. The developed circulatory system model is shown in Figure 3.3.



(a) Cardiovascular regulator



(b) Cardiovascular mechanical system

Figure 3.2: Cardiovascular system model [2].

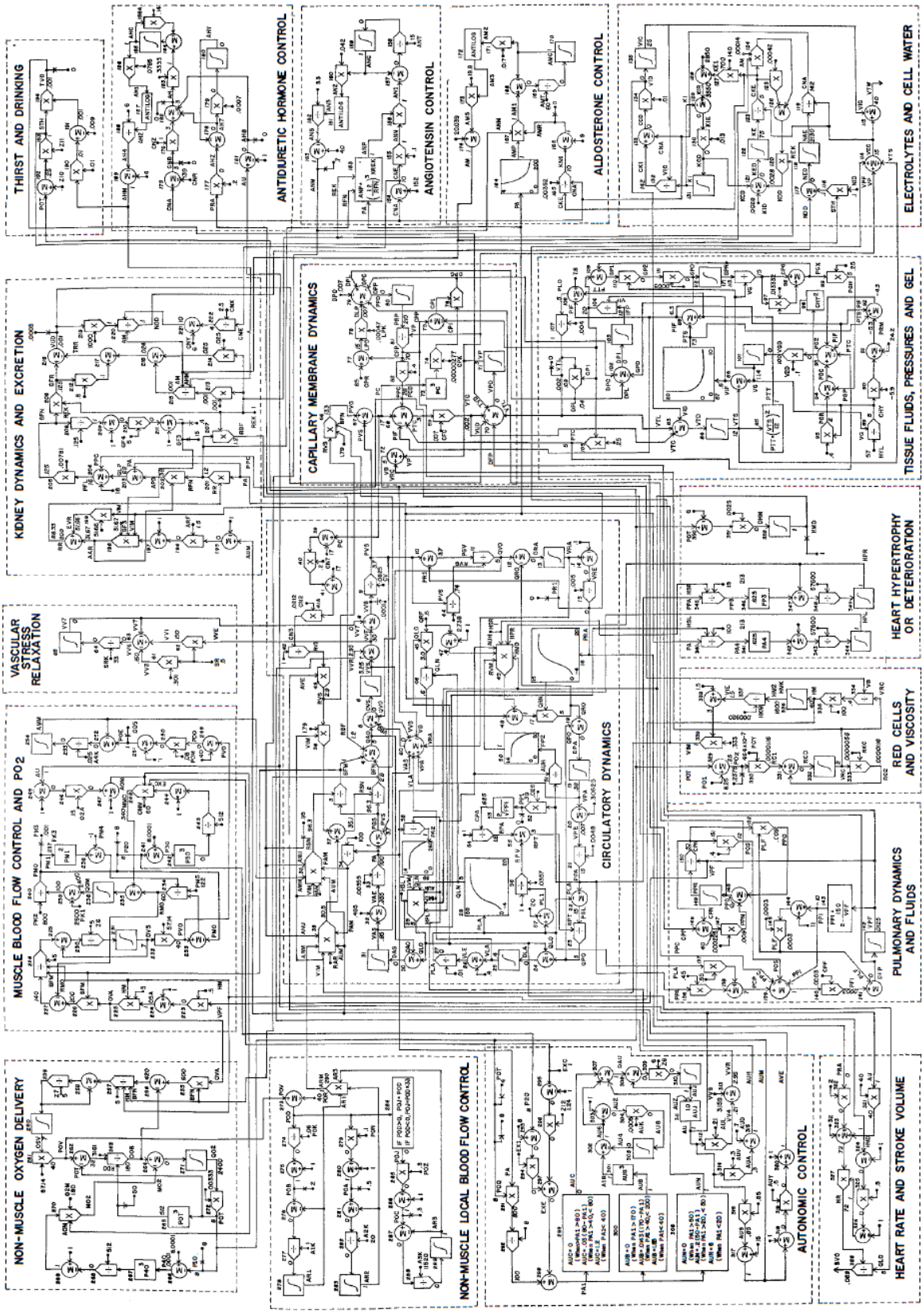


Figure 3.3: Circulatory system model [3].

Kappel and Peer develop a model of the response of the cardiovascular system to constant workload on a person after a period of complete rest [4]. Under these conditions, the cardiovascular system responds by increasing the heart rate in order to increase blood flow and oxygen supply to the muscles. In this process, the baroreceptor control loop plays a central role. This is one the local control systems (auto regulation mechanism of the cardiovascular system). This local control mechanism is modeled in [4], using the four compartment model of the cardiovascular system. The cardiovascular system is divided into four basic compartments: arterial systemic part, venous systemic part, arterial pulmonary part, and venous pulmonary part. These four compartments are connected by the left and right ventricles, and peripheral regions. The system model is shown in Figure 3.4. The variation of heart rate with respect to time is considered as the control input variable (u). The pressures and volumes in different parts of this system are considered as state variables. The main task of the baroreceptor control loop is to stabilize the arterial systemic pressure. This is mathematically formulated as an optimal control problem, where the goal is to find the control input u that minimizes a quadratic cost function.

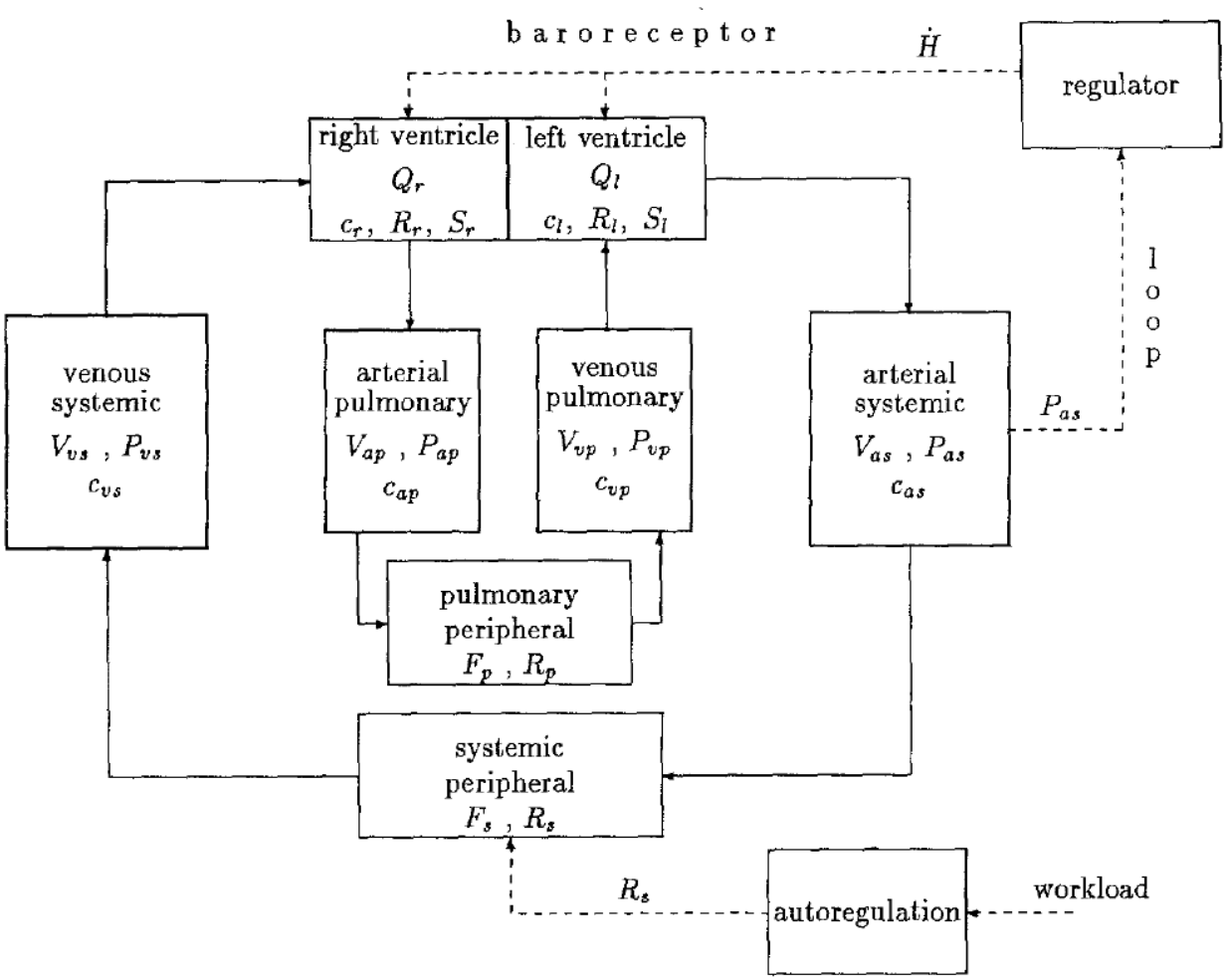
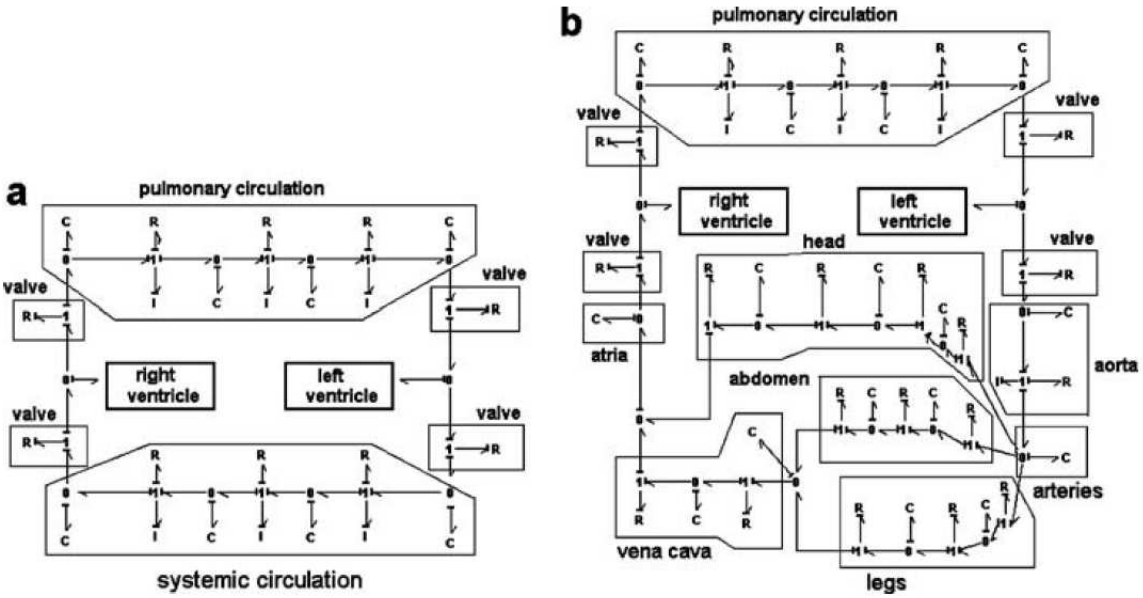


Figure 3.4: Cardiovascular system model [4].

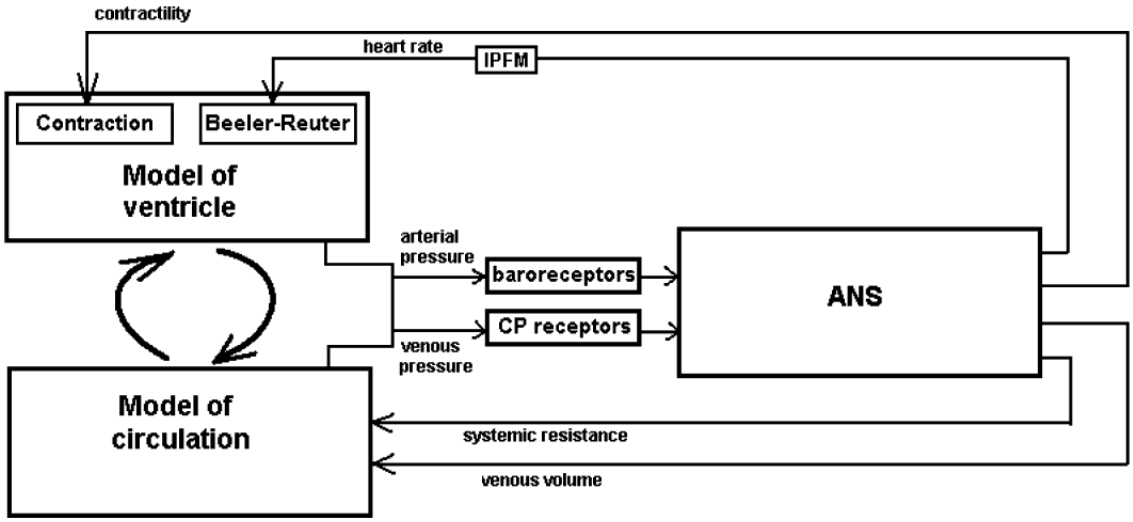
In [76], Batzel *et al.* model the cardiovascular-respiratory control system. Similar to [4], the cardiovascular and respiratory control systems are modeled by a linear negative feedback control which minimizes a quadratic cost function (denoting an optimal system performance). The model is used to estimate several parameters (like heart rate, pulmonary resistance, etc.) in the case of a congestive heart failure. A person going from quiet awake state to stage 4 non-REM sleep is considered in the simulated experiments.

In [5], the interaction between the autonomic nervous system (ANS) and the cardiovascular system (CVS) is modeled using bond graph formalism. A bond graph

is used to represent multi-energy systems using power exchanges. The proposed model takes into account the three important subsystems: cardiac ventricular activity, circulatory system, and autonomic modulation (baroreflex loop). These subsystems are modeled using the bond graphs, as shown in Figure 3.5.



(a) Pulmonary and systemic circulation models



(b) Coupling between the ANS model and the models of the ventricles and the circulatory system

Figure 3.5: Bond graph model [5].

In [6], a closed-loop computational model of the cardiovascular system and its interaction with the autonomic nervous system during the Valsalva maneuver is developed. The cardiovascular system is represented by an electric analog circuit model with 42 compartments. The lumped parameter model of each compartment represent the dynamics of different hemodynamic variables of the cardiovascular system. The model diagram is shown in Figure 3.6.

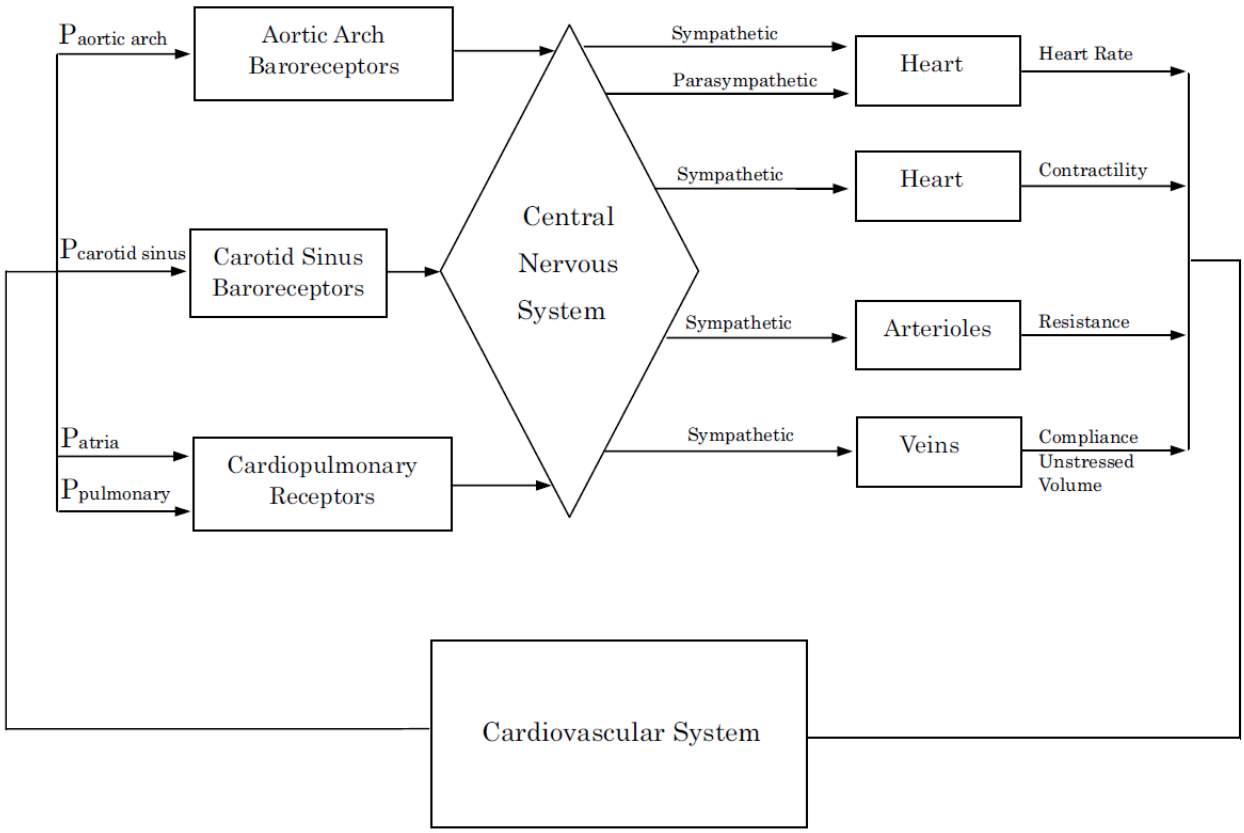
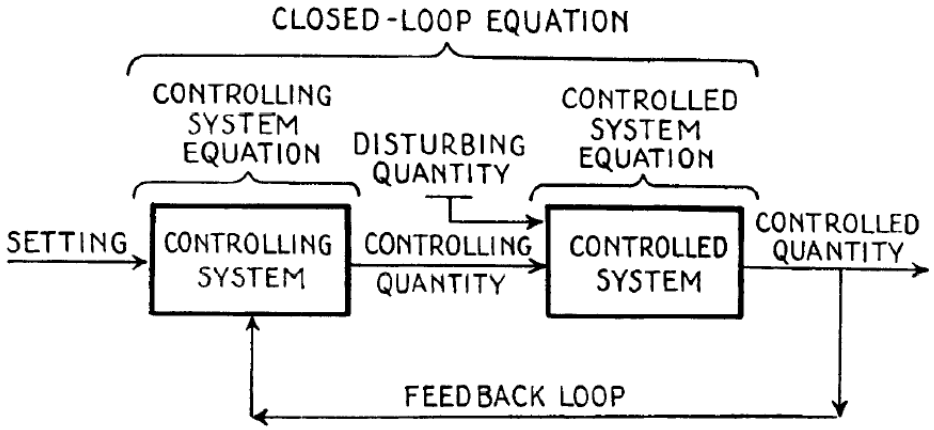


Figure 3.6: Autonomic nervous system and its interaction with the the cardiovascular system [6].

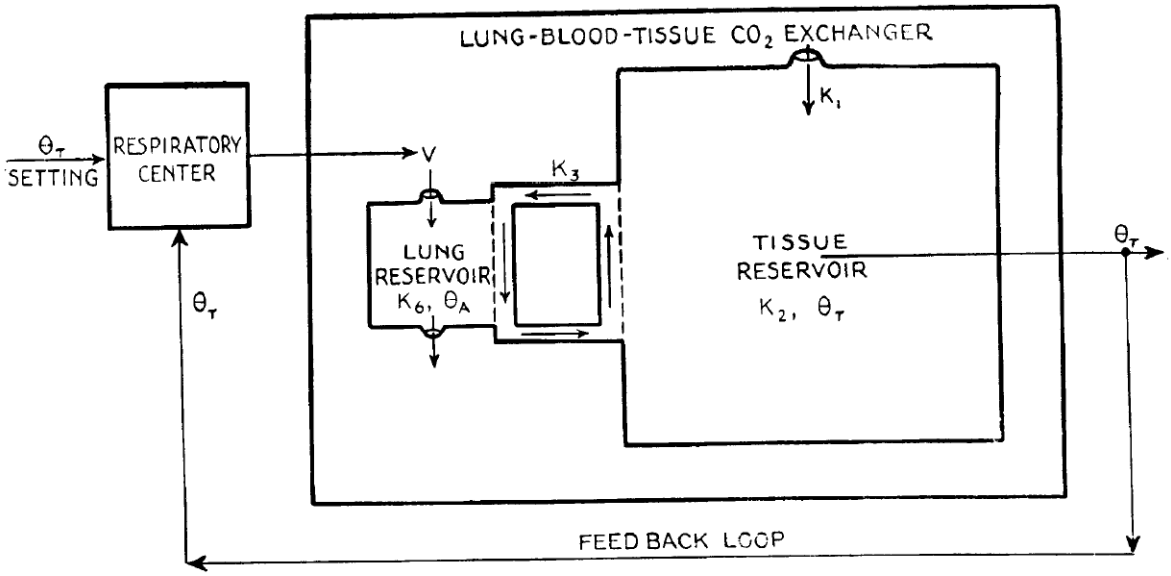
3.2 Respiratory System Models

Mathematical modeling of the respiratory physiology began with early works of Grodins *et al.* [7]. In [7], the respiratory system is modeled using a closed-loop feedback regulator in which CO_2 concentration controls the ventilation and the venti-

lation regulates the CO_2 concentration. Thus, the pulmonary ventilation acts as the controlling system, the CO_2 concentration is the controlled quantity, and a negative feedback exists between these two. An electric analog circuit model is developed for the pulmonary ventilation system. The system model is shown in Figure 3.7.



(a) Feedback regulator



(b) Closed-loop model

Figure 3.7: Respiratory system model [7].

A lumped parameter model of the respiratory control system is developed in [8]. The system contains three major compartments: lungs, brain, and tissue. The lung-

blood-tissue gas transport and exchange system is represented by a set of differential equations with time delays. A control function is defined to monitor the concentrations of different chemical quantities at carotid chemoreceptors. The developed control system model is shown in Figure 3.8.

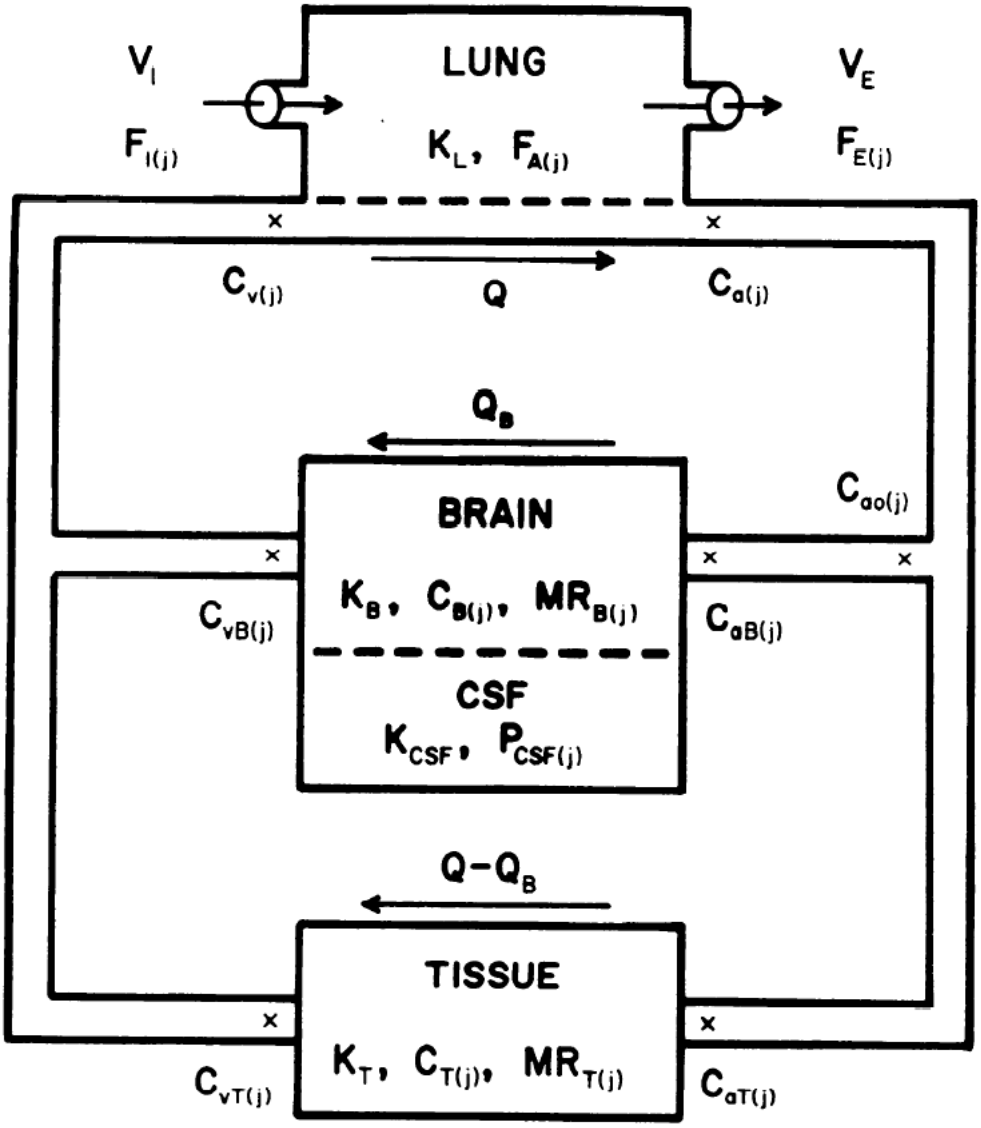


Figure 3.8: Respiratory control system model [8].

A mathematical model describing the pressure-flow relationship in the ventilatory system under conditions of constant lung volume is developed in [9]. A lumped parameter model of the ventilatory system is developed, as shown in Figure 3.9.

An electrical analog of the model is developed, and the airway dynamics equations are derived using the Kirchhoff's circuit laws. A least squares parameter estimation technique is used to adjust the model parameter values to fit the data from human subjects.

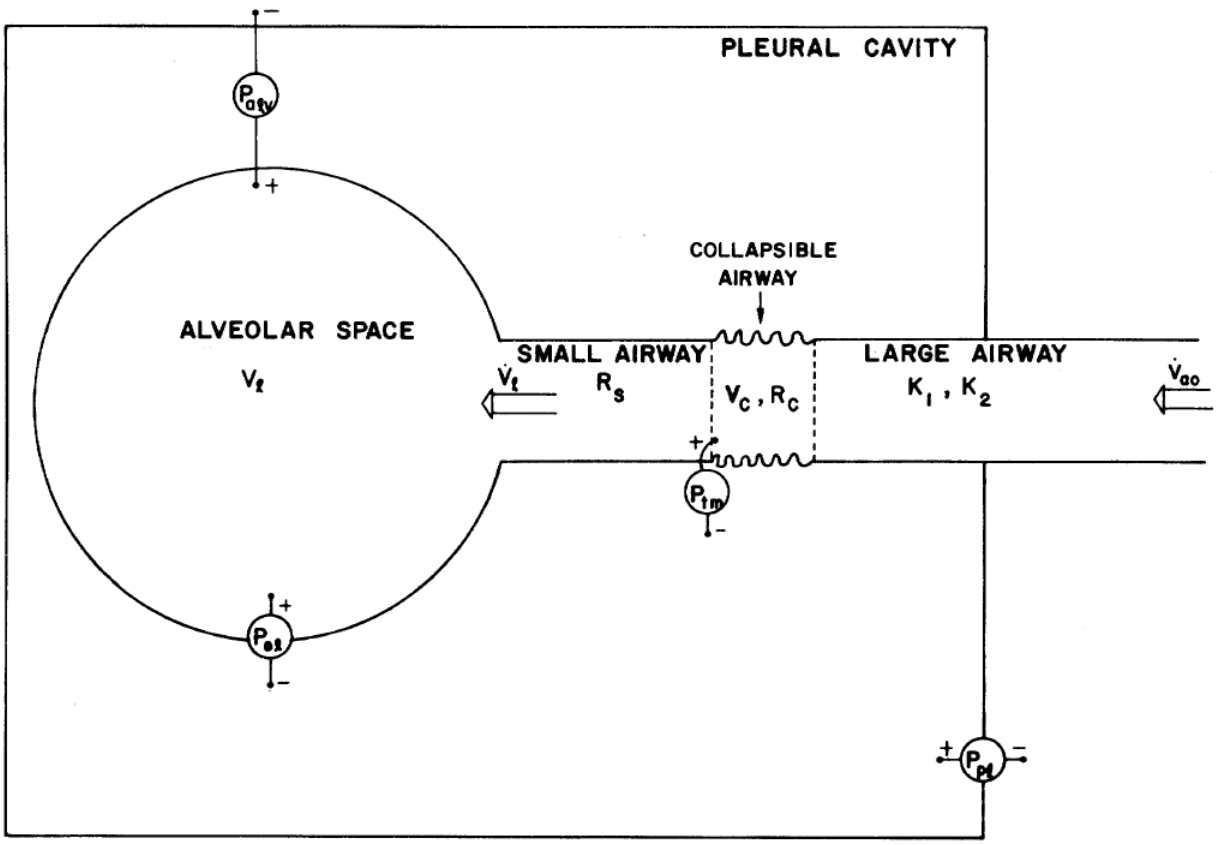


Figure 3.9: Ventilatory system lumped model [9].

A model defining the relation between respiratory neural and mechanical outputs is described in [77]. A model for the inspiratory muscle activity during breathing is developed, which is then used to calculate the respiratory volume and flow from different neural output profiles. In the model, different factors such as the relation between inspiratory activity and isometric pressure, muscle reaction time, relation between respiratory volume and pressure output, relation between respiratory flow and pressure output, expiratory muscle pressure and flow, are considered.

In [10], a model of the respiratory control system is developed. The model describes the gas exchange between pulmonary blood, tissue capillary blood, venous blood and tissue compartments. The state variables are the concentrations of carbon dioxide, acid, hemoglobin, bicarbonate, and a generic buffer. The proposed model has 23 ordinary differential equations (ODEs) containing terms of blood flow, gas diffusion, and chemical reactions. The ODEs have the standard form:

$$\dot{\mathbf{y}} = f(\mathbf{y}, \boldsymbol{\beta}),$$

$$\mathbf{y}(t_0) = \boldsymbol{\eta},$$

where $\boldsymbol{\beta}$ are the model parameters, and $\boldsymbol{\eta}$ are the initial values. Sensitivity analysis is performed to measure the effect of the model parameters on the respiratory control system. The model parameters and initial values are estimated by minimizing a quadratic error function between the estimated and observed trajectories of the system over a period of time. The developed system model is shown in Figure 3.10.

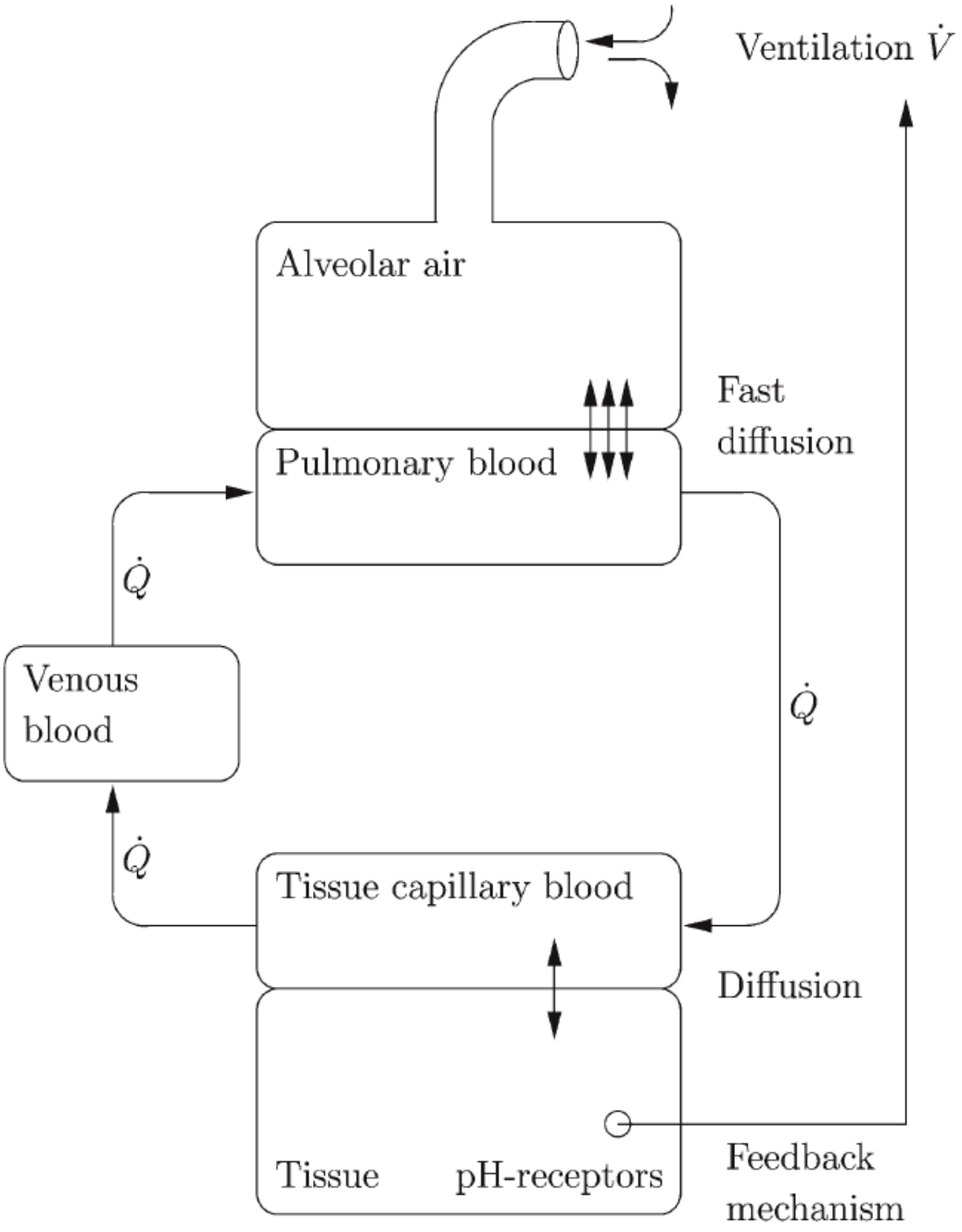


Figure 3.10: Respiratory system model [10].

Integrated Model

An integrated model of the cardiovascular and respiratory systems is constructed in [11]. This model is constructed by combining several existing models of the cardiovascular system, respiratory system, central neural control system, and sleep mechanism system. The model can simulate different conditions such as normal sleep-wake cycle, obstructive sleep apnea (OSA), Cheyne-Stokes respiration with CSA, hypoxia-induced periodic breathing, etc. and generate the corresponding physiological signals. The model has a total of 80 states and 472 parameters. The block diagram of this integrated model is shown in Figure 3.11. A Simulink model of this is also developed in [11]. The Simulink models of different parts of the system are shown in Figures 3.12 to 3.16.

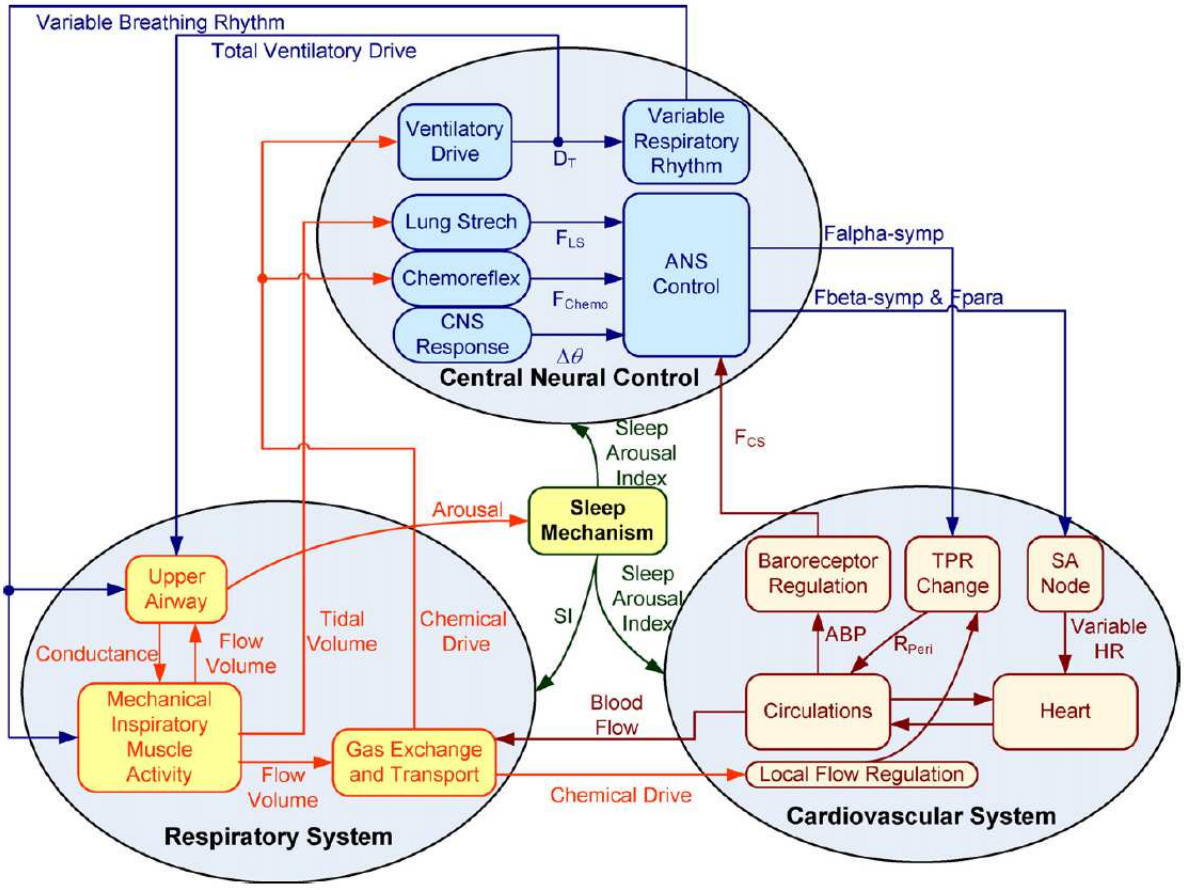


Figure 3.11: Block diagram of the integrated system model in [11].

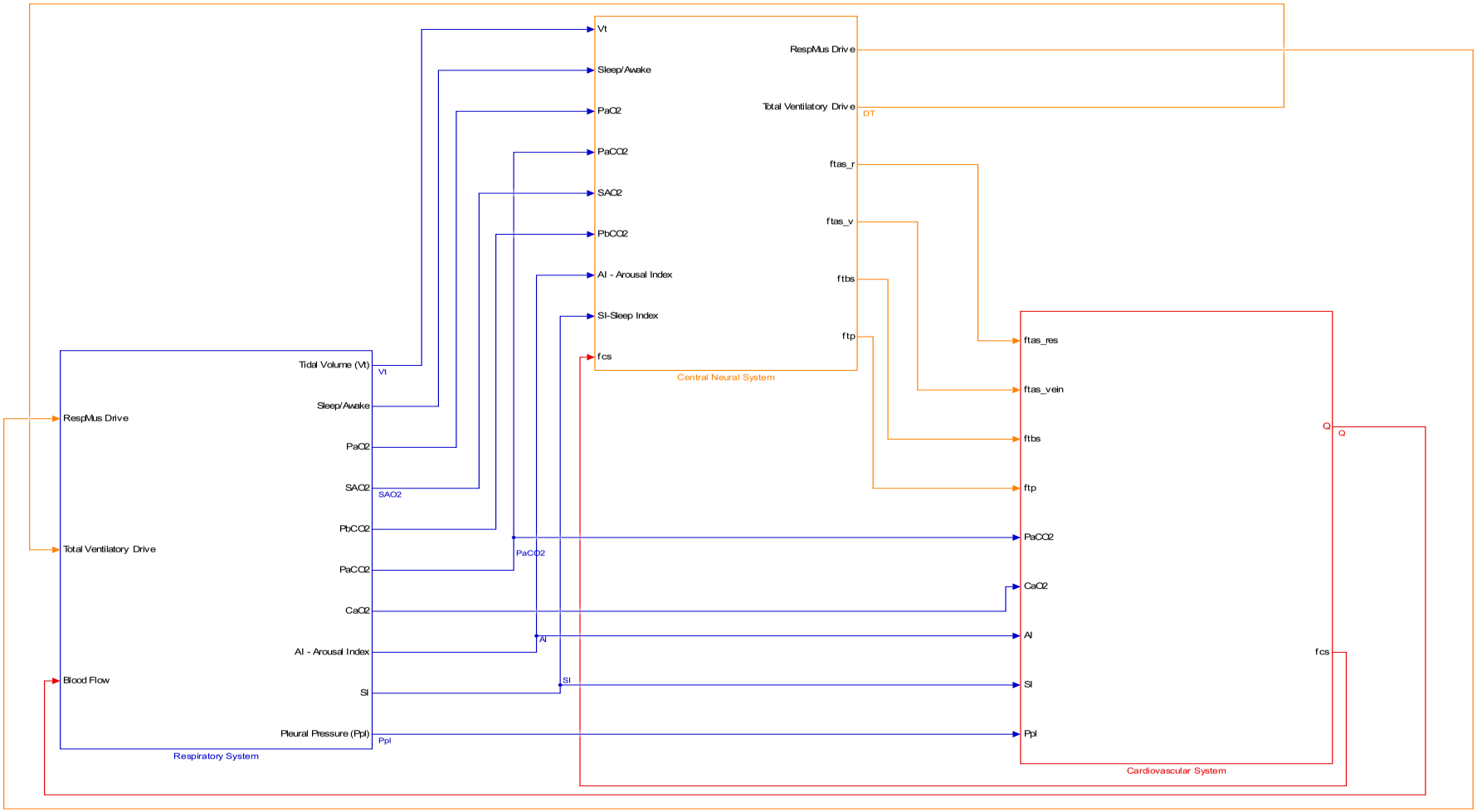


Figure 3.12: Simulink model of the integrated system [11].

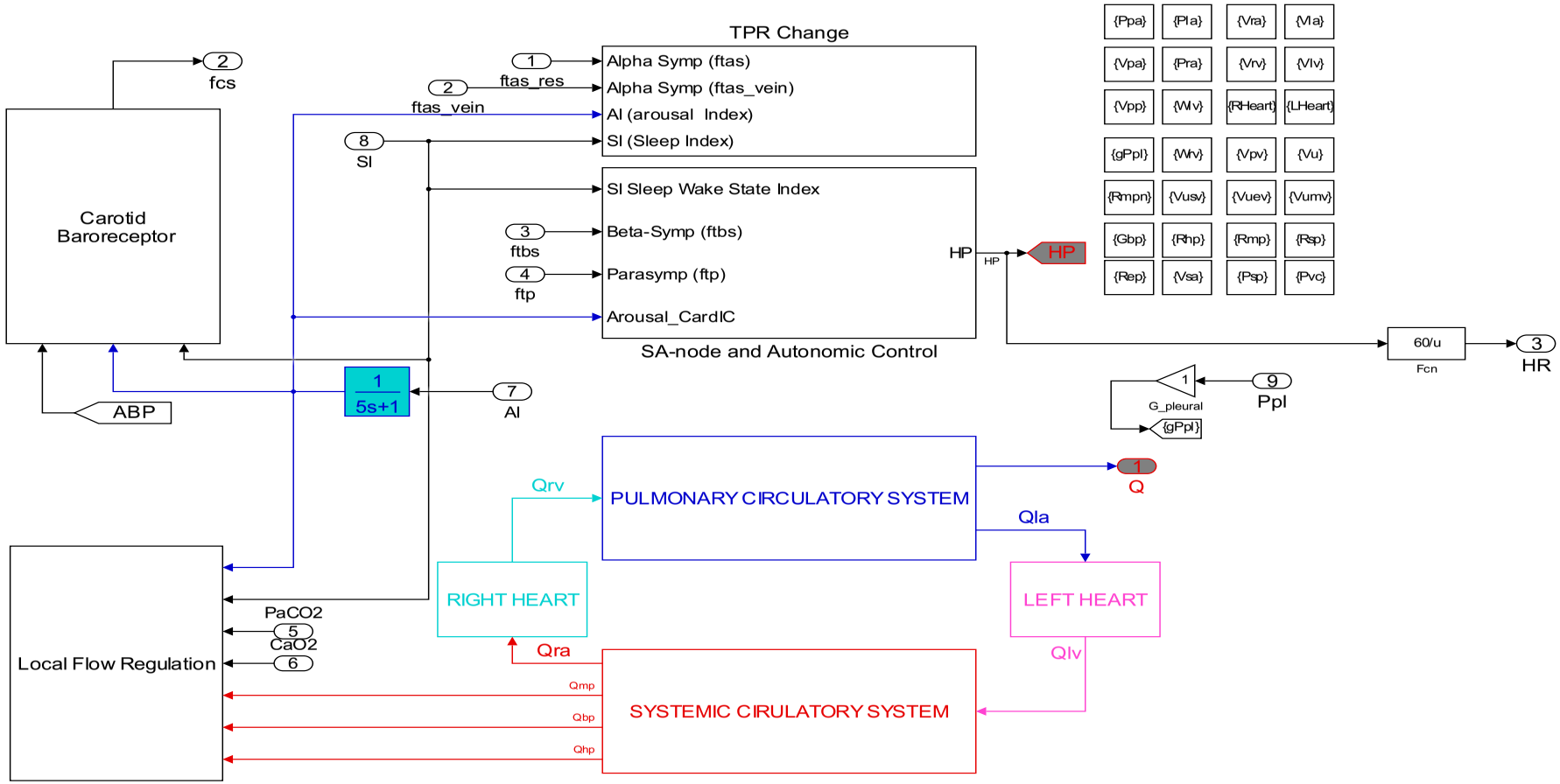


Figure 3.13: Simulink model of the cardiovascular system part [11].

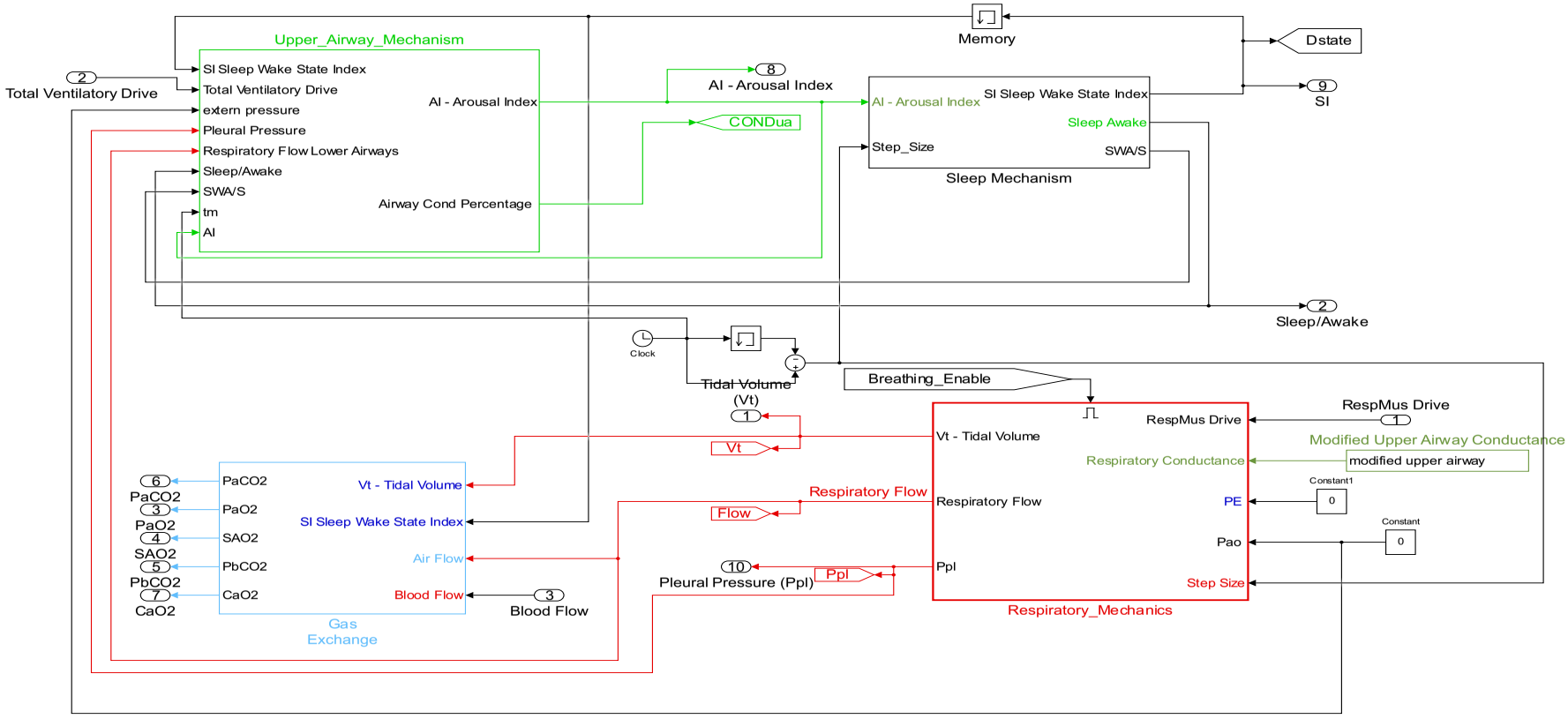


Figure 3.14: Simulink model of the respiratory system part [11].

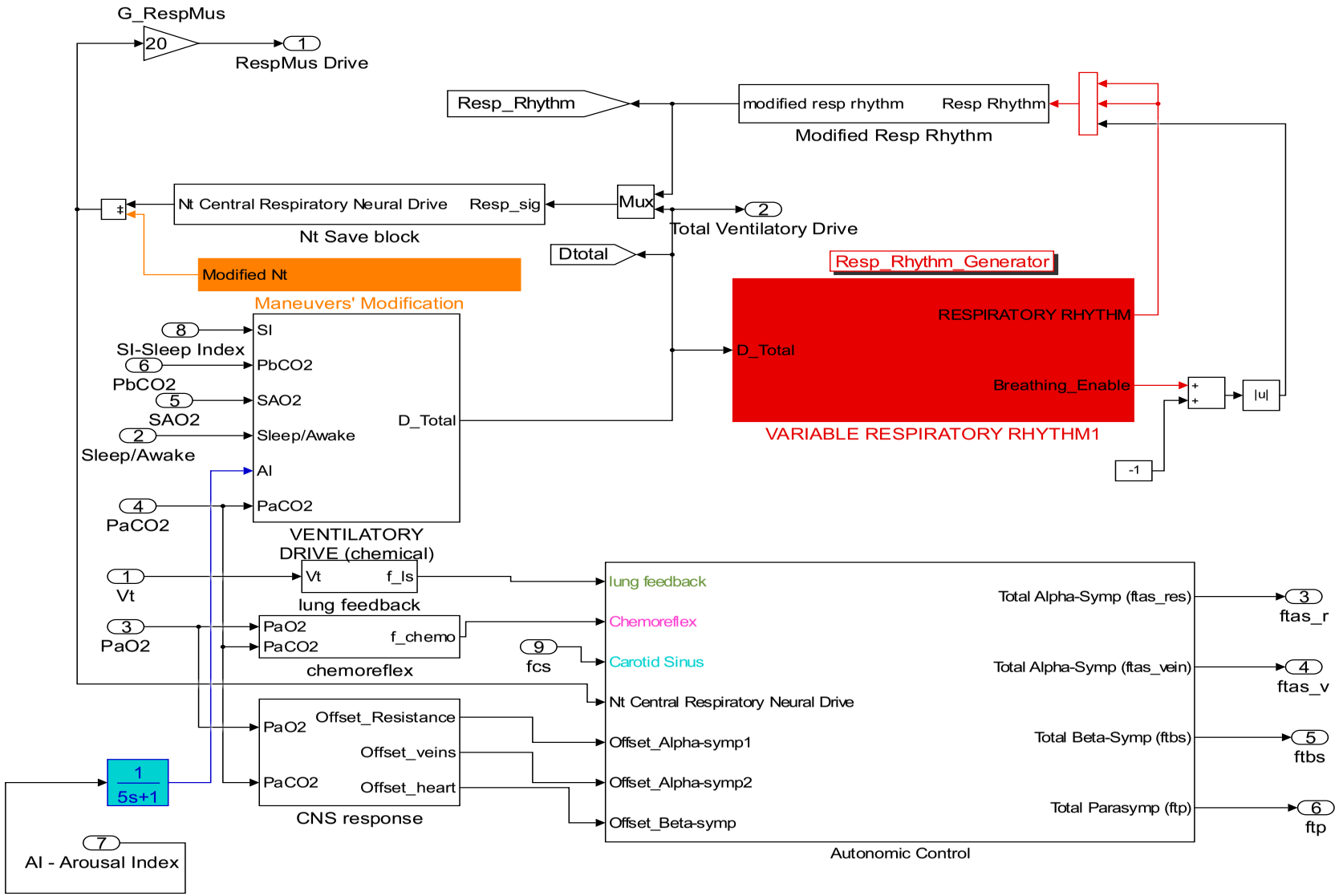


Figure 3.15: Simulink model of the central neural system part [11].

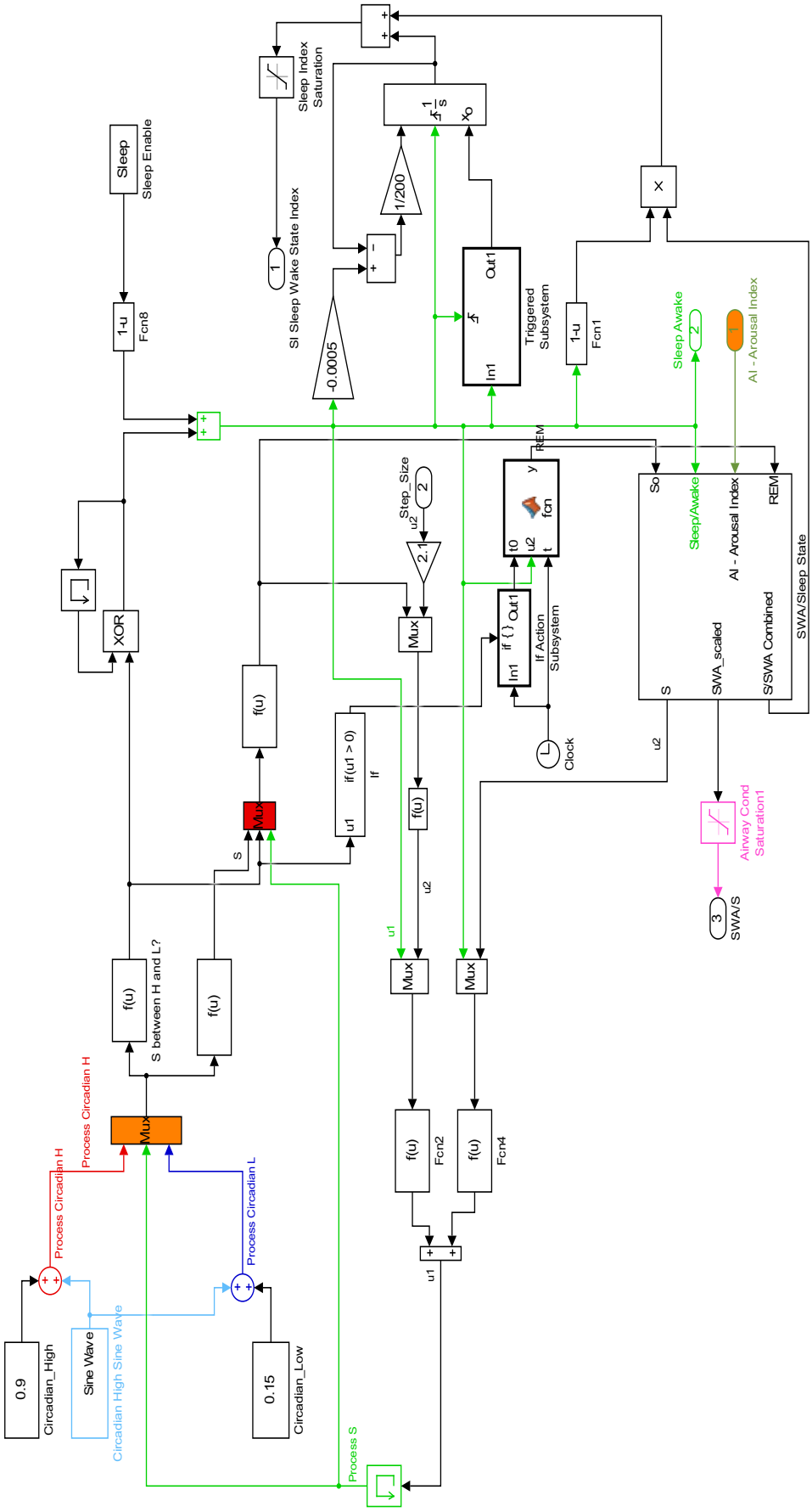


Figure 3.16: Simulink model of the sleep mechanism part [11].

3.3 Summary

In this chapter, we have reviewed several mathematical models of the cardiovascular and respiratory systems. Each model focuses on certain aspects of the cardiorespiratory system and tries to model the underlying physiological processes, mechanisms, and interactions between different subsystems. These models can be simulated using computer programs, and thus can be used to perform some virtual experiments and simulations. In the next chapters, we will use some of these mathematical models for modeling and statistical inference.

CHAPTER 4

CONTROL MECHANISM MODELING OF CARDIORESPIRATORY SYSTEM

Control mechanisms are vital for maintaining the homeostasis (stable equilibrium) in the human body. In this chapter, we focus on modeling the control mechanism of the cardiorespiratory system during sleep transition stage. The control mechanism model can be helpful in detecting certain sleep disorders such as difficulty initiating sleep and insomnia.

Specifically, we model the local control mechanism of the cardiorespiratory system as the human body goes from awake state to stage 4 non-REM sleep state. The transition from awake state to stage 4 non-REM sleep state can be modeled by stabilizing certain states of the cardiorespiratory system. Since the cardiorespiratory system model is nonlinear, solving the corresponding optimal control problem is hard.

In [76], Batzel *et al.* solve this continuous-time optimal control problem by linearizing the system at the final sleep state. However, linearizing the system at just one point is not optimal. Moreover, it is difficult to know the final steady state (sleep) values of all the states in practice. Here, we propose an iterative algorithm to solve the optimal control. We initially start with a nominal state and input sequences, and iteratively update these sequences to get the final optimal sequences. In each iteration, the system is linearized with the sequences obtained from the previous iteration. Using the linearized system, we formulate the optimal control problem as a convex optimization problem and efficiently solve it using interior-point methods.

4.1 Cardiorespiratory System Model

We adopt the cardiovascular-respiratory system model in [76]. The model is described by 13 ordinary differential equations as follows,

$$
\begin{aligned}
V_{ACO_2}\,\dot{P}_{aCO_2}(t) &= 863\,F_p(t)\,(C_{vCO_2}(t-\tau) - C_{aCO_2}(t)) \\
&\quad + \dot{V}_A(t)\,(P_{ICO_2} - P_{aCO_2}(t)), \\
V_{AO_2}\,\dot{P}_{aO_2}(t) &= 863\,F_p(t)\,(C_{vO_2}(t-\tau) - C_{aO_2}(t)) \\
&\quad + \dot{V}_A(t)\,(P_{IO_2} - P_{aO_2}(t)), \\
V_{TCO_2}\,\dot{C}_{vCO_2}(t) &= MR_{CO_2} \\
&\quad + F_s(t)\,(C_{aCO_2}(t-\tau) - C_{vCO_2}(t)), \\
V_{TO_2}\,\dot{C}_{vO_2}(t) &= -MR_{O_2} \\
&\quad + F_s(t)\,(C_{aO_2}(t-\tau) - C_{vO_2}(t)), \\
c_{as}\,\dot{P}_{as}(t) &= Q_l(t) - F_s(t), \\
c_{vs}\,\dot{P}_{vs}(t) &= F_s(t) - Q_r(t), \\
c_{vp}\,\dot{P}_{vp}(t) &= F_p(t) - Q_l(t), \\
\dot{S}_l(t) &= \sigma_l(t), \\
\dot{S}_r(t) &= \sigma_r(t), \\
\dot{\sigma}_l(t) &= -\gamma_l\,\sigma_l(t) - \alpha_l\,S_l(t) + \beta_l\,H(t), \\
\dot{\sigma}_r(t) &= -\gamma_r\,\sigma_r(t) - \alpha_r\,S_r(t) + \beta_r\,H(t), \\
\dot{H}(t) &= u_1(t), \\
\ddot{V}_A(t) &= u_2(t),
\end{aligned}
\tag{4.1}
$$

where

$$
\begin{aligned}
P_{ap}(t) &= \frac{V_0 - c_{as}\, P_{as}(t) - c_{vs}\, P_{vs}(t) - c_{vp}\, P_{vp}(t)}{c_{ap}}, \\
F_s(t) &= \frac{P_{as}(t) - P_{vs}(t)}{R_s(t)}, \\
R_s(t) &= A_{pesk}\, C_{VO_2}(t), \\
F_p(t) &= \frac{P_{ap}(t) - P_{vp}(t)}{R_p}, \\
Q_l(t) &= H(t)\, V_{str}(t) = H(t)\, S_l(t)\, \frac{c_l\, P_{vp}(t)}{P_{as}(t)}, \\
Q_r(t) &= H(t)\, V_{str}(t) = H(t)\, S_r(t)\, \frac{c_r\, P_{vs}(t)}{P_{ap}(t)}, \\
C_{aCO_2}(t) &= K_{CO_2}\, P_{aCO_2}(t) + k_{CO_2}, \\
C_{aO_2}(t) &= K_1\, \left(1 - e^{-K_2\, P_{aO_2}(t)}\right)^2.
\end{aligned}
\tag{4.2}
$$

The block diagram of the 13-state model is shown in Figure 4.1. The system states and parameters are described in Table 4.1 and Table 4.2. The delay (τ) in the differential equations represents the transport delay between the cardiovascular and respiratory systems.

Table 4.1: Cardiovascular system states and parameters

P_{as}	Mean arterial blood pressure in systemic part
P_{vs}	Mean venous blood pressure in systemic part
P_{ap}	Mean arterial blood pressure in pulmonary part
P_{vp}	Mean venous blood pressure in pulmonary part
Q	Cardiac output
R	Peripheral resistance
S	Ventricle contractility
V_{str}	Stroke volume
V_0	Total blood volume
H	Heart rate
F	Blood flow
c_a	Arterial compliance
c_v	Venous compliance
l, r	subscripts for left and right sides of the heart
s, p	subscripts for systemic and pulmonary parts
a, v	subscripts for arterial and venous circuits

Table 4.2: Respiratory system states and parameters

P_a	Partial pressure of arterial blood gas
P_v	Partial pressure of venous blood gas
P_I	Partial pressure of inspired gas
C_a	Concentration of arterial blood gas
C_v	Concentration of venous blood gas
MR	Metabolic production rate
V_A	Lung gas storage volume
$\dot{V}_A$	Ventilation rate
V_T	Tissue gas storage volume
τ	Transport delay
O_2, CO_2	subscripts for oxygen and carbon dioxide

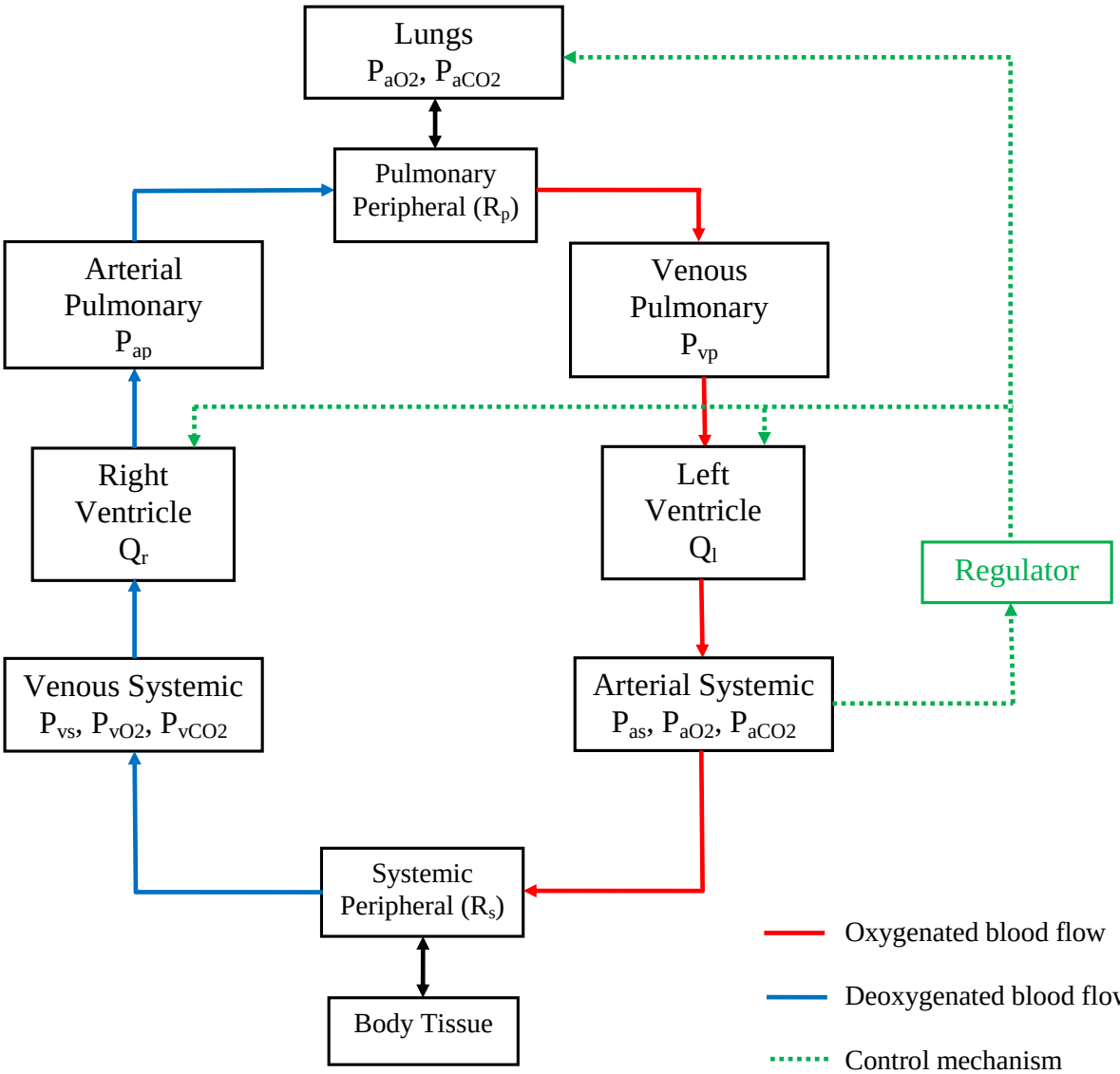


Figure 4.1: Block diagram of the cardiorespiratory system.

4.2 Problem Formulation

Based on the model described above, the cardiorespiratory system model has 13 state variables and 2 control input variables:

$$\begin{aligned}\mathbf{x}(t) = [&P_{aCO_2}(t), P_{aO_2}(t), C_{vCO_2}(t), C_{vO_2}(t), P_{as}(t), P_{vs}(t),\\ &P_{vp}(t), S_l(t), S_r(t), \sigma_l(t), \sigma_r(t), H(t), \dot{V}_A(t)]^T,\\ \mathbf{u}(t) = [&\dot{H}(t), \ddot{V}_A(t)]^T.\end{aligned} \tag{4.3}$$

The transition from awake state to stage 4 non-REM sleep state can be modeled by stabilizing P_{aCO_2}, P_{aO_2}, and P_{as} states of the cardiorespiratory system [76]. Thus, the optimal control problem that transfers the cardiorespiratory system from awake to sleep steady state can be formulated as,

$$\begin{aligned}\mathbf{u}^\star(t) \quad = \arg\min_{\mathbf{u}} \int_{t_0}^{t_f} \big(&q_1(x_1(t) - \bar{x}_1)^2 + q_2(x_2(t) - \bar{x}_2)^2 +\\ &q_5(x_5(t) - \bar{x}_5)^2 + r_1 u_1^2(t) + r_2 u_2^2(t)\big)\ dt,\end{aligned} \tag{4.4}$$

subject to the system model:

$$\begin{aligned}&\dot{x_i}(t) = F_i(\mathbf{x}(t), \mathbf{x}(t-\tau)) + \mathbf{b}_i^T\, \mathbf{u}(t), \quad t \in [t_0, t_f],\\ &\mathbf{x}(t) = \mathbf{x}_0(t), \quad t \in [t_0 - \tau, t_0],\\ &\mathbf{u}(t) \leq \mathbf{0},\end{aligned} \tag{4.5}$$

where $i = 1, 2, ..., 13$, $\bar{x}_i$ is the final steady state value of state i, $\mathbf{x}_0(t)$ is the given initial history, q_i and r_j's are positive coefficients that assign weight to the state and input terms in the above cost function. The inequality constraint, $\mathbf{u}(t) \leq \mathbf{0}$, models the fact that the heart and ventilation rates do not increase when going from awake to sleep state.

We first discretize the system using first-order Euler approximation as follows,

$$\begin{aligned}x_i[k+1] &= g_i\left(\mathbf{x}[k], \mathbf{x}[k-a], \mathbf{u}[k]\right)\\ &= x_i[k] + h\, F_i\left(\mathbf{x}[k], \mathbf{x}[k-a]\right) + h\, \mathbf{b}_i^T\, \mathbf{u}[k],\end{aligned} \tag{4.6}$$

where h is the step size, and the delay $a = \tau/h$. Using this discrete system, we reformulate the optimal control problem in (4.4) as follows,

$$\begin{aligned} \min_{\mathbf{U}} \quad & \sum_{k=0}^{N-1} \left(\mathbf{x}[k] - \bar{\mathbf{x}}\right)^T \mathbf{Q}_0 \left(\mathbf{x}[k] - \bar{\mathbf{x}}\right) + \mathbf{u}[k]^T \mathbf{R}_0 \mathbf{u}[k] \\ & \qquad + \left(\mathbf{x}[N] - \bar{\mathbf{x}}\right)^T \mathbf{Q}_0 \left(\mathbf{x}[N] - \bar{\mathbf{x}}\right) \\ \text{s.t.} \quad & \mathbf{x}[k+1] = \mathbf{g}\left(\mathbf{x}[k], \mathbf{x}[k-a], \mathbf{u}[k]\right), \\ & \mathbf{u}[k] \leq \mathbf{0}, \end{aligned} \tag{4.7}$$

where $N = \frac{t_f - t_0}{h}$, $\mathbf{U} = \{\mathbf{u}[0], \mathbf{u}[1], ..., \mathbf{u}[N-1]\}$ is the optimal control input sequence, $\bar{\mathbf{x}} = [\bar{x}_1, \bar{x}_2, 0, 0, \bar{x}_5, 0, ..., 0]^T$, $\mathbf{Q}_0$ is a 13×13 diagonal matrix with $[q_1, q_2, 0, 0, q_5, 0, ..., 0]^T$ as its main diagonal, and $\mathbf{R}_0$ is a 2×2 diagonal matrix with $[r_1, r_2]^T$ as its main diagonal. Since the matrices $\mathbf{Q}_0$ and $\mathbf{R}_0$ are positive definite, the objective function in (4.7) is quadratic and convex. However, note that the equality constraints in (4.7) are not linear. Also, the system difference equations in (4.6) are not first-order. In the next section, we discuss about solving the optimal control problem (4.7) with these issues.

4.3 Proposed Control Model

We first convert the higher order difference equations in (4.6) into first-order difference equations by augmenting the states $\{\mathbf{x}[k], \mathbf{x}[k-1], ..., \mathbf{x}[k-a]\}$ to construct a new state vector as, $\mathbf{z}[k] = \left[\mathbf{x}[k]^T, \mathbf{x}[k-1]^T, ..., \mathbf{x}[k-a]^T\right]^T$. Note that $\mathbf{z}[k] \in \mathbb{R}^{13(a+1)}$. With this new state vector, the system in (4.6) can be written as follows,

$$\begin{aligned} z_i[k+1] &= f_i\left(\mathbf{z}[k], \mathbf{u}[k]\right) = g_i\left(\mathbf{z}[k], \mathbf{u}[k]\right), \quad i = 1, 2, ..., 13, \\ z_i[k+1] &= f_i\left(\mathbf{z}[k], \mathbf{u}[k]\right) = z_{i-13}[k], \quad i = 14, ..., 13(a+1). \end{aligned} \tag{4.8}$$

We now define the new steady state as $\bar{\mathbf{z}} = [\bar{\mathbf{x}}^T, 0, 0, ..., 0]^T$. The optimal control problem in (4.7) can be reformulated using the new states as follows,

$$\begin{aligned} \min_{\mathbf{U}} \quad & J = \sum_{k=0}^{N-1} \left(\mathbf{z}[k] - \bar{\mathbf{z}}\right)^T \mathbf{Q} \left(\mathbf{z}[k] - \bar{\mathbf{z}}\right) + \mathbf{u}[k]^T \mathbf{R} \mathbf{u}[k] \\ & \qquad + \left(\mathbf{z}[N] - \bar{\mathbf{z}}\right)^T \mathbf{Q} \left(\mathbf{z}[N] - \bar{\mathbf{z}}\right) \\ \text{s.t.} \quad & \mathbf{z}[k+1] = \mathbf{f}\left(\mathbf{z}[k], \mathbf{u}[k]\right), \\ & \mathbf{u}[k] \leq \mathbf{0}, \end{aligned} \tag{4.9}$$

where $\mathbf{R} = \mathbf{R}_0$, and $\mathbf{Q}$ is a block diagonal matrix with the matrices $\{\mathbf{Q}_0, \mathbf{0}, ..., \mathbf{0}\}$ as its main diagonal.

Note that the system, $\mathbf{z}[k+1] = \mathbf{f}\left(\mathbf{z}[k], \mathbf{u}[k]\right)$, is nonlinear. So we cannot directly solve the optimal control problem in (4.9) using convex optimization. The iterative linear quadratic regulator method [78] can be adopted. We assume some nominal sequences $\mathbf{u}_0[k]$ and $\mathbf{z}_0[k]$ are available. The system is first linearized around these nominal sequences as follows,

$$\delta\mathbf{z}[k+1] = \mathbf{A}_k \, \delta\mathbf{z}[k] + \mathbf{B}_k \, \delta\mathbf{u}[k], \tag{4.10}$$

where $\mathbf{A}_k = \nabla_{\mathbf{z}} \mathbf{f}\left(\mathbf{z}_0[k], \mathbf{u}_0[k]\right)$ and $\mathbf{B}_k = \nabla_{\mathbf{u}} \mathbf{f}\left(\mathbf{z}_0[k], \mathbf{u}_0[k]\right)$. Here, $\delta\mathbf{u}[k]$ and $\delta\mathbf{z}[k]$ represent the deviations of the input and state variables from their nominal sequences.

Defining $\delta\bar{\mathbf{z}}[k] = \bar{\mathbf{z}} - \mathbf{z}_0[k]$ and $\delta\bar{\mathbf{u}}[k] = -\mathbf{u}_0[k]$, we reformulate the optimal control problem as a tracking problem as follows,

$$\begin{aligned} \min_{\delta\mathbf{U}} \quad & \tilde{J} = \sum_{k=0}^{N-1} \frac{1}{2} \left(\delta\mathbf{z}[k] - \delta\bar{\mathbf{z}}[k]\right)^T \mathbf{Q} \left(\delta\mathbf{z}[k] - \delta\bar{\mathbf{z}}[k]\right) \\ & \qquad + \frac{1}{2} \left(\delta\mathbf{u}[k] - \delta\bar{\mathbf{u}}[k]\right)^T \mathbf{R} \left(\delta\mathbf{u}[k] - \delta\bar{\mathbf{u}}[k]\right) \\ & \qquad + \frac{1}{2} \left(\delta\mathbf{z}[N] - \delta\bar{\mathbf{z}}[N]\right)^T \mathbf{Q} \left(\delta\mathbf{z}[N] - \delta\bar{\mathbf{z}}[N]\right) \\ \text{s.t.} \quad & \delta\mathbf{z}[k+1] = \mathbf{A}_k \, \delta\mathbf{z}[k] + \mathbf{B}_k \, \delta\mathbf{u}[k], \\ & \mathbf{u}_0[k] + \delta\mathbf{u}[k] \leq \mathbf{0}, \end{aligned} \tag{4.11}$$

where $\delta\mathbf{U} = \{\delta\mathbf{u}[0], \delta\mathbf{u}[1], ..., \delta\mathbf{u}[N-1]\}$.

We use interior-point methods [79] to solve the above convex problem. The above inequality constrained optimization problem is first converted into an equality constrained problem using barrier functions, which is then solved using the Newton's method. After computing the optimal $\delta\mathbf{u}[k]$, we update the nominal input sequence as, $\mathbf{u}[k] = \mathbf{u}_0[k] + \delta\mathbf{u}[k], k = 0, 1, ..., N-1$. To update the nominal state sequence, we simulate the nonlinear system, $\mathbf{z}[k+1] = \mathbf{f}(\mathbf{z}[k], \mathbf{u}[k])$, with the above updated nominal input sequence.

We initially start with a nominal input sequence, $\mathbf{u}_0[k] = \mathbf{0}$. We run the system model with this input sequence to get the initial nominal state sequence $\mathbf{z}_0[k]$. We find the optimal $\delta\mathbf{u}[k]$ as described above, and update the nominal input and state sequences. We repeat this until $||J^{(i)} - J^{(i-1)}||_2 < \epsilon$, where $J^{(i)}$ is the objective function value of ith iteration.

4.4 Experimental Results

We perform simulations of the optimal control of the cardiovascular and respiratory system under the transition from awake to stage 4 non-REM sleep using step size, $h = 1$ sec.

The steady state values of all the cardiovascular and respiratory system states in both awake and sleep stages are determined by running the discrete system in (4.6) for a long period of time with zero control inputs. The calculated steady state values during awake and sleep stages are respectively,

$$\begin{aligned} \bar{\mathbf{x}}_a = [&39.0974, 103.4, 0.5563, 0.1273, 104.5, 3.515, \\ &7.857, 61.54, 4.691, 0, 0, 75, 5.736]^T, \\ \bar{\mathbf{x}}_s = [&51.0767, 89.1, 0.6386, 0.1187, 91.23, 3.788, \\ &7.742, 55.79, 4.253, 0, 0, 68, 4.392]^T. \end{aligned} \tag{4.12}$$

These steady state values are in fact reasonable [76, 80–83]. The transition of the

cardiovascular-respiratory system from awake to sleep state is modeled by stabilizing the states $P_{aCO_2}(t)$, $P_{aO_2}(t)$, and $P_{as}(t)$ to their corresponding sleep steady state values: $\bar{x}_1 = 51.0767$, $\bar{x}_2 = 89.1$, $\bar{x}_5 = 91.23$. We consider $t_f = 40$ minutes.

The final optimal control input sequences of the cardiovascular and respiratory system are shown in Figure 4.2. The corresponding optimal state trajectories are shown in Figures 4.3 to 4.6. From the results, we can see that the optimal control inputs, $\dot{H}(t)$ and $\ddot{V}_A(t)$ (rate of change of heart rate and ventilation rate), stabilize to zero as the cardiovascular-respiratory system goes from awake to sleep state, which is expected. From the results, we can see that the states $P_{aCO_2}(t)$, $P_{aO_2}(t)$, and $P_{as}(t)$ converge to their sleep steady state values. Remaining system states also converge to their sleep steady state values.

To validate our simulation results, we collected real data from a healthy 25-year-old male subject using Hexoskin biometric smart shirt. The measured physiological signals of the subject during awake to sleep transition are shown in Figure 4.7. From the results, we can see that the real heart rate and ventilation rate trajectories are close to the simulated optimal state trajectories. Since the cardiorespiratory system model is not tuned to this specific subject, the real heart rate and ventilation rate values are different from the simulated model state values. However, the control mechanism model captures the cardiorespiratory system dynamics (during awake to sleep transition) in general. Comparisons with real physiological signals show that the control mechanism model can catch the system dynamics of the subject from awake to sleep state.

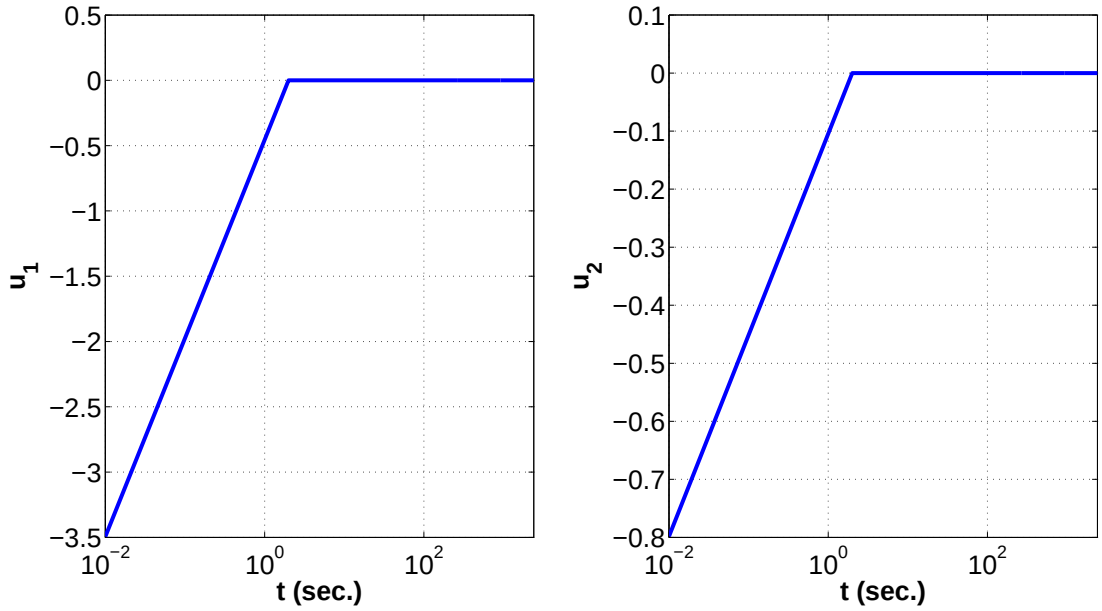


Figure 4.2: Optimal control input sequences of the cardiorespiratory system.

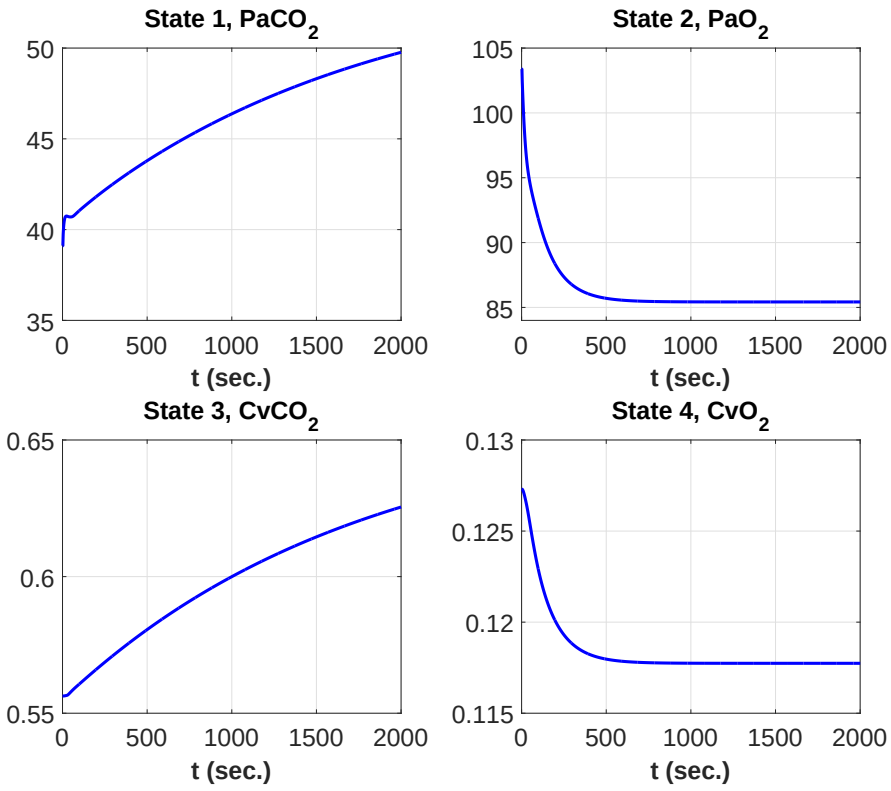


Figure 4.3: Optimal state trajectories of the cardiorespiratory system.

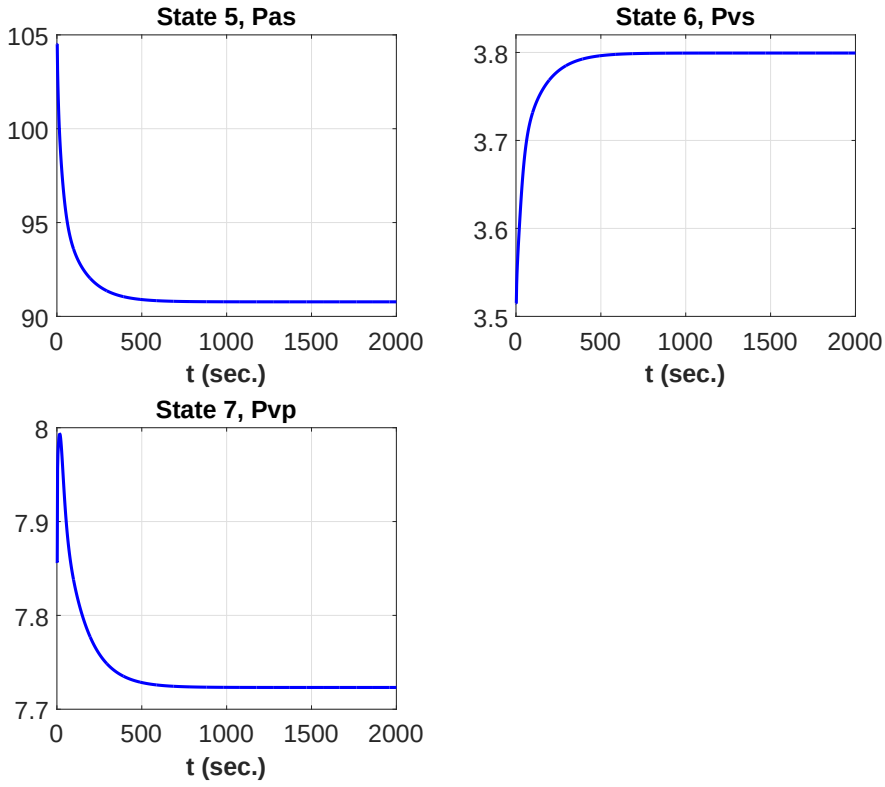


Figure 4.4: Optimal state trajectories of the cardiorespiratory system.

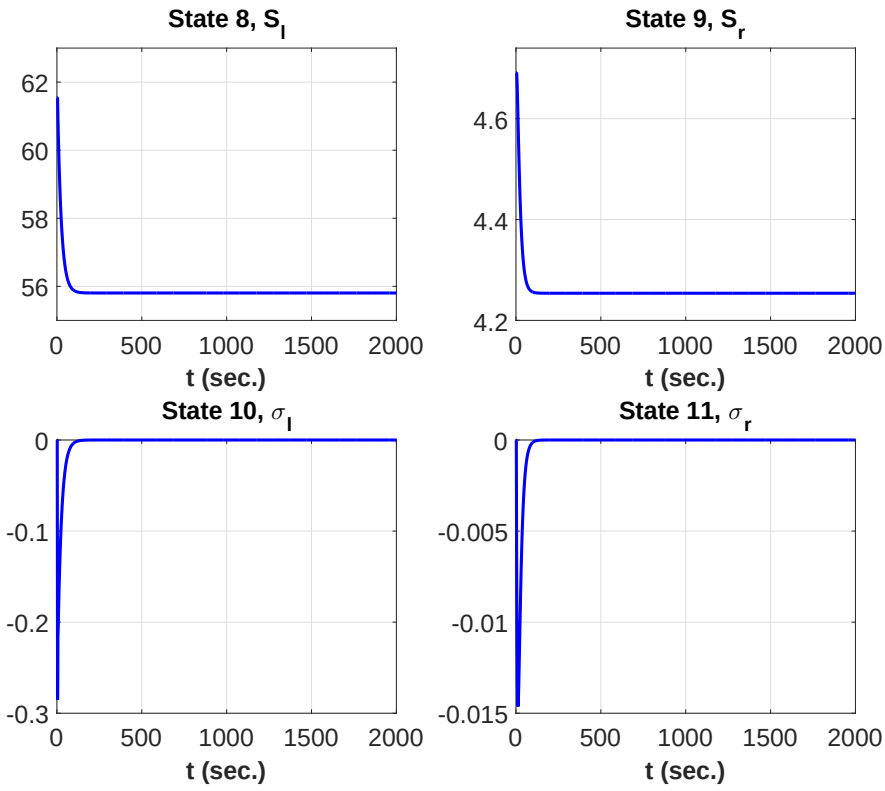


Figure 4.5: Optimal state trajectories of the cardiorespiratory system.

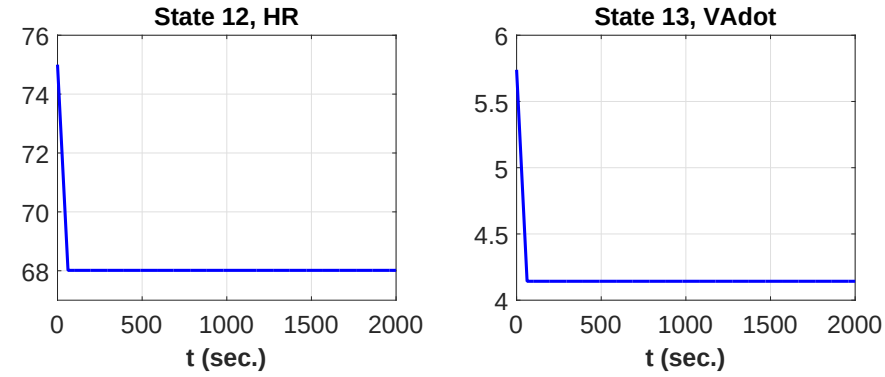


Figure 4.6: Optimal state trajectories of the cardiorespiratory system.

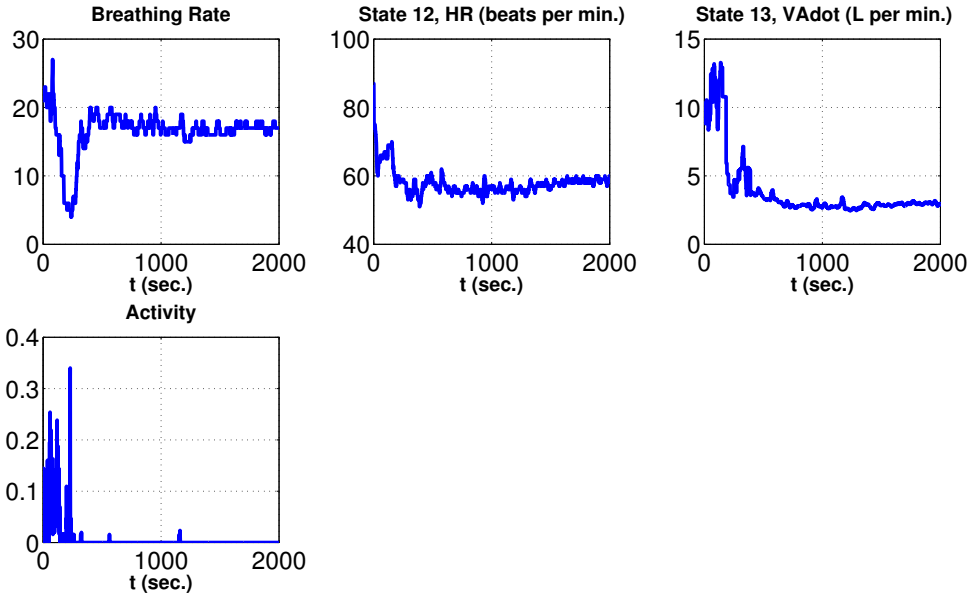


Figure 4.7: Real data from a healthy 25-year-old male subject during awake to sleep transition.

4.5 Summary

In this chapter, we focused on modeling the control mechanism of the cardiorespiratory system during the transition from an awake state to stage 4 non-REM sleep state. A cardiorespiratory system model with transport delays is considered. An iterative algorithm is proposed to find the optimal control inputs that drive the cardiorespiratory system from awake state to sleep state. Simulation results show the effectiveness of the proposed control mechanism model. The cardiorespiratory system states converge to their sleep steady state values. Comparisons with real physiological signals show that the control mechanism model captures the system dynamics of a subject during awake to sleep state transition.

CHAPTER 5

CARDIORESPIRATORY MODEL-BASED FUSION FRAMEWORK FOR SLEEP APNEA DETECTION

Sleep apnea is a serious chronic sleep disorder in which a person experiences frequent pauses in breathing during sleep. A sleep apnea episode is defined as the lack of airflow at the nose and mouth for at least 10 seconds [84]. There are two types of sleep apnea: central sleep apnea (CSA) and obstructive sleep apnea (OSA). In CSA, a person experiences breathing pauses due to the lack of respiratory effort neural signals from the brain. This usually occurs due to some feedback instabilities in the respiratory control system. In OSA, the breathing pauses are due to the upper airway obstruction. OSA is the most common type of sleep apnea affecting millions of people worldwide. According to multiple international surveys and studies, around 2-10% of adults have OSA [85].

Sleep apnea has some serious adverse physiological effects. Repetitive apneic episodes during sleep reduce the blood oxygen saturation level. Surges in heart rate and blood pressure after each apneic episode cause arousal from sleep, thereby affecting the overall sleep quality. Individuals with sleep apnea are rarely aware of the condition, and hence it is often left untreated. According to one study, around 93% of women and 82% of men with sleep apnea are not diagnosed [86]. When left untreated, sleep apnea increases the risk of hypertension, cardiac arrhythmia, heart attack, and stroke [87, 88]. Sleep apnea also increases the risk of motor vehicle collisions [89]. There is a great chance of the at-risk population to be left undiagnosed and untreated for a long time.

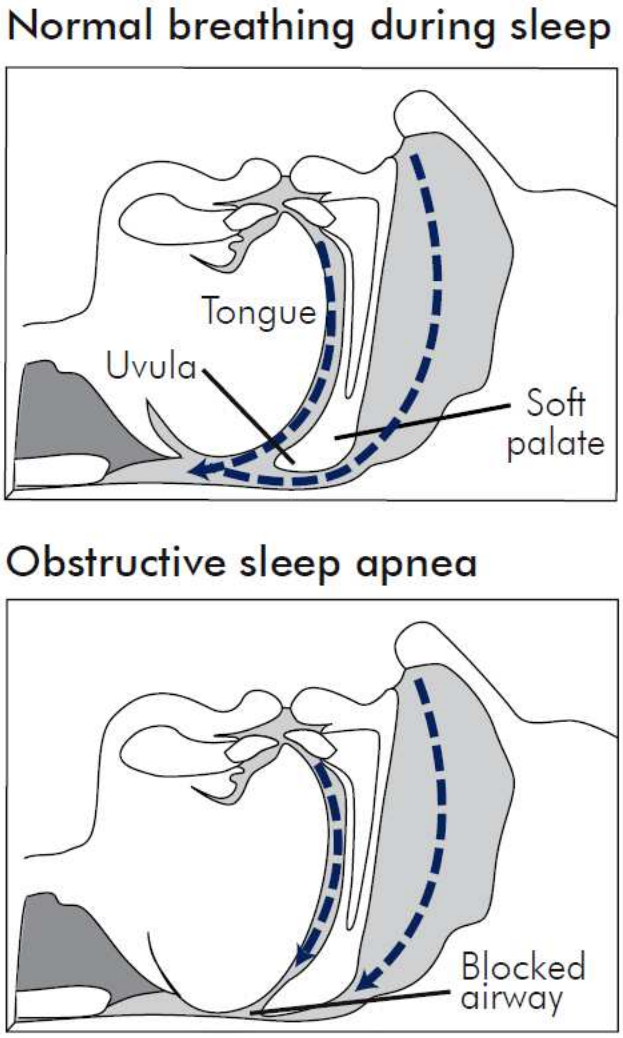


Figure 5.1: Anatomy of obstructive sleep apnea (OSA) [12].

Sleep apnea is usually diagnosed through a polysomnography (PSG) sleep study. It involves overnight monitoring of the patient's physiological signals like electrocardiogram (ECG), electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), peripheral oxygen saturation (SpO_2), airflow, thoracic and abdominal movement signals, etc. in a sleep lab at the hospital. The PSG signals are then analyzed by a sleep specialist to diagnose sleep apnea. Due to first-night effect [90,91] and night-to-night variability, multiple night studies may often be required. Polysomnography is expensive, labor-intensive, time consuming, and inconvenient to the patient. Polysomnography sleep study typically costs few thousand dollars [1, 85]. Moreover,

limited availability of sleep labs and sleep medicine specialists often leads to longer wait times for diagnosis and treatment of sleep apnea. The demand for sleep labs and waiting time for diagnosis and treatment of sleep apnea in different countries is shown in Table 5.1. Hence, there is a great need to develop a low-cost, noninvasive, safe and accurate method of detecting sleep apnea that is easily accessible to everyone [92].

Table 5.1: Sleep study rates and waiting time for diagnosis and treatment of sleep apnea in different countries [1]

Country	Population (million)	Number of Sleep Labs	Number of Sleep Studies/year/100,000	Waiting Time (months)
United States	280	1,292	427	2 to 10
United Kingdom	58.8	84	42.5	7 to 60
Canada	31.4	100	370.4	4 to 36
Australia	18.97	65	282	3 to 16
Belgium	10	50	177.2	2

With the availability of several low-cost wearable sensors that can conveniently collect different physiological signals such as ECG, SpO_2, respiration (airflow), etc. in a noninvasive manner, it is now possible to develop low-cost and easy-to-implement methods for detection of OSA. In this chapter, we propose a new framework for OSA detection combining cardiorespiratory model-based and data-driven approaches. We combine the available sensor measurement data with the cardiorespiratory model information to further improve the OSA detection performance. The proposed framework can combine measurement signals of different sensor modalities with the cardiorespiratory system model information for better detection performance. Mathematical models of the cardiovascular, respiratory, and sleep regulation systems can capture the physiological system dynamics, mechanisms, and interactions to certain extent and this model information can be used to increase the sleep apnea detection

accuracy. Since the models represent the basic physiological dynamics in general and do not reflect the actual dynamics in any specific individual, in order to model the physiological variations among different individuals, vector-valued Gaussian processes (GPs) are adopted and imposed upon existing mathematical models of multiple physiological signals. The proposed OSA detection framework can integrate different measurement signals such as heart rate, peripheral oxygen saturation (SpO_2), airflow, etc. which can be easily acquired using the available wearable sensors noninvasively, and hence can be used to detect sleep apnea at home. The main advantage of such home sleep apnea test (HSAT) is less cost compared to the traditional in-lab PSG test. Other potential advantages of HSAT include no first-night effect and collection of more representative and accurate physiological data during sleep.

5.1 Related Work

Several sleep apnea detection methods are proposed in the literature recently, and the sensors used, features extracted and classifiers adopted are summarized in Table 5.2. In [93], Mendez *et al.* present a method for detecting OSA based on a single-lead ECG signal. A bivariate time-varying autoregressive model is used to define features from the power spectral densities of RR interval and QRS complex area signals. Artificial neural networks and k-nearest neighbor algorithm are then used to classify the apneic and non-apneic episodes. Khandoker *et al.* apply support vector machine (SVM) classifier for detecting OSA using the wavelet-based features of the ECG signal [94]. The features are extracted from the wavelet decomposition of heart rate variability (HRV) and ECG-derived respiration (EDR) signals. In [95], Bsoul *et al.* develop a real-time sleep apnea monitoring system using the support vector machine (SVM) classifier and implement it on an Android operation system based smartphone. A total of 111 temporal and spectral features are extracted from a single-lead ECG signal. Xie *et al.* propose a sleep apnea detection method based on the features extracted from both

ECG and SpO_2 signals [96]. For classification of apneic episodes, several classifiers (such as support vector machine, k-nearest neighbor, decision tree, multilayer perceptron) are combined using ensemble methods such as Adaptive Boosting, Bagging, etc. In [97], Nguyen *et al.* propose an online sleep apnea detection method based on recurrence quantification analysis (RQA) statistics of the heart rate variability (HRV) signal extracted from the ECG signal. For classification of apneic episodes, a soft decision fusion rule is proposed to combine the decisions of support vector machine and artificial neural network classifiers. Koley and Dey propose an online sleep apnea detection method using the SpO_2 signal measured at the fingertip through a pulse oximeter [98]. A total of 34 time domain features are extracted from the SpO_2 signal. Support vector machine classifier is used for classification. Varon *et al.* propose an algorithm for automatic detection of sleep apnea from the ECG signal [99]. From the ECG signal, the authors extract RR interval signal and three ECG derived respiration (EDR) signals. A total of 28 features are defined from these ECG-derived signals. For classification, linear discriminant analysis (LDA) and support vector machine (SVM) classifiers are used.

All these existing sleep apnea detection methods are purely data-driven approaches. First, certain features are defined and extracted from the commonly available physiological signals such as ECG, SpO_2, etc. Then, supervised machine learning methods are used to classify the apneic and non-apneic data. The main drawback of these data-driven approaches is that their performance greatly depends on the quality and quantity of the sensor data available for training. Also, the data-driven methods require a lot of feature engineering to define features for each physiological signal, and their detection performance also depends on the type of features extracted from the physiological signals. Moreover, these methods treat different physiological signals as independent data sources and perform feature extraction for each physiological signal separately. However, in reality the physiological signals are dependent. For

Table 5.2: Related work on sleep apnea detection

	Sensors	Features	Classifiers
Mendez *et al.* '09 [93]	One-lead ECG	Power spectral densities of RR interval and QRS complex area signals	Neural network, k-nearest neighbor
Khandoker *et al.* '09 [94]	One-lead ECG	Wavelet decomposition of HRV and EDR signals	Support vector machine
Bsoul *et al.* '11 [95]	One-lead ECG	111 temporal and spectral features extracted from ECG signal	Support vector machine
Xie *et al.* '12 [96]	One-lead ECG, SpO_2	Temporal and spectral features	Ensemble classifier
Nguyen *et al.* '14 [97]	One-lead ECG	Recurrence quantification analysis (RQA) statistics of the HRV signal	Soft decision fusion combining support vector machine and neural network
Koley and Dey '14 [98]	SpO_2	34 time domain features extracted from the SpO_2 signal	Support vector machine
Varon *et al.* '15 [99]	One-lead ECG	28 features extracted from RR interval and EDR signals	Linear discriminant analysis, support vector machine

instance, ECG signal contains respiratory effort information and the photoplethysmography (PPG) signal of pulse oximeter contains heart rate and respiration rate information [100].

5.2 Measurement Models

5.2.1 Heart rate

The heart rate measurement signal can be modeled as a measurement of the actual heart rate of the heart. We model the sensor measurement noise as additive white Gaussian noise (AWGN). The heart rate measurement model is given by,

$$z_{t+1}^{HR} = x_{t+1}^{HR} + n, \quad t \geq 0, \tag{5.1}$$

where x_t^{HR} is the true heart rate, z_t^{HR} is the measured heart rate, and $n \sim \mathcal{N}(0, \sigma_n^2)$ is the independent and identically distributed (i.i.d.) measurement noise following a Gaussian distribution with zero mean and variance σ_n^2.

5.2.2 Peripheral oxygen saturation (SpO_2)

Oxygen is vital to the human metabolism. It reacts with carbohydrates, fats, and proteins to release the energy necessary for cellular function in the human body [101]. Hemoglobin (Hb) in the red blood cells transports oxygen from the lungs to the tissues in the human body. Oxygen level in the human body can be defined using the arterial oxygen saturation (SaO_2), which is the percentage of hemoglobin saturated with oxygen molecules in the arterial blood. SaO_2 plays a vital role in the detection of sleep apnea. Due to the lack of airflow during each sleep apnea episode, the oxygen saturation levels in the human body vary greatly during apnea episodes. However, measurement of SaO_2 is invasive and requires sampling blood from arteries [102]. A noninvasive estimate of SaO_2 can be obtained by measuring the peripheral oxygen

saturation (SpO_2) near extremities (such as fingers, ears, forehead, etc.) using a pulse oximeter [103].

Even though SpO_2 is related to SaO_2, modeling their relationship is hard as it is difficult to obtain SaO_2 measurements in practice. The SpO_2 signal is also related to the respiration signal. For instance, during sleep apnea episodes, the lack of airflow for a certain amount of time decreases the amount of available oxygen in the lungs, and the SpO_2 value drops below a certain level. In this section, we propose a mathematical model relating the SpO_2 signal to the respiration signal such as tidal volume signal, airflow signal, etc. These respiration signals can be easily acquired using noninvasive techniques such as respiratory inductance plethysmography (RIP) and impedance pneumography (IP). The proposed model tries to capture the physiological interactions between respiration and peripheral oxygen saturation (SpO_2), and it takes into account the time delay between these two signals.

Respiration-SpO_2 model

Let $x_r(t)$ and $x_s(t)$ be the respiration and SpO_2 signals, respectively. We first construct a signal $\mu_r(t)$, where $\mu_r(t)$ is the mean of $x_r(t)$ over the time window $t \in [t - t_w, t]$. We choose the window size $t_w = 4$ sec., which is the average duration of a single breath (one inhalation and exhalation). Generally, changes in the respiration signal are reflected in the SpO_2 signal with a certain time delay. For example, the respiration, respiration mean, and SpO_2 signals of a subject with sleep apnea are shown in Figure 5.2. From the figure, we can clearly see that the SpO_2 value changes according to the respiration mean signal with a time delay.

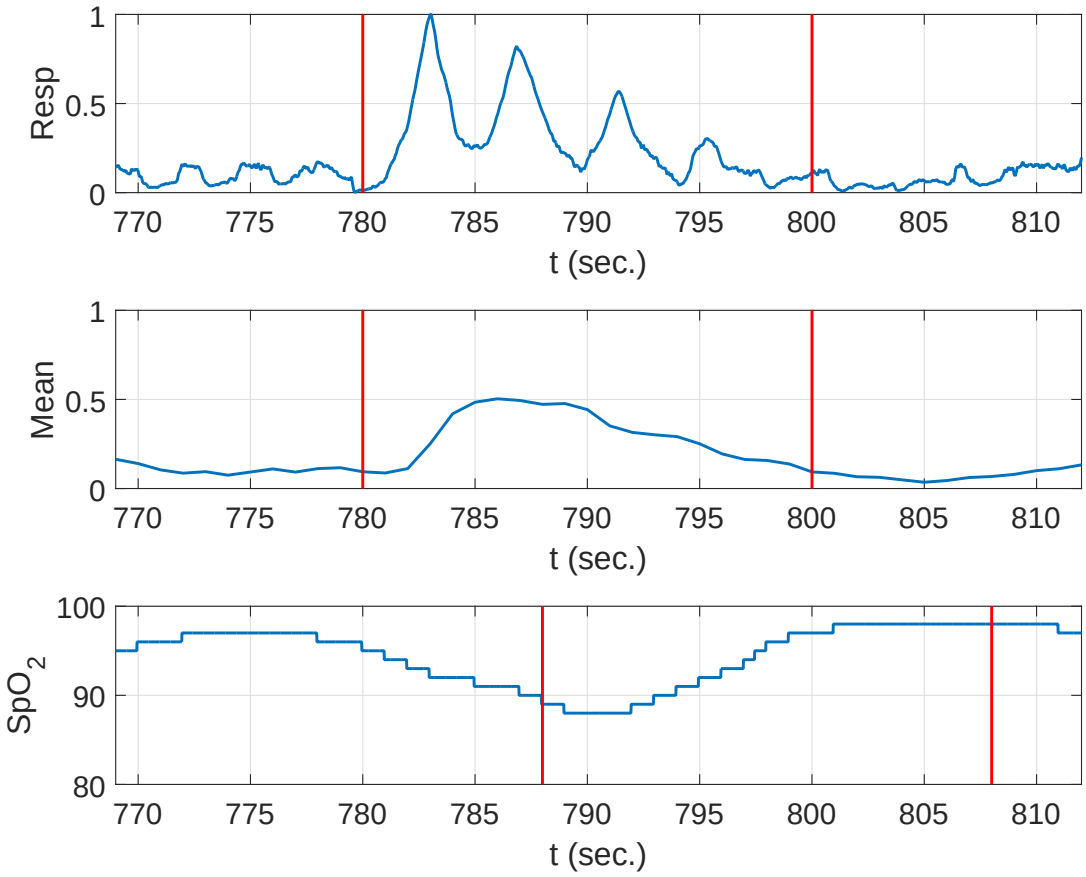


Figure 5.2: Sample respiration $x_r(t)$, respiration mean $\mu_r(t)$, and SpO_2 $x_s(t)$ signals from PhysioNet [13]. A sleep apnea episode is indicated by the red vertical lines.

Let $x_r[k]$ and $x_s[k]$ be the discrete-time respiration and SpO_2 signals, respectively. We model the relation between the SpO_2 and respiration signals as follows:

- We first generate a signal $y_s[\cdot]$ as,

$$\begin{aligned} y_s[k] &= f(x_r[k-d]) \\ &= y_s[k-1] + (\mu_r[k-d] - c)\, t_s, \end{aligned} \tag{5.2}$$

  where $k \geq 1$, $\mu_r[k]$ is the discrete-time respiration mean signal, c is a threshold, t_s is the sampling time interval, $d > 0$ is the time delay, and the initial value $y_s[0] = 0$.

- We then scale and shift the above signal $y_s[\cdot]$ to get the SpO_2 signal,

$$x_s[k] = a\, y_s[k] + b, \qquad k \geq 0, \tag{5.3}$$

  where a is the scaling parameter and b is the shift parameter (bias).

The model parameters, $\boldsymbol{\theta} = \{a, b, c, d\}$, are estimated by solving the following non-linear least square problem,

$$\min_{\boldsymbol{\theta}} \;\; E(\boldsymbol{\theta}) = \|\mathbf{x}_s - \mathbf{x}\|_2^2, \tag{5.4}$$

where $\mathbf{x} = [x[1], x[2], ..., x[N]]^T$ is the available SpO_2 measurement signal. For ease of optimization, we fix the time delay d and optimize over the other parameters.

Using the respiration signal $\mathbf{x}_r = [x_r[1], x_r[2], ..., x_r[N]]^T$, the model SpO_2 signal with time delay $d > 0$ is given by,

$$x_s[k] = a \left(\sum_{j=2}^{k-d} \mu_r[j] - (k - d - 1)\, c \right) t_s + b, \tag{5.5}$$

where $k = d+1, d+2, ..., N$, $\mu_r[j]$ is the mean of the respiration signal $x_r[\cdot]$ over the last breath window at time instant j. Specifically, $x_s[d+1] = b$. We estimate the model parameters a, b, c, and d by minimizing the residual sum of squares,

$$\min_{a,b,c,d} \;\; E(a, b, c, d) = \|\mathbf{x}_s - \mathbf{x}\|_2^2, \tag{5.6}$$

where $\mathbf{x} = [x[d+1], x[d+2], ..., x[N]]^T$ is the discrete-time measured SpO_2 signal, $\mathbf{x}_s = [x_s[d+1], x_s[d+2], ..., x_s[N]]^T$ is the discrete-time model SpO_2 signal with time delay $d > 0$.

For ease of optimization, we fix the time delay d, and estimate the corresponding optimal parameters $a^\star$, $b^\star$, and $c^\star$ by minimizing the sum of squares objective function,

$$\begin{aligned} &\min_{a,b,c} \;\; E(a, b, c) \\ &\qquad = \sum_{k=d+1}^{N} \left(x_s[k] - x[k] \right)^2 \\ &\qquad = \sum_{k=d+1}^{N} \left(a \, \left(\mu[k-d] - (k - d - 1)\, c \right) \, t_s + b - x[k] \right)^2, \end{aligned} \tag{5.7}$$

where $\mu[k] = \sum_{j=2}^{k} \mu_r[j]$. Note that the above optimization problem has implicit constraints on the scale and shift parameters: $a \geq 0$ and $b > 0$.

By equating the derivatives of the sum of squares function $E(\cdot)$ with respect to a, b, and c to zero, we get the following system of nonlinear equations,

$$\begin{aligned}
&\frac{\partial E}{\partial a} = 0 \\
&\implies \qquad (5.8a) \\
&a\,t_s(Q_{\mu 2} + c\,Q_{\mu k}) + c(a\,t_s(Q_{\mu k} + c\,Q_{k2}) - Q_{xk}) = Q_{x\mu}, \\
&\frac{\partial E}{\partial b} = 0 \implies (N-d)\,b + a\,t_s\,P_\mu + c\,a\,t_s\,P_k = P_x, \qquad (5.8b) \\
&\frac{\partial E}{\partial c} = 0 \implies a\,(a\,t_s\,(Q_{\mu k} + c\,Q_{k2}) - Q_{xk}) = 0, \qquad (5.8c)
\end{aligned}$$

where the partial sums are

$$\begin{aligned}
P_x &= \sum_{k=d+1}^{N} x[k], \\
P_\mu &= \sum_{k=d+2}^{N} \mu[k-d], \quad P_k = \sum_{k=d+2}^{N} (1-k+d), \\
P_{\mu 2} &= \sum_{k=d+2}^{N} \mu[k-d]^2, \quad P_{k2} = \sum_{k=d+2}^{N} (1-k+d)^2, \\
P_{x\mu} &= \sum_{k=d+2}^{N} x[k]\,\mu[k-d], \quad P_{xk} = \sum_{k=d+2}^{N} x[k]\,(1-k+d), \qquad (5.9) \\
P_{\mu k} &= \sum_{k=d+2}^{N} \mu[k-d]\,(1-k+d), \\
Q_{\mu 2} &= (N-d)\,P_{\mu 2} - P_\mu^2, \;\; Q_{k2} = (N-d)\,P_{k2} - P_k^2, \\
Q_{x\mu} &= (N-d)\,P_{x\mu} - P_x\,P_\mu, \;\; Q_{xk} = (N-d)\,P_{xk} - P_x\,P_k, \\
Q_{\mu k} &= (N-d)\,P_{\mu k} - P_\mu\,P_k.
\end{aligned}$$

From (5.8c), we have

$$\begin{aligned}
a &= \frac{Q_{xk}}{t_s\,(Q_{\mu k} + c\,Q_{k2})}, \quad \text{or} \\
a &= 0.
\end{aligned} \qquad (5.10)$$

Accordingly, we get the following two sets of solutions:

- When $a \neq 0$, from (5.8a), we get the optimal threshold,

$$c^\star = \frac{Q_{xk}\, Q_{\mu 2} - Q_{x\mu}\, Q_{\mu k}}{Q_{x\mu}\, Q_{k2} - Q_{xk}\, Q_{\mu k}}. \tag{5.11}$$

  By substituting the optimal threshold $c^\star$ in (5.10) and (5.8b), we get

$$\begin{aligned} a^\star &= \frac{Q_{xk}}{t_s\,(Q_{\mu k} + c^\star\, Q_{k2})}, \\ b^\star &= \frac{P_x - a^\star\, t_s\, P_\mu - c^\star\, a^\star\, t_s\, P_k}{N - d}. \end{aligned} \tag{5.12}$$

- When $a = 0$, from (5.8a) and (5.8b), we get

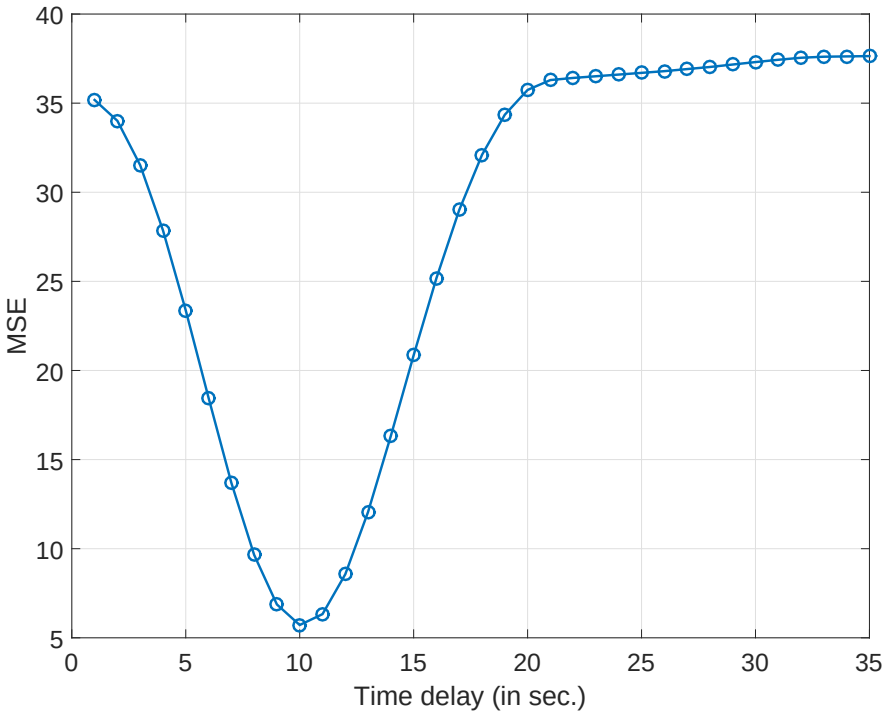
$$a^\star = 0, \quad b^\star = \frac{P_x}{N-d}, \quad c^\star = -\frac{Q_{x\mu}}{Q_{xk}}. \tag{5.13}$$

We choose the feasible set $\{a^\star, b^\star, c^\star\}$ with the least objective function value. To find the optimal time delay ($d^\star$), we try different delay values from a finite set of time delays (typically ranging from 1 sec. to 60 sec.) and choose the time delay with the minimum mean squared error (MSE).

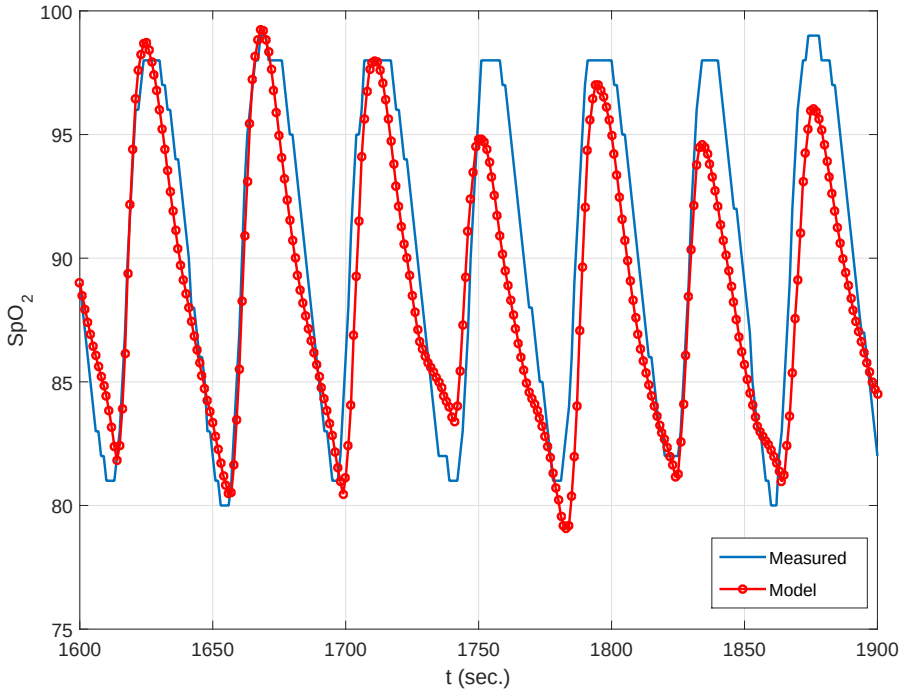
To test the proposed model, we use real sleep apnea data from Apnea-ECG database in PhysioNet [13]. The database contains continuous digitized ECG signals of 32 subjects with OSA. For eight subject records, four additional physiological signals are available: SpO_2, oronasal airflow signal, chest and abdominal respiratory inductance plethysmography signals. We consider the abdominal respiratory inductance plethysmography signal as the respiration signal in the proposed model.

In our experiments, we consider four subjects (a01-a04) and use the first two hours of sleep apnea data for testing the proposed model. For each subject, we divide the entire data into five-minute windows. We then estimate the model parameters model parameters a, b, c, and d using the respiration and SpO_2 signal data from one window as training data, and then test the model using the data from the next window. For each training window, we try different delay values from 1 sec. to 35 sec. and choose the time delay with the minimum mean squared error (MSE). The MSE for different

time delays for a training window and the model vs. measurement SpO_2 signals of a test window of subjects 1 and 4 are shown in Figure 5.3 and Figure 5.4, respectively. From the results, we can see that the model SpO_2 signal accurately tracks the actual SpO_2 measurement signal. Also, the estimated model parameters over a series of subject 1 windows is shown in Figure 5.5. From the figure, we can see that the estimated model parameters of consecutive windows are very close to each other.

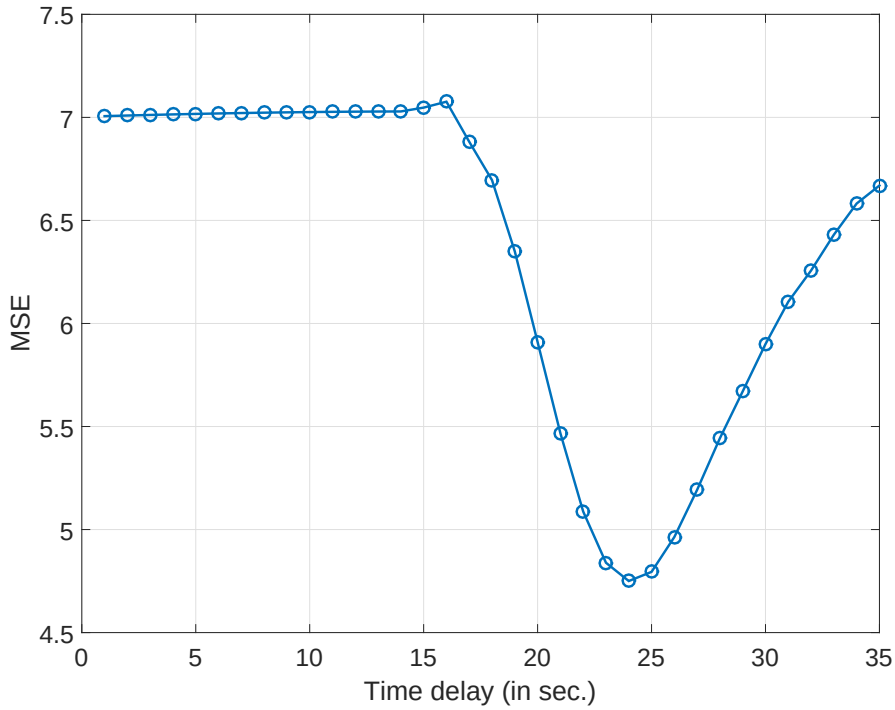


(a) Training MSE vs. time delay (d)

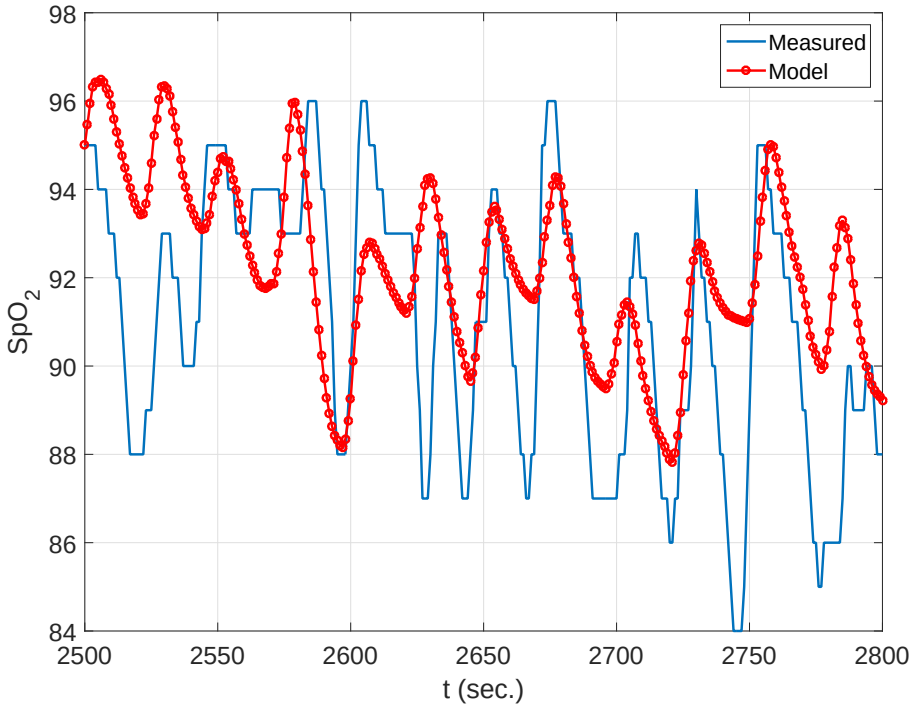


(b) Model SpO_2 signal $x_s(t)$ and measured SpO_2 signal $x(t)$

Figure 5.3: Subject 1 results.



(a) Training MSE vs. time delay (d)



(b) Model SpO_2 signal $x_s(t)$ and measured SpO_2 signal $x(t)$

Figure 5.4: Subject 4 results.

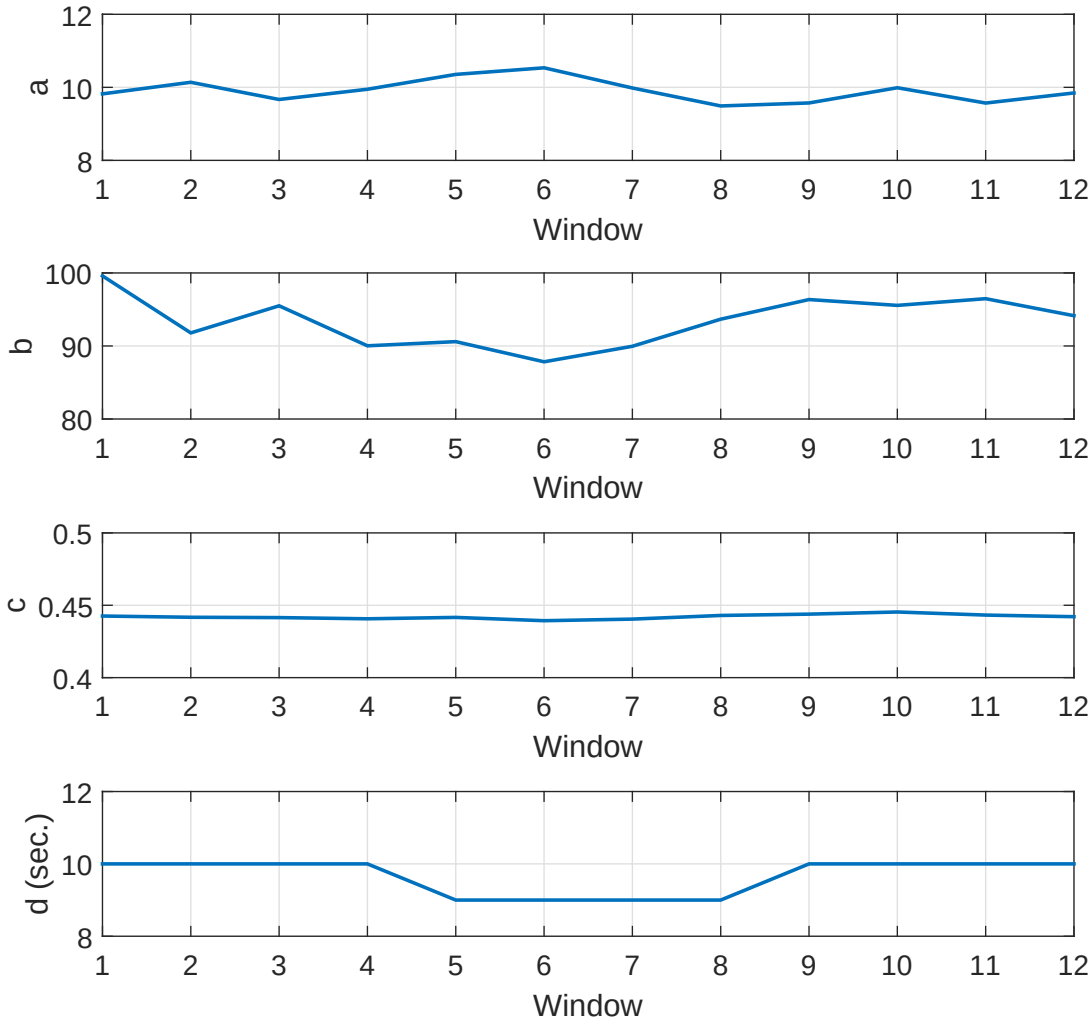


Figure 5.5: Model parameter estimates of Subject 1 windows.

5.3 State Models

We adopt the integrative model of the cardiovascular and respiratory systems in [11] to define the cardiorespiratory system state-space model. By setting the upper airway sensitivity parameter value to at least 0.38 (from its normal value of 0.01), the model generates each OSA episode as follows. Increased upper airway sensitivity leads to a decrease in the airflow. This leads to hypoxia (decreased O_2 concentration) and

hypercapnia (elevated CO_2 concentration). The respiratory effort will then increase, which increases the pleural pressure in the lungs. This triggers an arousal from sleep, which leads to an increase in the sympathetic nervous activity and decrease in the parasympathetic nervous activity. This finally leads to an increase in heart rate and blood pressure. Also, the hypoxia and hypercapnia states additionally lead to an increase in the sympathetic nervous activity through the chemoreflex mechanism. Some key physiological signals generated by the model under normal and OSA conditions are shown in Figure 5.6 and Figure 5.7, respectively. From the figure, we can see that the model generated signals represent the OSA dynamics and trends in general.

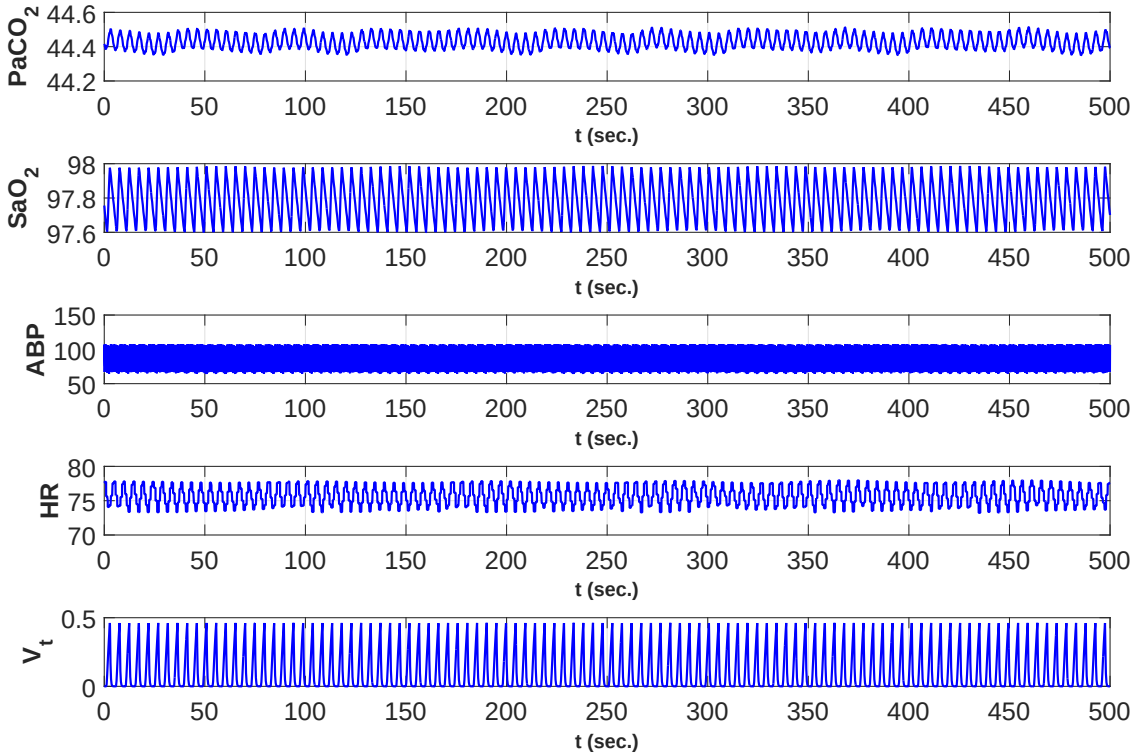


Figure 5.6: Normal sleep simulation signals: partial pressure of carbon dioxide in arterial blood ($PaCO_2$), oxygen saturation (SaO_2), arterial blood pressure (ABP), heart rate (HR), tidal volume (V_t).

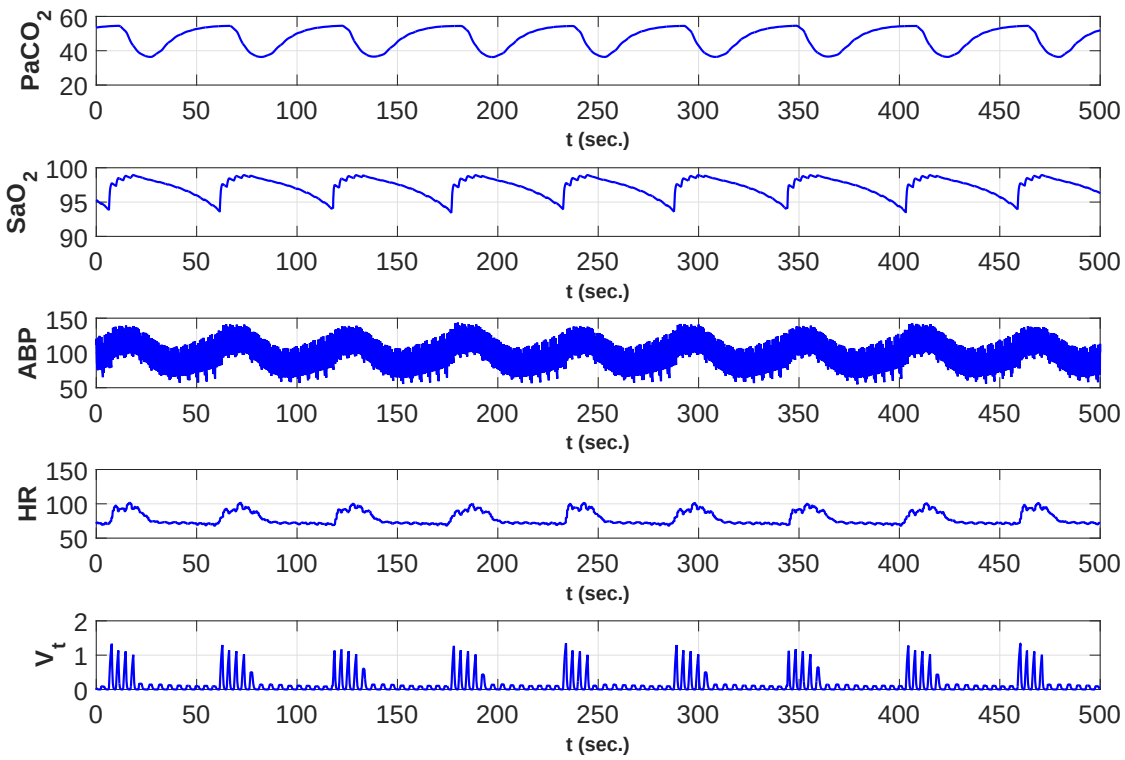


Figure 5.7: OSA simulation signals: partial pressure of carbon dioxide in arterial blood ($PaCO_2$), oxygen saturation (SaO_2), arterial blood pressure (ABP), heart rate (HR), tidal volume (V_t).

5.3.1 Introducing virtual oxygen saturation state

In the cardiorespiratory system model, we consider the following states: partial pressure of carbon dioxide in arterial blood ($PaCO_2$), arterial oxygen saturation (SaO_2), arterial blood pressure (ABP), heart rate (HR), and tidal volume (V_t), since they are most relevant to OSA symptoms. Note that not all these states can be noninvasively measured in practice. Even though SpO_2 is related to SaO_2, modeling their relationship is hard as it is difficult to obtain SaO_2 measurements in practice. From the previous section, we observe that the SpO_2 can also be related to the respiration signal such as airflow, tidal volume, etc. So in addition to the cardiorespiratory sys-

tem states, we define a new virtual oxygen saturation state (VSO_2) using the tidal volume state. Note that there is usually a time delay between the respiration and SpO_2 signals. We do not consider the time delay in the state model as the time delay is not constant and often varies from subject to subject. The time delay will first be estimated and measurement synchronization will be conducted accordingly before detection. The new VSO_2 state is generated as follows:

- Let $x_{vt}[k]$ and $x_{vs}[k]$ be the discrete-time tidal volume and VSO_2 states, respectively. First, construct a signal $\mu_{vt}[k]$, where $\mu_{vt}[k]$ is the mean of $x_{vt}[k]$ over the time window $[k - k_w, k]$. We choose the window size $k_w = 4$ sec., which is the average duration of a single breath (one inhalation and exhalation).

- Generate a signal $y_{vs}[\cdot]$ as,

$$\begin{aligned} y_{vs}[k] &= f(x_{vt}[k]) \\ &= y_{vs}[k-1] + (\mu_{vt}[k] - c)\, t_s, \end{aligned} \tag{5.14}$$

  where $k \geq 1$, c is a threshold, t_s is the sampling time interval, and the initial value $y_{vs}[0] = 0$.

- We then scale and shift the above signal $y_{vs}[\cdot]$ to produce the VSO_2 state,

$$x_{vs}[k] = a\, y_{vs}[k] + b, \qquad k \geq 0, \tag{5.15}$$

  where a is the scaling parameter and b is the shift parameter (bias).

Based on our empirical study, we use the following parameter values to generate reasonable oxygen saturation state values under both normal and apnea conditions:

- Normal: $a = 2$, $b = 94$, $c = 0.16$.

- OSA: $a = 2$, $b = 86$, $c = 0.16$.

The generated VSO_2 state sequences from tidal volume state under both normal and OSA conditions are shown in Figure 5.8. Note that the actual SpO_2 state will be a delayed version of this new VSO_2 state, $SpO_2(t) = VSO_2(t - d)$, where $d > 0$ is the time delay.

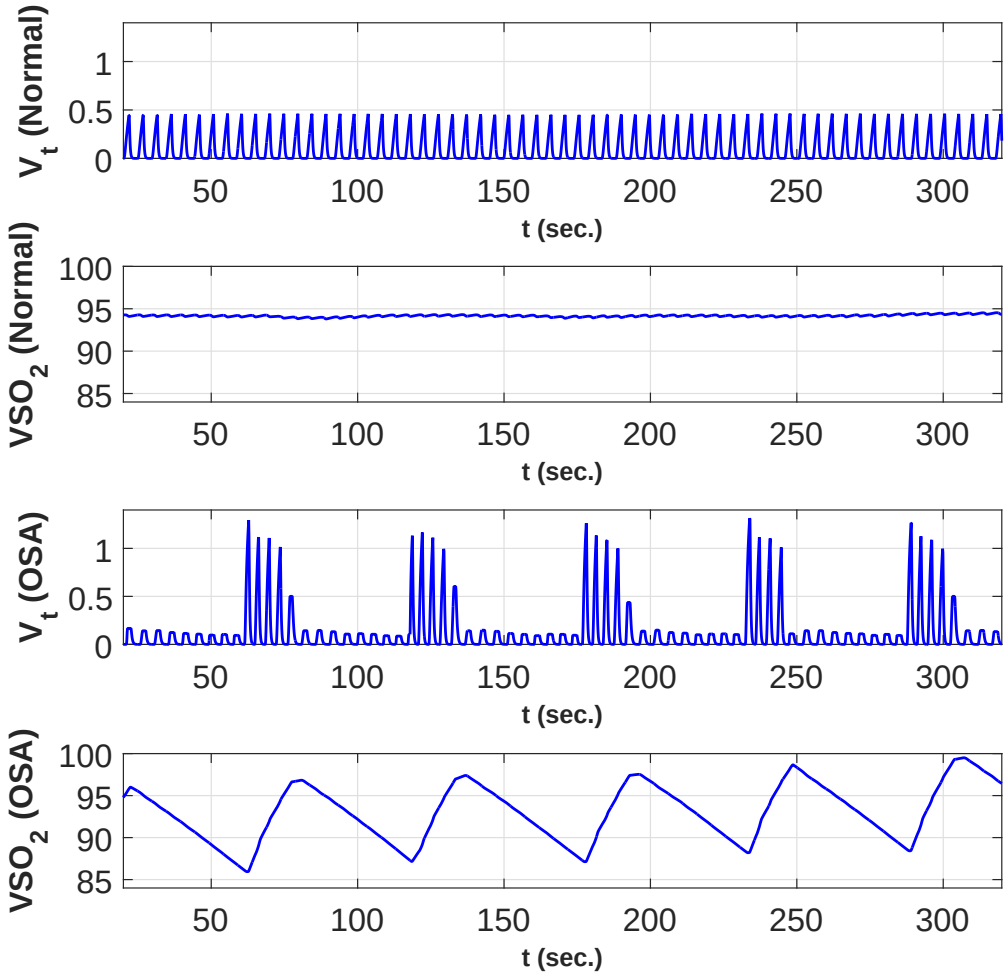


Figure 5.8: Tidal volume (V_t) and generated virtual oxygen saturation (VSO_2) state signals under normal and OSA conditions.

5.3.2 Gaussian process state-space model

We represent the cardiorespiratory system model as a discrete-time nonlinear state-space model,

$$\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t) + \boldsymbol{\epsilon}_t, \quad t \geq 0, \tag{5.16}$$

where $\mathbf{x}_t \in \mathbb{R}^D$ are the states in the model at time t, D is the total number of states in the model, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \Sigma)$ is the independent and identically distributed (i.i.d.) process noise following Gaussian distribution with zero mean and diagonal covariance matrix $\Sigma \in \mathbb{R}^{D\times D}$ with noise variances $\{\sigma_1^2, \sigma_2^2, ..., \sigma_D^2\}$ on its main diagonal, and $\mathbf{f}(\cdot)$ represents the cardiorespiratory system state functions.

For the set of system state functions, $\{f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_D(\mathbf{x})\}$, where $\mathbf{x} \in \mathbb{R}^D$, a Gaussian process (GP) prior is applied,

$$\mathbf{f}(\cdot) \sim GP\left(\boldsymbol{\mu}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x}')\right), \tag{5.17}$$

where $\boldsymbol{\mu}(\mathbf{x}) \in \mathbb{R}^D$ is the GP mean function and $\mathbf{K}(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^{D\times D}$ is the GP covariance matrix. Note that the GP prior tries to model the variations of system states among different individuals. The GP mean function values may come from the cardiorespiratory system model in [11]. This way the existing system model information is incorporated into our model through the GP mean function. For any $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^D$, $\mathbf{K}(\mathbf{x}, \mathbf{x}')$ is defined using the sum of separable kernels (SoS kernels) structure [104] as,

$$\mathbf{K}(\mathbf{x}, \mathbf{x}') = \sum_{q=1}^{Q} k_q(\mathbf{x}, \mathbf{x}')\, \mathbf{B}_q, \tag{5.18}$$

where $k_q(\cdot, \cdot)$ is a scalar kernel function and $\mathbf{B}_q \in \mathbb{R}^{D\times D}$ are symmetric positive semi-definite matrices. The following widely used Gaussian kernel is adopted as the scalar kernel function,

$$k_q(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\delta_q^2}\right). \tag{5.19}$$

Let $\mathbf{y}_t = \mathbf{x}_{t+1}$. The state-space model is rewritten as

$$\mathbf{y}_t = \mathbf{f}(\mathbf{x}_t) + \boldsymbol{\epsilon}_t. \tag{5.20}$$

Let $\boldsymbol{\theta}$ denote the set of all hyperparameters. Note that the hyperparameters include process noise variances $\{\sigma_1^2, \sigma_2^2, ..., \sigma_D^2\}$, parameters of the Gaussian kernel function $\{\delta_q^2\}$, and the elements of all $\mathbf{B}_q$ matrices. The state-space models under the normal condition and the OSA condition are different in both the upper airway sensitivity parameter and the hyperparameters. For OSA detection, we first need to estimate $\boldsymbol{\theta}$ using the training data $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$ and $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N\}$ obtained by simulating the cardiorespiratory system model in [11] under both conditions. Then OSA detection is performed using both the measurements and the GP state-space model. The main steps are shown in the flowchart in Figure 5.9.

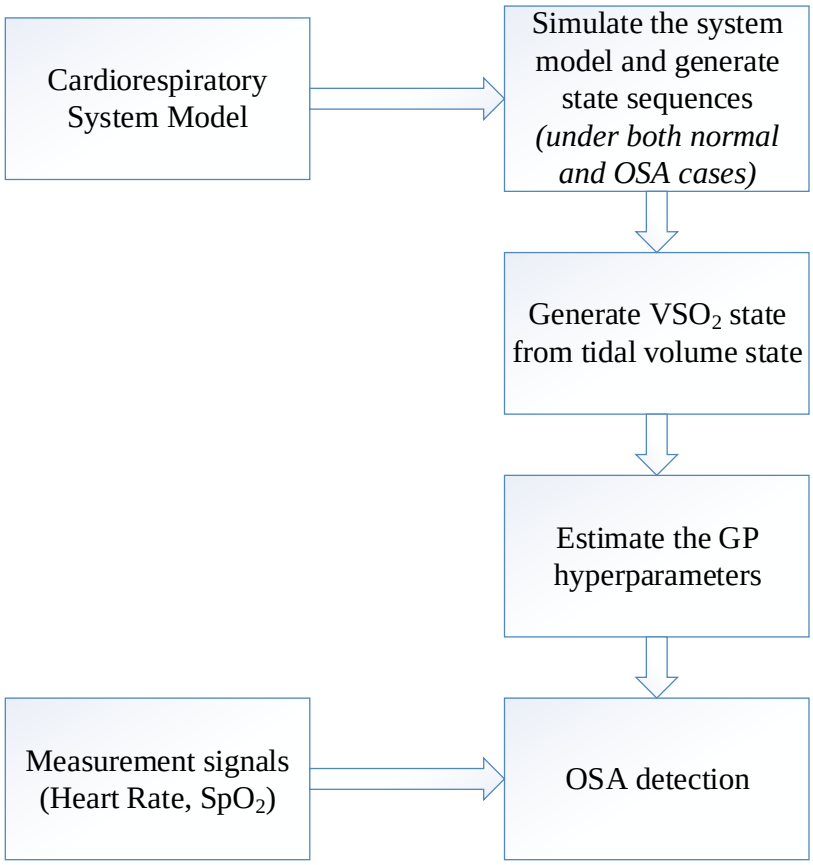


Figure 5.9: Flowchart of model-based and data-driven framework for OSA detection.

5.4 Model Hyperparameters Estimation

Let $\mathbf{f} = \left[\mathbf{f}(\mathbf{x}_1)^T, \mathbf{f}(\mathbf{x}_2)^T, ..., \mathbf{f}(\mathbf{x}_N)^T\right]^T$. The prior distribution over $\mathbf{f}$ is given by $\mathbf{f} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}_{ND})$. Note that the mean vector $\boldsymbol{\mu} \in \mathbb{R}^{ND}$ can be computed using the training data generated by running the cardiorespiratory system model. The $ND \times ND$ covariance matrix $\mathbf{K}_{ND}$ is constructed using the $D \times D$ GP covariance matrix as follows,

$$\mathbf{K}_{ND} = \begin{bmatrix} \mathbf{K}(\mathbf{x}_1, \mathbf{x}_1) & \mathbf{K}(\mathbf{x}_1, \mathbf{x}_2) & \dots & \mathbf{K}(\mathbf{x}_1, \mathbf{x}_N) \\ \mathbf{K}(\mathbf{x}_2, \mathbf{x}_1) & \mathbf{K}(\mathbf{x}_2, \mathbf{x}_2) & \dots & \mathbf{K}(\mathbf{x}_2, \mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}(\mathbf{x}_N, \mathbf{x}_1) & \mathbf{K}(\mathbf{x}_N, \mathbf{x}_2) & \dots & \mathbf{K}(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}. \tag{5.21}$$

We first subtract the mean from $\mathbf{f}$ to construct a set of zero-mean function values, $\mathbf{g} = \mathbf{f} - \boldsymbol{\mu}$. So the prior distribution over $\mathbf{g}$ is given by $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_{ND})$. Let the corresponding state-space model be,

$$\mathbf{v}_t = \mathbf{g}(\mathbf{x}_t) + \boldsymbol{\epsilon}_t. \tag{5.22}$$

Let $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$ and $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_N\}$ be the corresponding training data, where $\mathbf{v}_i = \mathbf{y}_i - \boldsymbol{\mu}(\mathbf{x}_i)$.

The conditional likelihood of $\mathbf{v}$ given $\mathbf{X}$ is given by

$$p(\mathbf{v}|\mathbf{X}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{v}|\mathbf{0}, \mathbf{K}_{ND} + \boldsymbol{\Sigma}), \tag{5.23}$$

where $\mathbf{v} = \left[\mathbf{v}_1^T, \mathbf{v}_2^T, ..., \mathbf{v}_N^T\right]^T$, $\boldsymbol{\Sigma} = \mathbf{I}_N \otimes \Sigma$, $\otimes$ represents tensor product, $\mathbf{I}_N$ is an identity matrix of size N. Taking logarithm on both sides of (5.23), we have

$$\begin{aligned} \log p(\mathbf{v}|\mathbf{X}, \boldsymbol{\theta}) = &-\frac{1}{2}\mathbf{v}^T(\mathbf{K}_{ND} + \boldsymbol{\Sigma})^{-1}\mathbf{v} - \frac{1}{2}\log|\mathbf{K}_{ND} + \boldsymbol{\Sigma}| \\ &- \frac{ND}{2}\log 2\pi. \end{aligned} \tag{5.24}$$

Generally,

$$\hat{\boldsymbol{\theta}}_{MLE} = \arg\max_{\boldsymbol{\theta}} \ \log p(\mathbf{v}|\mathbf{X}, \boldsymbol{\theta}). \tag{5.25}$$

Note that to compute the log-likelihood function in (5.24), we need to invert the $ND \times ND$ matrix, $\mathbf{K}_{ND} + \mathbf{\Sigma}$, the complexity of which is $O(N^3D^3)$.

To reduce the computational complexity, we construct an approximated likelihood function using a small set of pseudo-input points. This is an extension of the pseudo-inputs approach in [105] to our vector-valued Gaussian process model. Let $\bar{\mathbf{X}} = \{\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, ..., \bar{\mathbf{x}}_M\}$ be the pseudo-input points and let $\bar{\mathbf{g}} = \{\bar{\mathbf{g}}_1, \bar{\mathbf{g}}_2, ..., \bar{\mathbf{g}}_M\}$ be their corresponding function values, where M can be much smaller than N, and each $\bar{\mathbf{x}}_i, \bar{\mathbf{g}}_i \in \mathbb{R}^D$. Let the prior distribution over the pseudo-input function values be $\bar{\mathbf{g}} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_{MD})$, where $\mathbf{K}_{MD} \in \mathbb{R}^{MD \times MD}$ is the covariance matrix of $\bar{\mathbf{g}}$ with entries $K_{d,d'}(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_j)$, where $d, d' = 1, 2, ..., D$ and $i, j = 1, 2, ..., M$.

The single data point likelihood conditioned on the pseudo-input points is given by

$$\begin{aligned} & p(\mathbf{v}_i|\mathbf{x}_i, \bar{\mathbf{X}}, \bar{\mathbf{g}}, \boldsymbol{\theta}) \\ & = \mathcal{N}\left(\mathbf{v}_i|\mathbf{K}_{\mathbf{x}_i}^T \mathbf{K}_{MD}^{-1} \bar{\mathbf{g}}, \mathbf{K}(\mathbf{x}_i, \mathbf{x}_i) - \mathbf{K}_{\mathbf{x}_i}^T \mathbf{K}_{MD}^{-1} \mathbf{K}_{\mathbf{x}_i} + \Sigma\right), \end{aligned} \tag{5.26}$$

where

$$\mathbf{K}_{\mathbf{x}_i} = \begin{bmatrix} \mathbf{K}(\bar{\mathbf{x}}_1, \mathbf{x}_i) \\ \mathbf{K}(\bar{\mathbf{x}}_2, \mathbf{x}_i) \\ \vdots \\ \mathbf{K}(\bar{\mathbf{x}}_M, \mathbf{x}_i) \end{bmatrix}_{MD \times D}.$$

The above likelihood is the GP predictive distribution with $\{\bar{\mathbf{X}}, \bar{\mathbf{g}}\}$ as the training data, $\mathbf{x}_i$ as the test input, and $\mathbf{v}_i$ as the test output [106]. Given the pseudo-input data $\bar{\mathbf{X}}$ and $\bar{\mathbf{g}}$, we assume the set of data points $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_N\}$ are independent.

Hence, we can approximate the likelihood function $p(\mathbf{v}|\mathbf{X}, \bar{\mathbf{X}}, \bar{\mathbf{g}}, \boldsymbol{\theta})$ as follows,

$$\begin{aligned} & p(\mathbf{v}|\mathbf{X}, \bar{\mathbf{X}}, \bar{\mathbf{g}}, \boldsymbol{\theta}) \\ &= \prod_{i=1}^{N} p(\mathbf{v}_i|\mathbf{x}_i, \bar{\mathbf{X}}, \bar{\mathbf{g}}, \boldsymbol{\theta}) \\ &= \prod_{i=1}^{N} \mathcal{N}\left(\mathbf{v}_i|\mathbf{K}_{\mathbf{x}_i}^T \mathbf{K}_{MD}^{-1} \bar{\mathbf{g}}, \mathbf{K}(\mathbf{x}_i, \mathbf{x}_i) - \mathbf{K}_{\mathbf{x}_i}^T \mathbf{K}_{MD}^{-1} \mathbf{K}_{\mathbf{x}_i} + \Sigma\right) \\ &= \mathcal{N}\left(\mathbf{v}|\mathbf{K}_{NM} \mathbf{K}_{MD}^{-1} \bar{\mathbf{g}}, \boldsymbol{\Lambda} + \boldsymbol{\Sigma}\right), \end{aligned} \tag{5.27}$$

where $\boldsymbol{\Lambda} \in \mathbb{R}^{ND \times ND}$ is a block diagonal matrix with the matrices $\{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, ..., \boldsymbol{\lambda}_N\}$, $\boldsymbol{\lambda}_i = \mathbf{K}(\mathbf{x}_i, \mathbf{x}_i) - \mathbf{K}_{\mathbf{x}_i}^T \mathbf{K}_{MD}^{-1} \mathbf{K}_{\mathbf{x}_i} \in \mathbb{R}^{D \times D}$,

$$\mathbf{K}_{NM} = \begin{bmatrix} \mathbf{K}_{\mathbf{x}_1}^T \\ \mathbf{K}_{\mathbf{x}_2}^T \\ \vdots \\ \mathbf{K}_{\mathbf{x}_N}^T \end{bmatrix}_{ND \times MD}.$$

The approximated marginal likelihood is given by

$$\begin{aligned} & p(\mathbf{v}|\mathbf{X}, \bar{\mathbf{X}}, \boldsymbol{\theta}) \\ &= \int p(\mathbf{v}|\mathbf{X}, \bar{\mathbf{X}}, \bar{\mathbf{g}}, \boldsymbol{\theta})\, p(\bar{\mathbf{g}}|\bar{\mathbf{X}}, \boldsymbol{\theta})\, d\bar{\mathbf{g}} \\ &= \int \mathcal{N}\left(\mathbf{v}|\mathbf{K}_{NM} \mathbf{K}_{MD}^{-1} \bar{\mathbf{g}}, \boldsymbol{\Lambda} + \boldsymbol{\Sigma}\right) \mathcal{N}(\bar{\mathbf{g}}|\mathbf{0}, \mathbf{K}_{MD})\, d\bar{\mathbf{g}} \\ &= \mathcal{N}\left(\mathbf{v}|\mathbf{0}, \mathbf{K}_{NM} \mathbf{K}_{MD}^{-1} \mathbf{K}_{NM}^T + \boldsymbol{\Lambda} + \boldsymbol{\Sigma}\right). \end{aligned} \tag{5.28}$$

Taking logarithm on both sides of (5.28), we have

$$\log p(\mathbf{v}|\mathbf{X}, \bar{\mathbf{X}}, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{v}^T \mathbf{Q}^{-1} \mathbf{v} - \frac{1}{2}\log|\mathbf{Q}| - \frac{ND}{2}\log 2\pi, \tag{5.29}$$

where $\mathbf{Q} = \mathbf{K}_{NM} \mathbf{K}_{MD}^{-1} \mathbf{K}_{NM}^T + \boldsymbol{\Lambda} + \boldsymbol{\Sigma}$. In general, the assumption in (5.27) is not true. We need to find $\bar{\mathbf{X}}$ so that the product of marginals is as close to the joint distribution as possible. Therefore, both the hyperparameters $\boldsymbol{\theta}$ and pseudo-input points $\bar{\mathbf{X}}$ need to be optimized by maximizing (5.29). Let $\boldsymbol{\alpha} = \{\boldsymbol{\theta}, \bar{\mathbf{X}}\}$ be all the

parameters we need to estimate. Genetic algorithms [107] are adopted to find the global optimum ($\boldsymbol{\alpha}^\star$) of the above log-likelihood function.

Note that to compute (5.29), $\mathbf{Q}^{-1}$ is required. Using the Woodbury matrix inversion lemma [108], we have

$$\begin{aligned}\mathbf{Q}^{-1} &= \left(\boldsymbol{\Lambda} + \boldsymbol{\Sigma} + \mathbf{K}_{NM}\,\mathbf{K}_{MD}^{-1}\,\mathbf{K}_{NM}^{T}\right)^{-1} \\ &= (\boldsymbol{\Lambda}+\boldsymbol{\Sigma})^{-1} - (\boldsymbol{\Lambda}+\boldsymbol{\Sigma})^{-1}\,\mathbf{K}_{NM}\,\mathbf{R}^{-1}\,\mathbf{K}_{NM}^{T}\,(\boldsymbol{\Lambda}+\boldsymbol{\Sigma})^{-1},\end{aligned} \tag{5.30}$$

where $\mathbf{R} = \mathbf{K}_{MD} + \mathbf{K}_{NM}^{T}\,(\boldsymbol{\Lambda}+\boldsymbol{\Sigma})^{-1}\,\mathbf{K}_{NM}$. Since $\boldsymbol{\Sigma}$ is diagonal and $\boldsymbol{\Lambda}$ is block diagonal, the complexity of computing $(\boldsymbol{\Lambda}+\boldsymbol{\Sigma})^{-1}$ is $O(ND^3)$. Here, the computationally intensive step is the inversion of the $MD \times MD$ matrix $\mathbf{R}$, whose complexity is $O(M^3D^3)$. So by using pseudo-input points and the approximate marginal likelihood, we have reduced the complexity from inverting a matrix of dimension ND with complexity $O(N^3D^3)$ to inverting a matrix of dimension MD with complexity $O(M^3D^3)$.

By tuning the upper airway sensitivity parameter value, the cardiorespiratory system model can generate the state sequences of both normal sleep and OSA conditions. Using the generated state sequences as training data, the corresponding optimal GP hyperparameters ($\boldsymbol{\alpha}^\star$) under both conditions can be estimated. Let $\hat{\boldsymbol{\alpha}}_1$ and $\hat{\boldsymbol{\alpha}}_0$ be the sets of estimated hyperparameters corresponding to the OSA and normal sleep conditions, respectively.

5.5 Sleep Apnea Detection

We propose to detect sleep apnea using the heart rate and peripheral oxygen saturation (SpO_2) measurement signals. The heart rate measurement signal can be modeled as a measurement of the heart rate state, and the SpO_2 measurement signal can be modeled as a delayed measurement of the VSO_2 state. We model the sensor measurement noise as additive white Gaussian noise (AWGN). Accordingly, the measurement model is given by,

$$\mathbf{z}_{t+1} = \mathbf{H}\,\mathbf{x}_{t+1} + \mathbf{n}, \quad t \geq 0, \tag{5.31}$$

where $\mathbf{x}_t \in \mathbb{R}^D$ is the state vector at time t, $\mathbf{H} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \end{bmatrix}$ is a $2 \times D$ matrix, and $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \Sigma_n)$ is the independent and identically distributed (i.i.d.) measurement noise vector following Gaussian distribution with zero mean and covariance matrix Σ_n.

We first divide the entire measurement signal vector into a set of non-overlapping windows of fixed duration. Then, each window is classified into two classes: apneic and non-apneic (normal).

From the state-space model in (5.16), we have

$$\begin{aligned} p(\mathbf{x}_{t+1}|\mathbf{x}_t, \hat{\boldsymbol{\theta}}) &= \mathcal{N}(\mathbf{x}_{t+1}|\boldsymbol{\mu}(\mathbf{x}_t), \mathbf{K}(\mathbf{x}_t, \mathbf{x}_t) + \Sigma) \\ &= \mathcal{N}(\mathbf{x}_{t+1}|\boldsymbol{\mu}(\mathbf{x}_t), \mathbf{K}_D + \Sigma), \end{aligned} \tag{5.32}$$

where $\hat{\boldsymbol{\theta}}$ are the estimated hyperparameters. From the measurement model in (5.31), we have

$$p(\mathbf{z}_{t+1}|\mathbf{x}_{t+1}) = \mathcal{N}(\mathbf{z}_{t+1}|\mathbf{H}\,\mathbf{x}_{t+1}, \Sigma_n). \tag{5.33}$$

Given a test window of length L, $\{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_L\}$, we perform the detection using the likelihood ratio test as follows.

Let $p(\mathbf{x}_0)$ be the initial state distribution and $t \in \{1, 2, ..., L-1\}$ be the time index. At time t, assume the updated state distribution under hypothesis $H \in \{0, 1\}$ is given by $p(\mathbf{x}_t|Z_t, \hat{\boldsymbol{\theta}})$, where $Z_t = \{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_t\}$ is the set of measurements till time t. The predicted state distribution for time $t+1$ is given by

$$\begin{aligned} p(\mathbf{x}_{t+1}|Z_t, \hat{\boldsymbol{\theta}}) &= \int p(\mathbf{x}_{t+1}|\mathbf{x}_t, \hat{\boldsymbol{\theta}})\, p(\mathbf{x}_t|Z_t, \hat{\boldsymbol{\theta}})\, d\mathbf{x}_t \\ &\approx \frac{1}{S} \sum_{i=1}^{S} p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \hat{\boldsymbol{\theta}}) \\ &= \frac{1}{S} \sum_{i=1}^{S} \mathcal{N}(\mathbf{x}_{t+1}|\boldsymbol{\mu}(\mathbf{x}_t^i), \mathbf{K}_D + \Sigma), \end{aligned} \tag{5.34}$$

where samples $\{\mathbf{x}_t^1, \mathbf{x}_t^2, ..., \mathbf{x}_t^S\}$ drawn from the distribution $p(\mathbf{x}_t|Z_t, \hat{\boldsymbol{\theta}})$ are used to simplify the integration. The predicted measurement distribution for time $t+1$ is

given by

$$
\begin{aligned}
& p(\mathbf{z}_{t+1}|Z_t, \hat{\boldsymbol{\theta}}) \\
&= \int p(\mathbf{z}_{t+1}|\mathbf{x}_{t+1})\, p(\mathbf{x}_{t+1}|Z_t, \hat{\boldsymbol{\theta}})\, d\mathbf{x}_{t+1} \\
&= \frac{1}{S}\sum_{i=1}^{S} \int p(\mathbf{z}_{t+1}|\mathbf{x}_{t+1})\, p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \hat{\boldsymbol{\theta}})\, d\mathbf{x}_{t+1} \\
&= \frac{1}{S}\sum_{i=1}^{S} \int \mathcal{N}(\mathbf{z}_{t+1}|\mathbf{H}\,\mathbf{x}_{t+1}, \Sigma_n)\, \mathcal{N}(\mathbf{x}_{t+1}|\boldsymbol{\mu}(\mathbf{x}_t^i), \mathbf{K}_D + \Sigma)\, d\mathbf{x}_{t+1} \\
&= \frac{1}{S}\sum_{i=1}^{S} \mathcal{N}\left(\mathbf{z}_{t+1}|\mathbf{H}\,\boldsymbol{\mu}(\mathbf{x}_t^i), \mathbf{H}\,(\mathbf{K}_D + \Sigma)\,\mathbf{H}^T + \Sigma_n\right).
\end{aligned}
\tag{5.35}
$$

When the measurement $\mathbf{z}_{t+1}$ at time $t+1$ becomes available, from the Bayes' theorem, the state distribution can be updated,

$$
\begin{aligned}
& p(\mathbf{x}_{t+1}|Z_{t+1}, \hat{\boldsymbol{\theta}}) \\
&\propto p(\mathbf{z}_{t+1}|\mathbf{x}_{t+1}) \cdot p(\mathbf{x}_{t+1}|Z_t, \hat{\boldsymbol{\theta}}) \\
&\propto \frac{1}{S}\sum_{i=1}^{S} p(\mathbf{z}_{t+1}|\mathbf{x}_{t+1}) \cdot p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \hat{\boldsymbol{\theta}}) \\
&\propto \frac{1}{S}\sum_{i=1}^{S} \mathcal{N}(\mathbf{z}_{t+1}|\mathbf{H}\,\mathbf{x}_{t+1}, \Sigma_n) \cdot \mathcal{N}(\mathbf{x}_{t+1}|\boldsymbol{\mu}(\mathbf{x}_t^i), \mathbf{K}_D + \Sigma) \\
&\propto \frac{1}{S}\sum_{i=1}^{S} \mathcal{N}\left(\mathbf{x}_{t+1}|\boldsymbol{\mu}_{t+1}^i, \boldsymbol{\Sigma}_{t+1}\right),
\end{aligned}
\tag{5.36}
$$

where

$$
\begin{aligned}
\boldsymbol{\mu}_{t+1}^i &= \boldsymbol{\mu}(\mathbf{x}_t^i) + \mathbf{R}\,(\mathbf{z}_{t+1} - \mathbf{H}\,\boldsymbol{\mu}(\mathbf{x}_t^i)), \\
\boldsymbol{\Sigma}_{t+1} &= \mathbf{K}_D + \Sigma - \mathbf{R}\,\mathbf{H}\,(\mathbf{K}_D + \Sigma)^T, \\
\mathbf{R} &= (\mathbf{K}_D + \Sigma)\,\mathbf{H}^T\,\left(\mathbf{H}\,(\mathbf{K}_D + \Sigma)\,\mathbf{H}^T + \Sigma_n\right)^{-1}.
\end{aligned}
$$

Finally, the test window, $\{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_L\}$, is classified as apneic using the likelihood ratio test,

$$
\sum_{t=1}^{L-1} \log \frac{p(\mathbf{z}_{t+1}|Z_t, \hat{\boldsymbol{\theta}}_1)}{p(\mathbf{z}_{t+1}|Z_t, \hat{\boldsymbol{\theta}}_0)} > \gamma,
\tag{5.37}
$$

where γ is a threshold. The threshold γ is set based on the validation performance. Based on the classification result of the current window, the initial state distribution

for the next window is chosen. That is, if the current window is classified as apneic, then for the next window we set $p(\mathbf{x}_0) = p(\mathbf{x}_L|Z_L, \hat{\boldsymbol{\theta}}_1)$. If the current window is classified as normal, then we set $p(\mathbf{x}_0) = p(\mathbf{x}_L|Z_L, \hat{\boldsymbol{\theta}}_0)$ for the next window. For the first window, we assume the initial state distribution, $p(\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, where $\boldsymbol{\mu}_0$ is set using the state values from the system model under normal condition and $\boldsymbol{\Sigma}_0$ is set as an identity matrix.

5.6 Experimental Results

We choose the five states ($D = 5$) in the state-space model: $PaCO_2$, SaO_2, ABP, HR, and VSO_2. We simulate the cardiorespiratory system model in [11] in Simulink and generate the state sequences under both normal sleep and OSA conditions at a sampling rate of 1 Hz. Since we use only five states in our work, we cannot directly use the cardiorespiratory system model equations to compute the GP mean functions as the system model equations depend on other states. Hence, we use artificial neural networks to approximate the GP mean functions ($\boldsymbol{\mu}(\cdot)$) of these five states under both normal sleep and OSA conditions. We use the standard feedforward neural network with one hidden layer of twenty neurons and hyperbolic tangent sigmoid transfer function. The Levenberg-Marquardt backpropagation algorithm is used for training. Since the state values have different ranges, we normalize each state data to $[-1, 1]$.

Next, we estimate the GP covariance hyperparameters by maximizing the approximated marginal log-likelihood function in (5.29). Note that the total number of hyperparameters depends on the number of kernels in the GP covariance matrix (Q) and the number of pseudo-input points (M). We try different sets of Q and M values and pick the best set based on the validation performance. Specifically, $Q = \{2, 3, 4\}$ and $M = \{60, 120, 180\}$ are experimented.

5.6.1 Synthetic data

One hour of both normal and apnea data for training and one hour of both normal and apnea data for testing are generated by simulating the cardiorespiratory system model. To test the robustness of the proposed OSA detection method, we add Gaussian noise with zero mean and variance σ_n^2 to the heart rate and SpO_2 data. Note that here in the synthetic datasets, the generated SpO_2 signal is already synchronized.

We also compare the performance of our proposed OSA detection method with other data-driven OSA detection methods in [97] and [98]. In [97], recurrence quantification analysis (RQA) is performed for every one-minute heart rate variability (HRV) data window, and the corresponding RQA statistics such as recurrence rate, determinism, maximum diagonal line length, maximum vertical line length, entropy of the distribution of diagonal lines, laminarity, mean diagonal line length, etc. are used for OSA/normal classification. A fifteen-neuron hidden layer neural network is used for classification. In [98], time domain features such as mean, variance, maximum, minimum, deviations of mean and median values from the maximum and minimum values, etc. are extracted for each one-minute window of the SpO_2 signal. Support vector machine with the Gaussian kernel function is used for classification. The detection performances of the proposed and the data-driven methods are presented in Table 5.3 and Table 5.4. From the results, we can see that the proposed OSA detection method outperforms the data-driven methods for various noise levels. This demonstrates the advantage of combining the cardiorespiratory system model information with the sensor measurement data. Note that the detection performance of data-driven methods greatly depends on the quality of the training data.

Table 5.3: Comparison of detection performances on simulated data using heart rate measurement signal only

		Sensitivity (True Positive Rate)	Specificity (True Negative Rate)	Accuracy
$\sigma_n = 0.8$	Fusion of model and measurements	93.33 %	95 %	94.17 %
	Measurements only [97]	88.33 %	90 %	89.17 %
$\sigma_n = 0.9$	Fusion of model and measurements	93.33 %	88.33 %	90.83 %
	Measurements only [97]	80 %	88.33 %	84.17 %
$\sigma_n = 1$	Fusion of model and measurements	86.67 %	83.33 %	85 %
	Measurements only [97]	76.67 %	80 %	78.33 %

Table 5.4: Comparison of detection performances on simulated data using SpO_2 measurement signal only

		Sensitivity (True Positive Rate)	Specificity (True Negative Rate)	Accuracy
$\sigma_n = 0.8$	Fusion of model and measurements	96.67 %	93.33 %	95 %
	Measurements only [98]	95 %	86.67 %	90.83 %
$\sigma_n = 0.9$	Fusion of model and measurements	95 %	88.33 %	91.67 %
	Measurements only [98]	88.33 %	85 %	86.67 %
$\sigma_n = 1$	Fusion of model and measurements	88.33 %	83.33 %	85.83 %
	Measurements only [98]	81.67 %	78.33 %	80 %

The detection performance of the proposed method using both heart rate and SpO_2 measurement signals is shown in Table 5.5. From the results, we can notice the improvement in detection performance of the proposed method when using both heart rate and SpO_2 measurement signals. Since the simulated data do not have the ECG signal, we do not have comparison results with the data-driven method in this case.

Table 5.5: Detection performance of the proposed method on simulated data using both heart rate and SpO_2 measurement signals

		Sensitivity (True Positive Rate)	Specificity (True Negative Rate)	Accuracy
$\sigma_n = 0.8$	Fusion of model and measurements	98.33 %	95 %	96.67 %
$\sigma_n = 0.9$	Fusion of model and measurements	95 %	91.67 %	93.33 %
$\sigma_n = 1$	Fusion of model and measurements	90 %	88.33 %	89.17 %

5.6.2 Real data

We also test the performance of the proposed method using the real data from the Apnea-ECG database in PhysioNet [13]. The database contains digitized ECG signals of 32 subjects with OSA. For eight subjects, four additional signals are available: SpO_2, oronasal airflow signal, chest and abdominal respiratory inductance plethysmography signals. For each subject, normal/apnea labels are provided at 60-second intervals by sleep experts. Each label indicates whether an apnea episode is in progress or not at the beginning of the corresponding minute. In our experiments, we consider the first four subjects (a01-a04) out of the eight subjects as they have both normal and apnea data.

First, we compute the heart rate signal from the available ECG signal as follows. From the R peaks of the ECG signal, we get the RR interval time series signal. From the RR interval signal, we calculate the corresponding heart rate signal as,

$$HR(t) = \frac{60}{RR(t)}, \tag{5.38}$$

where $HR(t)$ is the heart rate (in beats per minute) at time t and $RR(t)$ is the RR interval duration (in seconds) at time t. We then resample the heart rate signal at a sampling rate of 1 Hz using linear interpolation. In practice, the measured SpO_2 signal has a certain time delay. To estimate the time delay for each subject, we fit the proposed SpO_2-respiration model with different delay values and then choose the time delay with the minimum mean squared error. We then time shift the SpO_2 measurement signal according to the corresponding time delay.

Since the given labels in the database indicate the presence or absence of an apnea episode at a particular time instant instead of in a window, we perform OSA detection every one second, i.e., we set $L = 1$. To test the performance of OSA detection methods under situations with limited availability of labeled data, we consider different scenarios as follows:

- Scenario A: For each subject, we use one hour of data (around 12% of data) for training and three hours of data for testing.
- Scenario B: For each subject, we use three hours of data (around 35% of data) for training and three hours of data for testing.

The detection performances of the proposed method and the data-driven methods are presented in Tables 5.6 to 5.8. Note that the proposed method does not require large amount of training data. In our experiments, we use a small portion of the data (around 30-35 minutes per subject) to set the threshold γ and estimate the SpO_2 time delay for each subject. In general, the proposed method can be directly used in situations with no training data available.

From the results, we can see that the proposed method outperforms the data-driven methods, especially in situations of limited availability of labeled data for training. This is usually the case in practice due to the limited number of sleep medicine specialists and sleep labs. Data-driven methods often assume the availability

of a lot of labeled data for training, and they do not perform well when the amount of labeled data available is low. Since we combine the sensor measurement signals with the cardiorespiratory system model information, the proposed method does not require large amount of labeled training data.

Table 5.6: Comparison of detection performances on real data using ECG measurement signal only

	Sensitivity (True Positive Rate)	Specificity (True Negative Rate)	Accuracy
Fusion of model and measurements	85.38 %	70.41 %	82.44 %
Measurements only [97] (Scenario A)	46.56 %	34.64 %	44.03 %
Measurements only [97] (Scenario B)	86.42 %	77.10 %	84.72 %

Table 5.7: Comparison of detection performances on real data using SpO_2 measurement signal only

	Sensitivity (True Positive Rate)	Specificity (True Negative Rate)	Accuracy
Fusion of model and measurements	87.82 %	72.82 %	84.86 %
Measurements only [98] (Scenario A)	62.43 %	50.33 %	59.86 %
Measurements only [98] (Scenario B)	73.34 %	71.76 %	73.06 %

Table 5.8: Comparison of detection performances on real data using both ECG and SpO_2 measurement signals

	Sensitivity (True Positive Rate)	Specificity (True Negative Rate)	Accuracy
Fusion of model and measurements	90.46 %	76.71 %	87.76 %
Measurements only [96] (Scenario A)	65.08 %	53.59 %	62.64 %
Measurements only [96] (Scenario B)	77.59 %	73.28 %	76.81 %

From the results, we can also see that the proposed method has better detection performance when using both heart rate (from ECG) and SpO_2 measurement signals.

This is due to the fact that sometimes when one measurement signal is more noisy, the other measurement signal with less noise can help in increasing the detection accuracy. For example, some apnea segments of the data that are detected only with both heart rate and SpO_2 signals are shown in Figure 5.10. These segments are not detected using the heart rate signal alone. From the figure, we can see that the heart rate signal is more noisy and hence it is difficult to detect these apnea segments using the heart rate signal only.

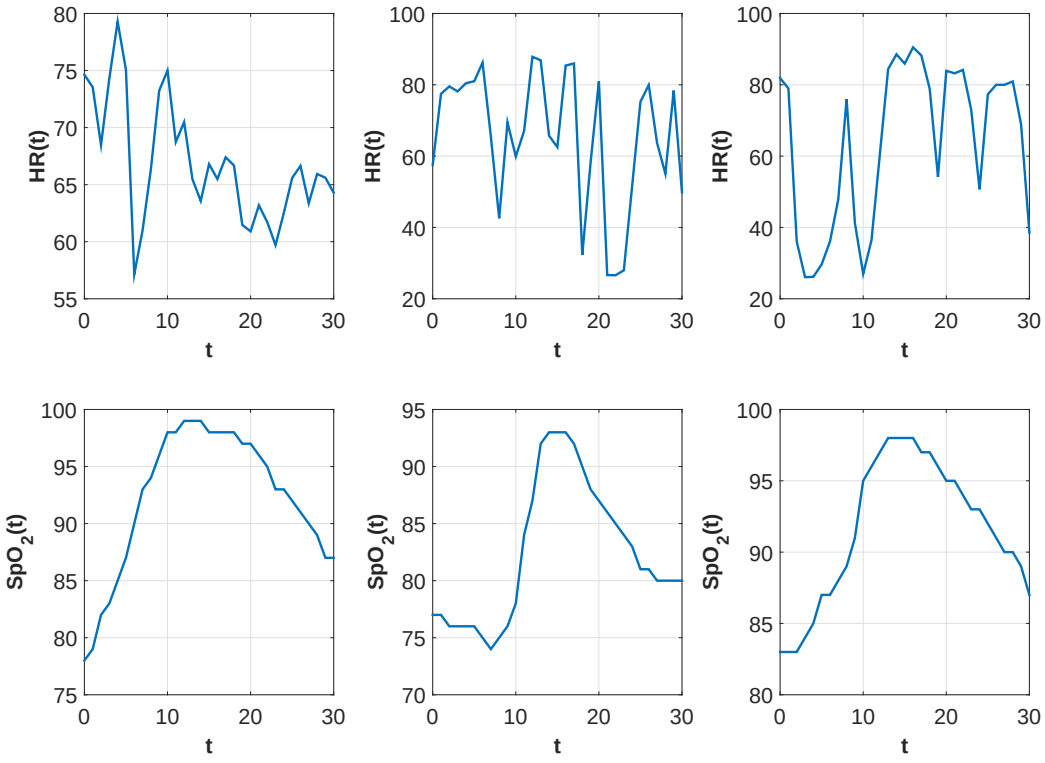


Figure 5.10: Apnea segments of the real data.

5.7 Summary

In this chapter, a new framework for OSA detection is proposed, in which the cardiorespiratory system model information is combined with the sensor measurement data. A GP-based state-space model is proposed using the existing mathematical

models of the cardiorespiratory system. A mathematical model relating the SpO_2 signal to respiration signal is developed. A likelihood ratio test is developed to detect OSA on a window-by-window basis using both heart rate and SpO_2 measurement signals. Experimental results on both synthetic and real data demonstrate the effectiveness of the proposed OSA detection framework over existing purely data-driven detection methods. Comparisons with the purely data-driven methods, especially in situations of limited training data, show the advantage of combining the cardiorespiratory system model information with the sensor measurement data in the proposed framework.

CHAPTER 6

CONCLUSIONS

Development of a noninvasive smart health monitoring system capable of providing accurate diagnostic information is a critical problem. In this research, different important aspects of the monitoring system are considered such as physiological signal processing, control mechanism modeling, and disease detection. We mainly focused on the detection of sleep apnea, a serious chronic sleep disorder affecting the cardiorespiratory system.

In Chapter 2, we focused on the design of a biometric recognition system using the ECG signal. A new joint feature extraction and classifier design method is proposed for the biometric recognition problem. Non-fiducial features are extracted from the ECG signal. The biometric classification problem is converted into a set of binary classification problems. A multitask learning framework in which feature extraction and classifier design for all the binary classification tasks are conducted simultaneously. For each binary classification task, probabilistic nonlinear kernel classifiers are used in which the features are weighted according to their relevance. The matrix consisting of the feature weights for all the tasks is decomposed into a sparse component and a low-rank component. The sparse component gives the features relevant to each classification task, and the low-rank component gives the common feature subspace relevant to all the classification tasks. Experimental results on the real data demonstrate the effectiveness of the proposed biometric recognition system.

In Chapter 4, modeling the local control mechanism of the cardiorespiratory system transiting from awake state to stage 4 non-REM sleep state is considered. A 13-

state discrete-time cardiorespiratory system model with transport delays is adopted. The system model is nonlinear and described by a set of higher order difference equations. We converted the system into a set of first-order difference equations and proposed an iterative algorithm to find the optimal control inputs that drive the cardiorespiratory system from awake state to sleep state. In each iteration, the system is linearized using the state and input sequences from the previous iteration. Simulation results show the effectiveness of the proposed control mechanism model. We validate the simulation results using real data collected from a healthy subject.

The focus of Chapter 5 is the detection of sleep apnea. Sleep apnea is a serious sleep disorder affecting millions of people worldwide. We proposed a new framework for sleep apnea detection in which we combine the multiple sensor data with the physiological signal information from the cardiorespiratory system models. Since the cardiorespiratory models represent the physiological dynamics in general, Gaussian processes (GPs) are used to capture the physiological variations among different individuals. We adopt the sum of separable kernel functions form to define the vector-valued GP covariance function and estimate the hyperparameters by maximizing the GP marginal likelihood function. To reduce the computational complexity, we construct an approximated GP marginal likelihood function using a small set of pseudo-input points and then estimate the hyperparameters by maximizing the approximated marginal likelihood function. Sleep apnea detection using the heart rate and peripheral oxygen saturation (SpO_2) measurement signals is performed on a window-by-window basis using a likelihood ratio test. Experimental results on both synthetic and real data show the effectiveness of the proposed detection framework. Comparisons with other data-driven sleep apnea detection methods demonstrate the advantage of combining the sensor measurement data and the cardiorespiratory model information in the proposed framework.

6.1 Future Work

Here, we provide some suggestions for future work:

- In Chapter 2, we developed a biometric recognition system based on the ECG signal. For more security, the ECG-based biometric recognition system can be extended to include other biometric traits such as iris, retina, etc.

- In Chapters 4 and 5, we considered general cardiorespiratory system models. In the future, we will investigate the effect of tuning the system models to a particular person using sensitivity analysis and simulation experiments.

- In the future, we will work on developing real-time sleep apnea prediction methods by fusing the measurements and cardiorespiratory system model information. By predicting the occurrence of sleep apnea episodes in real time, we can intervene and avoid apnea episodes using positive airway pressure (PAP) ventilator masks.

APPENDIX A

GAUSSIAN PROCESS

A Gaussian process (GP) is a generalization of the Gaussian probability distribution to functions. In other words, a GP is an infinite collection of random variables, where any finite number of them have a joint Gaussian distribution [106]. A GP on a function $f : \mathbb{R}^d \longmapsto \mathbb{R}$ is completely specified by its mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$,

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right), \tag{A.1}$$

where

$$\begin{aligned} m(\mathbf{x}) &= \mathbb{E}[f(\mathbf{x})], \\ k(\mathbf{x}, \mathbf{x}') &= \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))\,(f(\mathbf{x}') - m(\mathbf{x}'))]. \end{aligned} \tag{A.2}$$

Note that here the function values $f(\cdot)$ are the random variables, not the input $\mathbf{x}$. Given a finite set of inputs, $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$, the corresponding set of outputs have a joint Gaussian distribution,

$$\begin{bmatrix} f(\mathbf{x}_1) \\ f(\mathbf{x}_2) \\ \vdots \\ f(\mathbf{x}_N) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{x}_1) \\ m(\mathbf{x}_2) \\ \vdots \\ m(\mathbf{x}_N) \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & k(\mathbf{x}_1, \mathbf{x}_2) & \dots & k(\mathbf{x}_1, \mathbf{x}_N) \\ k(\mathbf{x}_2, \mathbf{x}_1) & k(\mathbf{x}_2, \mathbf{x}_2) & \dots & k(\mathbf{x}_2, \mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & k(\mathbf{x}_N, \mathbf{x}_2) & \dots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix} \right). \tag{A.3}$$

Using the compact notation, we can rewrite (A.3) as,

$$\mathbf{f}(\mathbf{X}) \sim \mathcal{N}\left(\mathbf{m}(\mathbf{X}), \mathbf{K}(\mathbf{X}, \mathbf{X})\right). \tag{A.4}$$

In practice, we often do not know the actual function $f(\cdot)$ of some real system or process. However, we have a set of observations, $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, available to us.

Generally, we assume the observed data is generated according to some model such as,

$$y_i = f(\mathbf{x}_i) + \epsilon, \tag{A.5}$$

where ϵ is the independent and identically distributed (i.i.d.) Gaussian noise with zero mean and variance σ_n^2. Now, our inference problem is to estimate the function $f(\cdot)$. In GP-based inference, we assume the function $f(\cdot)$ follows a GP with mean function $m(\cdot)$ and covariance function $k(\cdot,\cdot)$. Let $\mathbf{X}_* = \{\mathbf{x}_1^*, \mathbf{x}_2^*, ..., \mathbf{x}_M^*\}$ be the new input points (test inputs). Since any finite number of random variables of a GP are jointly Gaussian, we have

$$\begin{bmatrix} \mathbf{f}(\mathbf{X}) \\ \mathbf{f}(\mathbf{X}_*) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}(\mathbf{X}) \\ \mathbf{m}(\mathbf{X}_*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{X},\mathbf{X}) & \mathbf{K}(\mathbf{X},\mathbf{X}_*) \\ \mathbf{K}(\mathbf{X}_*,\mathbf{X}) & \mathbf{K}(\mathbf{X}_*,\mathbf{X}_*) \end{bmatrix} \right). \tag{A.6}$$

Using the conditional Gaussian identities, the GP predictive distribution is given by

$$\begin{aligned} \mathbf{f}(\mathbf{X}_*)|\mathbf{X}_*, \mathbf{X}, \mathbf{f}(\mathbf{X}) \sim \mathcal{N}(&\mathbf{m}(\mathbf{X}_*) + \mathbf{K}(\mathbf{X}_*,\mathbf{X})\,\mathbf{K}(\mathbf{X},\mathbf{X})^{-1}\,(\mathbf{f}(\mathbf{X}) - \mathbf{m}(\mathbf{X})), \\ &\mathbf{K}(\mathbf{X}_*,\mathbf{X}_*) - \mathbf{K}(\mathbf{X}_*,\mathbf{X})\,\mathbf{K}(\mathbf{X},\mathbf{X})^{-1}\,\mathbf{K}(\mathbf{X},\mathbf{X}_*)). \end{aligned} \tag{A.7}$$

Note that this is for the noise-free case. Similarly, for the noisy case, we have

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}(\mathbf{X}) \\ \mathbf{m}(\mathbf{X}_*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma_n^2\,\mathbf{I}_N & \mathbf{K}(\mathbf{X},\mathbf{X}_*) \\ \mathbf{K}(\mathbf{X}_*,\mathbf{X}) & \mathbf{K}(\mathbf{X}_*,\mathbf{X}_*) + \sigma_n^2\,\mathbf{I}_M \end{bmatrix} \right), \tag{A.8}$$

where $\mathbf{y} = [y_1, y_2, ..., y_N]^T$, $\mathbf{y}_* = [y_1^*, y_2^*, ..., y_M^*]^T$, and $\mathbf{I}_N$ is an $N \times N$ identity matrix. And the corresponding GP predictive distribution is given by

$$\mathbf{y}_*|\mathbf{X}_*, \mathbf{X}, \mathbf{y} \sim \mathcal{N}\left(\text{mean}(\mathbf{y}_*), \text{cov}(\mathbf{y}_*)\right), \tag{A.9}$$

where

$$\begin{aligned} \text{mean}(\mathbf{y}_*) &= \mathbf{m}(\mathbf{X}_*) + \mathbf{K}(\mathbf{X}_*,\mathbf{X})\,\mathbf{K}_\mathbf{y}^{-1}\,(\mathbf{y} - \mathbf{m}(\mathbf{X})), \\ \text{cov}(\mathbf{y}_*) &= \mathbf{K}(\mathbf{X}_*,\mathbf{X}_*) - \mathbf{K}(\mathbf{X}_*,\mathbf{X})\,\mathbf{K}_\mathbf{y}^{-1}\,\mathbf{K}(\mathbf{X},\mathbf{X}_*) + \sigma_n^2\,\mathbf{I}_M, \\ \mathbf{K}_\mathbf{y} &= \mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma_n^2\,\mathbf{I}_N. \end{aligned} \tag{A.10}$$

Covariance function

Generally in practice, the mean function $m(\cdot)$ is assumed to be a known deterministic function. Then, the GP is the completely specified by its covariance function $k(\cdot,\cdot)$. A function $k(\cdot,\cdot)$, mapping two input points to $\mathbb{R}$, is also referred to as kernel function. For a function $k(\cdot,\cdot)$ to be a valid kernel function (or covariance function), the following properties need to be satisfied:

- $k(\cdot,\cdot)$ is symmetric, i.e., $k(\mathbf{x},\mathbf{x}') = k(\mathbf{x}',\mathbf{x})$.
- Mercer's condition [109]: $k(\cdot,\cdot)$ should be positive semidefinite. That is, given a set of input points $\{\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_n\}$, the function $k(\cdot,\cdot)$ must satisfy

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j k(\mathbf{x}_i,\mathbf{x}_j) \geq 0,$$

  for all real numbers $\{c_1,c_2,...,c_n\}$. In other words, the $n \times n$ Gram matrix $\mathbf{K}$, with elements $K_{ij} = k(\mathbf{x}_i,\mathbf{x}_j)$, should be positive semidefinite.

Some types of covariance functions (kernel functions):

- *Stationary:* A covariance function is stationary if it is a function of $\mathbf{x}-\mathbf{x}'$. Note that the stationary kernel functions are invariant to translations in input space.
- *Isotropic (or homogeneous):* A covariance function is isotropic if it is a function of $|\mathbf{x}-\mathbf{x}'|$.

Some example covariance functions are as follows:

- Constant: $k(\mathbf{x},\mathbf{x}') = \sigma_c^2$.
- Dot product: $k(\mathbf{x},\mathbf{x}') = \sigma_d^2 + \mathbf{x}\cdot\mathbf{x}'$.
- Polynomial: $(\sigma_b^2 + \mathbf{x}\cdot\mathbf{x}')^p$.

- Exponential:

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|}{l}\right).$$

- Squared exponential (Gaussian function or radial basis function):

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\,l^2}\right).$$

- Rational quadratic:

$$k(\mathbf{x}, \mathbf{x}') = \left(1 + \frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\alpha l^2}\right)^{-\alpha}.$$

We can build new covariance functions from existing covariance functions using some properties. For example, given two covariance functions $k_1(\mathbf{x}, \mathbf{x}')$ and $k_2(\mathbf{x}, \mathbf{x}')$, the following will be valid covariance functions [50]:

- $c\,k_1(\mathbf{x}, \mathbf{x}')$, where $c > 0$.
- $k_1(\mathbf{x}, \mathbf{x}') + k_2(\mathbf{x}, \mathbf{x}')$.
- $k_1(\mathbf{x}, \mathbf{x}')\,k_2(\mathbf{x}, \mathbf{x}')$.
- $\exp\left(k_1(\mathbf{x}, \mathbf{x}')\right)$.
- $f(\mathbf{x})\,k_1(\mathbf{x}, \mathbf{x}')\,f(\mathbf{x}')$, where $f(\cdot)$ is any function.
- $g\left(k_1(\mathbf{x}, \mathbf{x}')\right)$, where $g(\cdot)$ is a polynomial with nonnegative coefficients.

Hyperparameter estimation

Note that the GP covariance function contains some free parameters (hyperparameters). These hyperparameters ($\boldsymbol{\theta}$) can be estimated by maximizing the marginal likelihood function $p(\mathbf{y}|\mathbf{X})$. The marginal likelihood is of the data $\{\mathbf{X}, \mathbf{y}\}$ is given by

$$p(\mathbf{y}|\mathbf{X}) = \int p(\mathbf{y}|\mathbf{f})\,p(\mathbf{f}|\mathbf{X})\,d\mathbf{f}, \tag{A.11}$$

where $\mathbf{f} = \mathbf{f}(\mathbf{X})$. Note that both the likelihood $p(\mathbf{y}|\mathbf{f})$ and prior $p(\mathbf{f}|\mathbf{X})$ are Gaussian,

$$\begin{aligned} \mathbf{f}|\mathbf{X} &\sim \mathcal{N}(\mathbf{m}(\mathbf{X}), \mathbf{K}(\mathbf{X}, \mathbf{X})), \\ \mathbf{y}|\mathbf{f} &\sim \mathcal{N}(\mathbf{f}, \sigma_n^2 \mathbf{I}_N). \end{aligned} \tag{A.12}$$

Hence, the marginal likelihood $p(\mathbf{y}|\mathbf{X})$ is also a Gaussian,

$$\mathbf{y}|\mathbf{X} \sim \mathcal{N}(\mathbf{m}(\mathbf{X}), \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}_N). \tag{A.13}$$

Note that the GP can be extended to more general vector-valued functions ($f : \mathbb{R}^d \longmapsto \mathbb{R}^b$) as well.

REFERENCES

[1] W. W. Flemons, N. J. Douglas, S. T. Kuna, D. O. Rodenstein, and J. Wheatley, "Access to diagnosis and treatment of patients with suspected sleep apnea," *American Journal of Respiratory and Critical Care Medicine*, vol. 169, no. 6, pp. 668–672, 2004.

[2] F. S. Grodins, "Integrative cardiovascular physiology: A mathematical synthesis of cardiac and blood vessel hemodynamics," *The Quarterly Review of Biology*, vol. 34, pp. 93–116, June 1959.

[3] A. C. Guyton, T. G. Coleman, and H. J. Granger, "Circulation: overall regulation," *Annual Review of Physiology*, vol. 34, pp. 13–46, Mar. 1972.

[4] F. Kappel and R. O. Peer, "A mathematical model for fundamental regulation processes in the cardiovascular system," *Journal of Mathematical Biology*, vol. 31, no. 6, pp. 611–631, 1993.

[5] V. L. Rolle, A. I. Hernandez, P. Y. Richard, J. Buisson, and G. Carrault, "A bond graph model of the cardiovascular system," *Acta Biotheoretica*, vol. 53, no. 4, pp. 295–312, 2005.

[6] F. Liang and H. Liu, "Simulation of hemodynamic responses to the valsalva maneuver: an integrative computational model of the cardiovascular system and the autonomic nervous system," *The Journal of Physiological Sciences*, vol. 56, no. 1, pp. 45–65, 2006.

[7] F. S. Grodins, J. S. Gray, K. R. Schroeder, A. L. Norins, and R. W. Jones, "Respiratory responses to CO2 inhalation; a theoretical study of a nonlinear

biological regulator," *Journal of Applied Physiology*, vol. 7, no. 3, pp. 283–308, 1954.

[8] F. S. Grodins, J. Buell, and A. J. Bart, "Mathematical analysis and digital simulation of the respiratory control system," *Journal of Applied Physiology*, vol. 22, no. 2, pp. 260–276, 1967.

[9] J. F. Golden, J. W. Clark, and P. M. Stevens, "Mathematical modeling of pulmonary airway dynamics," *IEEE Transactions on Biomedical Engineering*, vol. 20, no. 6, pp. 397–404, 1973.

[10] T. Aittokallio, M. Gyllenberg, O. Polo, and A. Virkki, "Parameter estimation of a respiratory control model from noninvasive carbon dioxide measurements during sleep," *Mathematical Medicine and Biology*, vol. 24, pp. 225–249, June 2007.

[11] L. Cheng, O. Ivanova, H.-H. Fan, and M. C. Khoo, "An integrative model of respiratory and cardiovascular control in sleep-disordered breathing," *Respiratory Physiology and Neurobiology*, vol. 174, pp. 4–28, Nov. 2010.

[12] S. C. Wallace, "Assessment tools help diagnose obstructive sleep apnea," *Pennsylvania Patient Safety Advisory*, vol. 11, pp. 168–171, Dec. 2014.

[13] A. L. Goldberger, L. Amaral, L. Glass, J. M. Hausdorff, P. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[14] CDC, "Health expenditures." [Online]. Available: `http://www.cdc.gov/nchs/fastats/health-expenditures.htm`.

[15] PCAST, "Better health care and lower costs: Accelerating improvement through systems engineering (Report to the President)." [Online]. Available: https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_systems_engineering_in_healthcare_-_may_2014.pdf.

[16] NAM, "Crossing the quality chasm: A new health system for the 21st century." [Online]. Available: https://www.nationalacademies.org/hmd/~/media/Files/Report%20Files/2001/Crossing-the-Quality-Chasm/Quality%20Chasm%202001%20%20report%20brief.pdf.

[17] S. Wegner, A. Cowsky, C. Davis, D. James, D. Yang, R. Fontaine, and J. Morrison, "Apple iPhone 7 teardown." [Online]. Available: http://www.chipworks.com/about-chipworks/overview/blog/apple-iphone-7-teardown.

[18] A. L. Shimpi, "Arm partners ship 50 billion chips since 1991 - where did they go?." [Online]. Available: http://www.anandtech.com/show/7909/arm-partners-ship-50-billion-chips-since-1991-where-did-they-go.

[19] Apple, "Apple introduces ResearchKit, giving medical researchers the tools to revolutionize medical studies." [Online]. Available: http://www.apple.com/pr/library/2015/, Mar. 2015.

[20] Apple, "Apple advances health apps with CareKit." [Online]. Available: http://www.apple.com/pr/library/2016/, Mar. 2016.

[21] E. Singer, "The measured life," *MIT Technology Review*, June 2011.

[22] H. J. Wilson, "You, by the numbers," *Harvard Business Review*, Sept. 2012.

[23] CarreTechnologies, "Hexoskin." [Online]. Available: http://www.hexoskin.com/.

[24] AliveCor, "New Kardia Band for Apple Watch delivers medical-grade electrocardiogram (EKG) anytime, anywhere." [Online]. Available: https://www.alivecor.com/press/pressrelease/new-kardia/, Mar. 2016.

[25] Beddit, "Beddit Sleep Tracker." [Online]. Available: http://www.beddit.com/.

[26] Nymi, "Nymi Band." [Online]. Available: https://nymi.com/.

[27] Omron, "Project Zero." [Online]. Available: https://omronhealthcare.com/blood-pressure/zero/.

[28] G. Acampora, D. J. Cook, P. Rashidi, and A. V. Vasilakos, "A survey on ambient intelligence in healthcare," *Proceedings of the IEEE*, vol. 101, pp. 2470–2494, Dec. 2013.

[29] S.-W. Lee and K. Mase, "Activity and location recognition using wearable sensors," *IEEE Pervasive Computing*, vol. 1, pp. 24–32, July-Sept. 2002.

[30] N. Kern, B. Schiele, H. Junker, P. Lukowicz, and G. Troster, "Wearable sensing to annotate meeting recordings," *Personal and Ubiquitous Computing*, vol. 7, pp. 263–274, Oct. 2003.

[31] P. Lukowicz, J. A. Ward, H. Junker, and T. Starner, "Recognizing workshop activity using bodyworn microphones and accelerometers," in *Proc. International Conference on Pervasive Computing*, pp. 18–23, Apr. 2004.

[32] P. Rashidi, D. Cook, L. Holder, and M. Schmitter-Edgecombe, "Discovering activities to recognize and track in a smart environment," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, pp. 527–539, Apr. 2011.

[33] F. Gouaux, L. Simon-Chautemps, J. Fayn, S. Adami, M. Arzi, D. Assanelli, M. Forlini, C. Malossi, A. Martinez, J. Placide, G. L. Ziliani, and P. Rubel,

"Ambient intelligence and pervasive systems for the monitoring of citizens at cardiac risk: New solutions from the epi-medics project," in *Proc. Computers in Cardiology*, pp. 289–292, 2002.

[34] M. Sung, R. DeVaul, S. Jimenez, J. Gips, and A. Pentland, "Shiver motion and core body temperature classification for wearable soldier health monitoring systems," in *Proc. Eighth International Symposium on Wearable Computers*, pp. 192–193, 2004.

[35] K. J. Heilman and S. W. Porges, "Accuracy of the LifeShirt (Vivometrics) in the detection of cardiac rhythms," *Biological Psychology*, vol. 75, pp. 300–305, July 2007.

[36] M. Mubashir, L. Shao, and L. Seed, "A survey on fall detection: Principles and approaches," *Neurocomputing*, vol. 100, no. 16, pp. 144–152, 2013.

[37] C. Debes, A. Merentitis, S. Sukhanov, M. Niessen, N. Frangiadakis, and A. Bauer, "Monitoring activities of daily living in smart homes: Understanding human behavior," *IEEE Signal Processing Magazine*, vol. 33, pp. 81–94, Mar. 2016.

[38] A. K. Jain, A. A. Ross, and K. Nandakumar, *Introduction to Biometrics*. Springer, 2011.

[39] C. Roberts, "Biometric attack vectors and defences," *Computers and Security*, vol. 26, pp. 14–25, Feb. 2007.

[40] A. Hern, "Hacker fakes German minister's fingerprints using photos of her hands." [Online]. Available: `https://www.theguardian.com/technology/2014/dec/30/hacker-fakes-german-ministers-fingerprints-using-photos-of-her-hands`.

[41] M. Sharif, S. Bhagavatula, L. Bauer, and M. K. Reiter, "Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition," in *Proc. 23rd ACM SIGSAC Conference on Computer and Communications Security*, Oct. 2016.

[42] H. Silva, A. Lourenco, A. Fred, and J. Filipe, "Clinical data privacy and customization via biometrics based on ECG signals," *Lecture Notes in Computer Science, Information Quality in e-Health*, vol. 7058, pp. 121–132, 2011.

[43] R. Hoekema, G. J. H. Uijen, and A. V. Oosterom, "Geometrical aspects of the interindividual variability of multilead ECG recordings," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 5, pp. 551–559, 2001.

[44] Y. Wang, F. Agrafioti, D. Hatzinakos, and K. N. Plataniotis, "Analysis of human electrocardiogram for biometric recognition," *EURASIP Journal on Advances in Signal Processing*, pp. 1–11, 2008.

[45] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "ECG analysis: A new approach in human identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 3, pp. 808–812, 2001.

[46] S. A. Israel, J. M. I. A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "ECG to identify individuals," *Pattern Recognition*, vol. 38, no. 1, pp. 133–142, 2005.

[47] G. Wübbeler, M. Stavridis, D. Kreiseler, R. D. Bousseljot, and C. Elster, "Verification of humans using the electrocardiogram," *Pattern Recognition Letters*, vol. 28, no. 10, pp. 1172–1175, 2007.

[48] J. M. Irvine, S. A. Israel, W. T. Scruggs, and W. J. Worek, "eigenPulse: Robust human identification from cardiovascular function," *Pattern Recognition*, vol. 41, no. 11, pp. 3427–3435, 2008.

[49] M. M. Tawfik and H. S. T. Kamal, "Human identification using QT signal and QRS complex of the ECG," *The Online Journal on Electronics and Electrical Engineering*, vol. 3, no. 1, pp. 383–387, 2011.

[50] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[51] M. K. Delano and C. G. Sodini, "A long-term wearable electrocardiogram measurement system," in *Proc. IEEE International Conference on Body Sensor Networks (BSN)*, pp. 1–6, 2013.

[52] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik, "Feature selection for SVMs," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, pp. 668–674, 2000.

[53] Y. Grandvalet and S. Canu, "Adaptive scaling for feature selection in SVMs," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, pp. 553–560, 2002.

[54] B. Krishnapuram, A. J. Hartemink, L. Carin, and M. A. T. Figueiredo, "A bayesian approach to joint feature selection and classifier design," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1105–1111, 2004.

[55] R. Caruana, *Multitask Learning*. PhD thesis, School of Computer Science, Carnegie Mellon University, 1997.

[56] B. Bakker and T. Heskes, "Task clustering and gating for Bayesian multitask learning," *Journal of Machine Learning Research*, vol. 4, pp. 83–99, May 2003.

[57] T. Jebara, "Multi-task feature and kernel selection for SVMs," in *Proc. International Conference on Machine Learning (ICML)*, 2004.

[58] A. Argyriou, T. Evgeniou, and M. Pontil, "Convex multi-task feature learning," *Machine Learning*, vol. 73, no. 3, pp. 243–272, 2008.

[59] J. Chen, J. Liu, and J. Ye, "Learning incoherent sparse and low-rank patterns from multiple tasks," *ACM Transactions on Knowledge Discovery from Data*, vol. 5, no. 4, pp. 22:1–22:31, 2012.

[60] I. Odinaka, P.-H.Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, and J. W. Rohrbaugh, "ECG biometric recognition: a comparative analysis," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1812–1824, 2012.

[61] P. McCullagh and J. A. Nelder, *Generalized Linear Models*. Chapman and Hall, 1989.

[62] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[63] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proc. American Control Conference*, pp. 4734–4739, 2001.

[64] V. Chandrasekaranan and M. I. Jordan, "Computational and statistical tradeoffs via convex relaxation," *Proceedings of the National Academy of Sciences*, vol. 110, no. 13, pp. E1181–E1190, 2013.

[65] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[66] S. Gutta and Q. Cheng, "Data-based distributed classification and its performance analysis," in *Proc. 15th International Conference on Information Fusion*, pp. 1519–1526, 2012.

[67] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[68] J. C. Platt, "Probabilistic outputs for support vector machines and comparison to regularized likelihood methods," *Advances in Large Margin Classifiers*, pp. 61–74, 1999.

[69] J. C. Platt, "Sequential minimal optimization: a fast algorithm for training support vector machines," *Advances in Kernel Methods - Support Vector Learning*, 1998.

[70] G. M. Friesen, T. C. Jannett, M. A. Jadallah, S. L. Yates, S. R. Quint, and H. T. Nagle, "A comparison of the noise sensitivity of nine QRS detection algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 1, pp. 85–98, 1990.

[71] H. Xu, C. Caramanis, and S. Mannor, "Robustness and regularization of support vector machines," *Journal of Machine Learning Research*, vol. 10, pp. 1485–1510, July 2009.

[72] D. Saad and S. A. Solla, "Learning with noise and regularizers in multilayer neural networks," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 260–266, 1997.

[73] S. Okamoto and Y. Nobuhiro, "An average-case analysis of the k-nearest neighbor classifier for noisy domains," in *Proc. 15th International Joint Conference on Artifical Intelligence - Volume 1*, pp. 238–243, 1997.

[74] WHO, "Noncommunicable diseases fact sheet." [Online]. Available: http://www.who.int/mediacentre/factsheets/fs355/en/.

[75] CDC, "Heart disease facts." [Online]. Available: http://www.cdc.gov/heartdisease/facts.htm.

[76] J. J. Batzel, F. Kappel, and S. Timischl-Teschl, "A cardiovascular-respiratory control system model including state delay with application to congestive heart failure in humans," *Journal of Mathematical Biology*, vol. 50, pp. 293–335, Mar. 2005.

[77] M. Younes and W. Riddle, "A model for the relation between respiratory neural and mechanical outputs. i. theory," *Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology*, vol. 51, no. 4, pp. 963–978, 1981.

[78] W. Li and E. Todorov, "Iterative linear quadratic regulator design for nonlinear biological movement systems," in *Proc. First International Conference on Informatics in Control, Automation and Robotics*, 2004.

[79] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[80] J. S. Jr., "Gas exchange and hemodynamics during sleep," *Medical Clinics of North America*, vol. 69, no. 6, pp. 1243–1264, 1985.

[81] J. Krieger, N. Maglasiu, E. Sforza, and D. Kurtz, "Breathing during sleep in normal middle-aged subjects," *Sleep*, vol. 13, no. 2, pp. 143–154, 1990.

[82] M. Khoo, A. Gottschalk, and A. Pack, "Sleep-induced periodic breathing and apnea: a theoretical study," *Journal of Applied Physiology*, vol. 70, no. 5, pp. 2014–2024, 1991.

[83] V. Somers, M. Dyken, A. Mark, and F. Abboud, "Sympathetic-nerve activity during sleep in normal subjects," *New England Journal of Medicine*, vol. 328, no. 5, pp. 303–307, 1993.

[84] C. Guilleminault, A. Tilkian, and W. C. Dement, "The sleep apnea syndromes," *Annual Review of Medicine*, vol. 27, pp. 465–484, 1976.

[85] D. Leger, V. Bayon, J. P. Laaban, and P. Philip, "Impact of sleep apnea on economics," *Sleep Medicine Reviews*, vol. 16, no. 5, pp. 455–462, 2012.

[86] T. Young, L. Evans, L. Finn, and M. Palta, "Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women," *Sleep*, vol. 20, no. 9, pp. 705–706, 1997.

[87] R. S. T. Leung and T. D. Bradley, "Sleep apnea and cardiovascular disease," *American Journal of Respiratory and Critical Care Medicine*, vol. 164, no. 12, pp. 2147–2165, 2001.

[88] H. K. Yaggi, J. Concato, W. N. Kernan, J. H. Lichtman, L. M. Brass, and V. Mohsenin, "Obstructive sleep apnea as a risk factor for stroke and death," *New England Journal of Medicine*, vol. 353, no. 19, pp. 2034–2041, 2005.

[89] C. F. P. George, "Reduction in motor vehicle collisions following treatment of sleep apnoea with nasal CPAP," *Thorax*, vol. 56, no. 7, pp. 508–512, 2001.

[90] H. Gouveris, O. Selivanova, U. Bausmer, B. Goepel, and W. Mann, "First-night-effect on polysomnographic respiratory sleep parameters in patients with sleep-disordered breathing and upper airway pathology," *European Archives of Oto-Rhino-Laryngology*, vol. 267, no. 9, pp. 1449–1453, 2010.

[91] M. Tamaki, J. W. Bang, T. Watanabe, and Y. Sasaki, "Night watch in one brain hemisphere during sleep associated with the first-night effect in humans," *Current Biology*, vol. 26, no. 9, pp. 1190–1194, 2016.

[92] S. T. Kuna, *Sleep Apnea Pathogenesis, Diagnosis and Treatment*, ch. Alternative strategies for diagnosis of patients with obstructive sleep apnea. Informa Healthcare, second ed., 2012.

[93] M. O. Mendez, A. M. Bianchi, M. Matteucci, S. Cerutti, and T. Penzel, "Sleep apnea screening by autoregressive models from a single ECG lead," *IEEE Transactions on Biomedical Engineering*, vol. 56, pp. 2838–2850, Dec. 2009.

[94] A. H. Khandoker, M. Palaniswami, and C. K. Karmakar, "Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, pp. 37–48, Jan. 2009.

[95] M. Bsoul, M. Minn, and L. Tamil, "Apnea MedAssist: Real-time sleep apnea monitor using single-lead ECG," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, pp. 416–427, May 2011.

[96] B. Xie and H. Minn, "Real-time sleep apnea detection by classifier combination," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, pp. 469–477, May 2012.

[97] H. D. Nguyen, B. A. Wilkins, Q. Cheng, and B. A. Benjamin, "An online sleep apnea detection method based on recurrence quantification analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, pp. 1285–1293, July 2014.

[98] B. L. Koley and D. Dey, "On-line detection of apnea/hypopnea events using SpO2 signal: a rule-based approach employing binary classifier models," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, pp. 231–239, Jan. 2014.

[99] C. Varon, A. Caicedo, D. Testelmans, B. Buyse, and S. V. Huffel, "A novel algorithm for the automatic detection of sleep apnea from single-lead ECG," *IEEE Transactions on Biomedical Engineering*, vol. 62, pp. 2269–2278, Sept. 2015.

[100] A. Roebuck, V. Monasterio, E. Gederi, M. Osipov, J. Behar, A. Malhotra, T. Penzel, and G. D. Clifford, "A review of signals used in sleep analysis," *Physiological Measurement*, vol. 35, no. 1, pp. R1–R57, 2014.

[101] J. E. Hall, *Guyton and Hall Textbook of Medical Physiology*. Saunders Elsevier, 2015.

[102] J.-A. Collins, A. Rudenski, J. Gibson, L. Howard, and R. O'Driscoll, "Relating oxygen partial pressure, saturation and content: the haemoglobin-oxygen dissociation curve," *Breathe*, vol. 11, no. 3, pp. 194–201, 2015.

[103] J. G. Webster, *Design of Pulse Oximeters*. CRC Press, 1997.

[104] M. A. Álvarez, L. Rosasco, and N. D. Lawrence, "Kernels for vector-valued functions: a review," *Foundations and Trends in Machine Learning*, vol. 4, no. 3, 2012.

[105] E. Snelson and Z. Ghahramani, "Sparse Gaussian processes using pseudo-inputs," in *Advances in Neural Information Processing Systems (NIPS)*, 2005.

[106] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[107] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Longman Publishing Co., 1989.

[108] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics (SIAM), 2002.

[109] J. Mercer, "Functions of positive and negative type, and their connection with the theory of integral equations," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 209, no. 441-458, pp. 415–446, 1909.

VITA

Sandeep Gutta

Candidate for the Degree of

Doctor of Philosophy

Thesis: BIOMEDICAL SIGNAL PROCESSING AND INFERENCE IN WEARABLE SENSING APPLICATIONS

Major Field: Electrical Engineering

Biographical:

Education:

Completed the requirements for the degree of Doctor of Philosophy in Electrical Engineering at Oklahoma State University, Stillwater, Oklahoma, USA, in December 2016.

Received the Master of Science degree in Electrical Engineering from Oklahoma State University, Stillwater, Oklahoma, USA, in 2011.

Received the Bachelor of Engineering degree in Electronics and Communication Engineering from Andhra University, Visakhapatnam, Andhra Pradesh, India, in 2008.

Experience:

2012 - 2016, Graduate Research Associate, Statistical Signal Processing Laboratory, Oklahoma State University, Stillwater, Oklahoma, USA.

2009 - 2011, Graduate Research Assistant, Statistical Signal Processing Laboratory, Oklahoma State University, Stillwater, Oklahoma, USA.

Professional Memberships:

Institute of Electrical and Electronics Engineers (IEEE)
Golden Key International Honour Society
Phi Kappa Phi